UNIVERSITY OF SAO PAULO
SCHOOL OF ARTS, SCIENCES AND HUMANITIES

STÉFFANI VIBANCO DE OLIVEIRA NEVES

Discovery of new epitopes of Trypanosoma cruzi in its interaction with humans.

Sao Paulo
2022

STÉFFANI VIBANCO DE OLIVEIRA NEVES

Discovery of new epitopes of Trypanosoma cruzi in its interaction with humans.

Translated from Portuguese version

Completion of course work presented to the Undergraduate Course in Biotechnology at the School of Arts, Sciences and Humanities of the University of São Paulo, to obtain the title of Bachelor of Biotechnology.

Advisor: Prof. doctor Joao Carlos Setubal

Co-advisor: Prof. doctor Luciano Antonio Digiampietri

Sao Paulo

2022

I dedicate this work to all the professors I had, who, without a doubt, made an infinitely greater contribution to this result than myself.

THANKS

First of all, I would like to express my sincere gratitude to my advisor, Prof. doctor João Carlos Setubal, as well as my co-advisor, Prof. doctor Luciano Antonio Digiampietri, and especially my doctoral student supervisor Gianluca Machado Major da Silva for his patience and support given to me throughout the project, without whom I would not have been able to complete this project.

Next, I would like to thank the masters, who guided us with so much affection during graduation, especially Prof. doctor Tiago Francoy, for teaching that graduation has much more to offer than theoretical classes and exams, and that you need to be able to balance responsibilities with relaxation, to doctoral student Celso Barbiéri, for following my path through graduation, always with a marsupial affection. Furthermore, I want to express my gratitude to Prof. doctor Felipe Chambergo, and to the entire committee of professors who accompanied me and gave me all the support I needed to be able to do my best within the Biotechnology course at the University of São Paulo.

Finally, I would like to thank my true friends, who walked the same arduous path with me, always offering support and a friendly shoulder in difficult times, and this, for me, played a key role in my development.

SUMMARY

Vibanco de Oliveira Neves; Stéffani. Discovery of new epitopes of Trypanosoma cruzi in its interaction with humans. Completion work of the bachelor's degree in biotechnology - School of Arts, Sciences and Humanity, University of São Paulo, São Paulo, 2022.

Large-scale mapping of antigens and epitopes is of fundamental importance for the development of various immunotherapies, but it becomes a major challenge, especially for eukaryotic pathogens, due to their large genomes. In this work, a process flow was developed, from genomic phages, to show that unbiased libraries of the eukaryotic parasite Trypanosoma cruzi allow the identification of antigens by serum samples from patients with Chagas disease. A comprehensive library of Chagas disease antibody response was constructed and validated, with the aim of showing how epitopes of linear and putative conformation (containing many repeating elements), allow the parasite to avoid an accumulation of neutralizing antibodies directed against domains of proteins that mediate the pathogenesis of the infection. Thus, this process flow is a reproducible and effective tool for the identification of epitopes and antigens, not only for Chagas disease, but perhaps also for emerging/reemerging pathogens globally.

Keywords: Antigens. Bioinformatics. Chagas Disease. Epitope. Phage. Immunotherapy.

ABSTRACT

Vibanco de Oliveira Neves; Stefani. Discovery of new epitopes of Trypanosoma cruzi in its interaction with humans. Completion work of the bachelor's degree in biotechnology – School of Arts, Sciences and Humanity, University of São Paulo, São Paulo, 2022.

Large-scale mapping of antigens and epitopes is of fundamental importance for the development of several immunotherapies, but it becomes a great challenge, especially for eukaryotic pathogens, due to their large genomes. In this work, a process flow was developed, starting from genomic phages, to show that unbiased libraries of the eukaryotic parasite Trypanosoma cruzi allow the identification of antigens by serum samples from patients with Chagas disease. A comprehensive library of the Chagas disease antibody response was constructed and validated, with the aim of showing how epitopes of linear and putative conformation (containing many repeated elements) allow the parasite to avoid an accumulation of neutralizing antibodies directed against protein domains. that mediate the pathogeny of the infection.

Keywords: Antigens. Bioinformatics. Chagas disease. epitope. Phage. Immunotherapy.

SUMMARY

1. INTRODUCTION

1.1. CHAGAS DISEASE

Chagas disease, also known as American trypanosomiasis, is a multisystem disorder that can affect the cardiovascular, digestive, and central nervous systems.[1] Chagas disease is caused by Trypanosoma cruzi, a hemoflagellate parasite that is transmitted by several species of insects hematophagous reduvids (kissing bugs) mainly in endemic areas.[2] The disease was described for the first time by Carlos Chagas in 1909, however Charles Darwin described his encounter with the vector and his own symptoms compatible with the disease, indicating that infection by T. cruzi happened before Dr. Chagas to describe.[3] The World Health Organization (WHO) considers Chagas disease one of the twenty neglected tropical diseases[4], and it is estimated that 6 to 7 million people are infected with T. cruzi worldwide, the vast majority in Latin America.[5]

Chagas disease can be considered a reemerging infection, as areas where there was no locally acquired infection are reporting autochthonous cases.[6]In the United States of America, when testing for Chagas disease in blood donors began in 2008, seropositivity for T. cruzi was 1 in 6,500 donors, with 36% of them having clinical evidence of Chagas cardiomyopathy. In at least 5 of these cases the infection is shown to have occurred as indigenous transmission in Texas.[7] Chagas disease is an important public health problem, affecting multiple systems, including the central nervous system (CNS), the digestive system, the immune system and, mainly, the heart. In Latin America, it is among the most frequent causes of heart failure (HF), and is supposedly responsible for up to 41% of cases in endemic areas.[8]

Parasite stages in mammalian hosts include bloodstream trypomastigotes and intracellular replicative, flagellaless, amastigotes, while vector stages include replicative epimastigotes and infectious metacyclic trypomastigotes. Infections in mammals occur when the parasite is in the trypomastigote stage. Trypomastigotes infect a variety of cells and convert into a replicative amastigote that multiplies in the host cell cytoplasm. The parasitized cells eventually rupture and release trypomastigotes that circulate and can infect other host cells.[9]

T. cruzi is transmitted in endemic areas by several species of three genera of triatomine blood-sucking insects, also known as kissing bugs (Triatoma, Panstrongylus, Rhodnius).[10]The three genera are widely distributed in Latin America, from Mexico to Argentina and Chile, and inhabit both forests and drier areas.[11] However, other infection mechanisms that are important, especially in non-endemic areas, include blood transfusion, organ transplantation, oral ingestion, laboratory accidents, mother-to-child vertical, or shared intravenous needles. Sexual transmission has been reported by in vivo mouse experiments, but no reports in humans are currently available.[12]

1.2.GENETIC DIVERSITY

T. cruzi is a heterogeneous species with seven strains, or discrete typing units (DTU), called TcI, TcII, TcIII, TcIV, TcV, TcVI and Tcbat.[13]This genetic diversity has been related to distribution, pathogenesis, clinical features and response to therapy. The parasites in each DTU are genetically similar and have similar characteristics, including the pathology they cause, biochemistry and immunogenicity, and resistance to treatment.[14]

TcI has a wide distribution, from the southern United States to northern Argentina and Chile, and this DTU is most frequently sampled in sylvatic cycles, but is also frequent in domestic cycles and is the dominant DTU responsible for disease transmission. Chagas disease in endemic countries located north of the Amazon basin. As for TcII, V and VI are more likely to be associated with domestic cycles and patients with chronic Chagas disease in Southern Cone countries and Bolivia. TcIII and IV are mainly sampled in the rainforest. And finally, Tcbat previously identified in bats, was recently found in humans. It's fineIt is known that several DTUs can coexist in the same vector and in a single host.[15]

The genetic variety presented by this parasite is also evident when analyzing the families of multigenes that it possesses. This occurs because the genome of T. cruzi has many repeated sequences, indicating that many of these genes are in linkage disequilibrium and that clonal reproduction of this population occurs. Furthermore, the presence of multigene families is related to its ability to invade cells and present tropism for different tissues, causing different types of heart diseases and mega syndromes associated with Chagas disease. Thus, this genetic variety is associated with the infectivity of T. cruzi, as many of these families code for genes present on the surface of the protozoan, such as trans-sialidases and mucins.

1.3. AUTOIMMUNITY IN CHAGAS DISEASE

T. cruzi has different escape strategies that allow it to evade the host's immune system, allowing its persistence and the establishment of chronic infection that leads to the development of chronic chagasic cardiomyopathy (CCC). The potent immune stimuli generated by the persistence of T. cruzi can result in tissue damage and an inflammatory response. In addition, molecular mimicry between parasite molecules and host proteins can result in cross-reaction with self molecules and, consequently, autoimmune features, including autoantibodies and self-reactive cells. Although controversial, there is evidence that demonstrates a role for autoimmunity in the clinical progression of CCC. Nonetheless,[16]

There are two mechanisms that try to explain autoimmunity in Chagas: one with the activation of B and T lymphocytes in an antigen-independent manner, and the other with molecular mimicry. The first mechanism, together with the presence of parasite antigens, can trigger major tissue damage, surpassing the self-tolerance threshold and inducing the production of autoantibodies. Mimicry, in turn, is due to the similarity between T. cruzi and human antigens, and thus cross-reaction of antibodies occurs.[16]

1.4. PHAGE DISPLAY

The phage display technique was initially developed and used to map antibody epitopes[17], identify antigens involved in diseases such as cancer[18], illnesses[19], and parasitic infections[20], including Chagas disease.[21]Phage Display involves inserting a DNA fragment into a genetically modified bacteriophage, which expresses a peptide on its viral capsid, so that the corresponding peptide (or antibody) encoded by the exogenous DNA fragment is displayed on the surface of the bacteriophage. If the peptide is a fragment of an antigen recognized by a given antibody, the bacteriophage particle can be captured by exposing the protein that interacts with the target ligand.[22]. In this affinity selection process called biopanning, the library is then presented with target molecules, usually immobilized on solid supports. Weak interactions between phage expressing the protein and the target are disrupted by successive washes, while phages containing molecules with high

target affinity are recovered by elution.[23]And thus the antigen is isolated from the pool of phage particles.[24]

1.5. IEDB (IMMUNE EPITOPE DATABASE)

The Immune Epitope Database (IEDB) is a free service with the aim of assisting immunological research. In it, it is possible to find results of more than 1.6 million experiments of adaptive immune response to epitopes, gathered mainly in the literature.[25]These data come from 19,500 publications, including all available literature from the inception of PubMed to the present. Searches are performed on PubMed every two weeks allowing for an update with new content.[25]

The IEDB has a great relevance for this project, it is from it that it becomes possible to validate that our data are about possible epitopes, since it is very likely that we will find epitopes that are already known. In addition, it is possible to have a better understanding of the characteristics of an epitope through a vast database.

1.6. CONTEXT OF THIS WORK

This course completion work is part of a collaborative project between the advisor Professor João Carlos Setubal and Professor Ricardo Giordano. This project, led by Prof. Ricardo, aims to identify T. cruzi epitopes using phage display. In this project, 8 datasets have been analyzed so far (Table 1).

A preliminary analysis of these data sets was published in the article "A refined genome phage display methodology delineates the human antibody response in patients with Chagas disease"[24]by Teixeira et al.,

This course completion work aims to refine the methodology used in the article by Teixeira et al., for a specific data set, the CCC_mild A and B set. This work is also associated with the doctoral work of student Gianluca Machado da Silva, guided by profs. Setubal and Giordano. Gianluca was a co-supervisor of this TCC work.

2. OBJECTIVES

2.1. MAIN GOAL

Generation of a new list of potential epitopes in the T. cruzi-human interaction, based on the analysis of a data set obtained by the Phage Display technique.

2.2. SPECIFIC OBJECTIVES

2.2.1. Familiarization with the following topics:

- Chagas disease and Trypanosoma cruzi;
- Phage display technique;
- Epitope concept;
- setulab computational environment;
- Sequence alignment;
- Sequence clustering;

●BLAST (Basic Local Alignment Search Tool);

●IEDB[25].

2.2.2. For the phage display CCC_mild A and B datasets: ●Extract inserts from reads;

●Comparing DNA sequences to each other for counting of frequency

●Remove inserts that do not map to the T. cruzi genome;

●Determination of the ORFs (Open Read Frame) of each insert;

●Comparison of ORFs with each other to remove duplications;

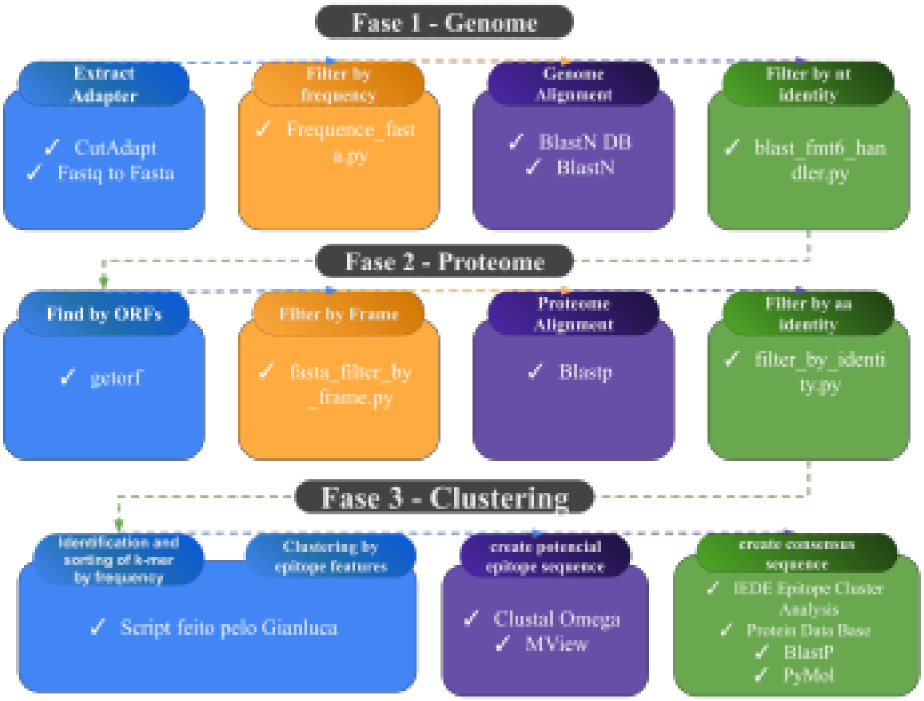
●Comparison of the resulting ORFs with T. cruzi proteome;

●Comparison of the resulting ORFs with T. cruzi epitopes in IEDB.

3. METHODOLOGY

The study was divided into 5 stages: Review and Understanding of the literature (3.1); Genome Treatment (3.2); Proteome Treatment(3.3); Validation (3.4), Clustering (3.5) and Sequence Consensus and Visualization in the protein (3.6). Among these steps, 3.2, 3.3, 3.5 and 3.6 form a pipeline for the identification of epitopes from antigen sequences, as shown in figure 1.

Figure 1: Flow diagram of pipeline steps for epitope identification from antigen sequences obtained using the Phage Display technique.



Source: Stéffani Vibanco de Oliveira Neves (2022).

3.1. LITERATURE REVIEW AND UNDERSTANDING

In order to obtain a better understanding of the processes and themes that were used, a survey of the scientific literature was carried out.

The understanding was fundamentally based on the article “A refined genome phage display methodology delineates the human antibody response in patients with Chagas disease” by Teixeira et al.[24], and also in the article “Protocol for design, construction, and selection of genome phage (gPhage) display libraries.” by Rodriguez Carnero et al..[26]

We decided to approach the issue from a set of data from a Phage Display library (Table 1) from the reference article.[24]With the intention of obtaining a greater genetic variety, the author used, as a source for this library, serum samples from patients contaminated by Chagas disease with different levels of symptoms, and who fit the following requirements.[24]

The. Patients with at least two positive results for the presence of anti-T. cruzi.

B. All candidate patients underwent electrocardiography (ECG) and echocardiography (ECHO) and those with abnormal ECG were classified as having mild cardiomyopathy when the left ventricular ejection fraction (LVEF) was greater than 40% (LVEF > 40%) , or severe cardiomyopathy, when the left ventricular ejection fraction was less than or equal to 40% (LVEF ≤ 40%).

ç. Patients without electrocardiographic changes were considered asymptomatic.

d. Serum samples were pooled into groups of 10 donors to form two independent sets (biological duplicates) for each disease condition:

i. control (2 x 10 donors)

ii. asymptomatic (2 x 10 donors)

iii. mild cardiomyopathy (2 x 10 donors)

iv. severe cardiomyopathy (2 x 10 donors).

Table 1: Available read sets. For each set (line) there are two lots (A and B)

control	control patients
asympto	patients who test positive for Chagas disease but have no symptoms
CCC_mild	patients who test positive for Chagas disease with mild symptoms
CCC_severe	patients who test positive for Chagas disease with severe symptoms

Source: , 2022[24]Stéffani Vibanco de Oliveira Neves

In this project, the set CCC_mild_A (also called K) and B (also called set O) was used.

3.2. DATA PROCESSING

3.2.1. EXTRACTING INSERTS FROM READS

From the download of the data set obtained by sequencing, it was necessary to remove the adapter sequences, these adapters contain the indices (short sequence of bases that identify each sample) and are at the beginning and end of the sequence. It was identified that for this sequencing, the sequence of adapters to be removed is from upstream “ATGACCATGGCAGTAC” and downstream “GTACCCGGTGCGCCGG” and for the removal, the command line tool Cutadapt was used.[27]

In addition, the command line converter from fastq to fasta file was used. In this way, we obtain data containing only the initial sequence of interest containing the possible T. cruzi antigens.

3.2.2. COMPARISON OF DNA SEQUENCES WITH EACH OTHER FOR FREQUENCY COUNTING

The sequences that are part of the set have great diversity. To ensure greater reliability, it was defined that they need to have a minimum frequency of two appearances, this information was observed in the article by Teixeira et al., which cites the reference by Dias-Neto et al, 2009.[28]For this filter, the script in Python was used[29]which checked the appearance of repeated sequences and created a new file containing only sequences with a minimum frequency of two.

3.2.3. REMOVAL OF INSERTS THAT DO NOT MAP IN THE T. CRUZI GENOME

In order to identify the sequences that represent the T. cruzi genome, an alignment of the possible antigen sequences with the T. cruzi sequences (CL Brener, Sylvio X10, DM28c, and Marinkellei strains downloaded from the NCBI[30]). For this, the blastN command line application was used.[31], with the parameters: number of alignments (num_alignments) in ten and maximum number of HSPs (High-Scoring Segment Pairs) in 1.

After the alignment, a Python script was used[32]to verify the percentage of identity of each sequence with the T. cruzi sequences and discard those that obtained a value lower than 90%. Ideally, we would use a value of 100% identity, but due to the variations of T. cruzi strains in the samples, and the one used for alignment, there was a greater tolerance for its identity.

3.3. PROTEOME TREATMENT

3.3.1. DETERMINATION OF ORFS OF EACH INSERT

The sequences that have been obtained so far consist of nucleotides, so for the construction of a proteome it was necessary to use a command line program called EMBOSS GetOrf[33]~~, with it, it was possible to obtain sequences of open reading frames (ORFs).~~

3.3.2. REMOVAL OF INSERTS THAT STARTED WITH A FRAME DIFFERENT FROM TWO

The inserts coming from the DNA phage assume that the correct frame for amino acid translation is frame two positive (+2). Therefore, in this step, it was necessary to discard the open reading frames (ORF) that did not start in frame +2, for which a script in Python was used[34]which filtered the sequences of ORFs that were in frame +2, creating a file with only the desired ones.

3.3.3. REMOVAL OF INSERTS THAT DO NOT MAP IN THE T. CRUZI PROTEOME

In order to identify the sequences that represent the T. cruzi proteome, an alignment of the possible antigen sequences with the T. cruzi sequences (strains CL Brener, Sylvio X10, DM28c, and Marinkellei downloaded from the NCBI[30]~~). For this, the blastP command-line application was used.~~[31]~~, with the parameters: number of alignments (num_alignments) in ten and maximum number of alignments (max_hsps) in 1, the size of matching sequences (word_size) of value 6, number of openings (gapopen) in 13 and the number of alignments that would be expected (evalue) at 100.~~

After the alignment, a Python script was used[35]to verify the percentage of identity of each sequence with the T. cruzi sequences and discard those that obtained a value lower than 70%. This value is due to the 90% filter made in step 3.2.3, since it was used for the nucleotide identity filter, there is a need for greater tolerance when dealing with peptides. This statement is justified by the occurrence of gene families, which may have similar proteins encoded by different genes. It is known that in the T. cruzi genome there are gene families, and therefore, using this tolerance, it was possible to map translated peptides in the proteome taking this phenomenon into account.

3.4. VALIDATION

With the intention of validating that the sequences that passed through the pipeline contain Trypanossoma cruzi epitopes, at this stage of the pipeline all epitope sequences that are present in the database of the IEDB website were downloaded.[reference], with the following parameters: Linear Epitope, Trypanossoma cruzi Organism, Human Host, Chagas disease and other parameters with standard values already inserted by the platform.

After unloading, a Blast database was created with these sequences, and thus the IEDB epitope sequences were aligned with the possible antigen sequences from previous flows, using the BlastP command-line application.[31]

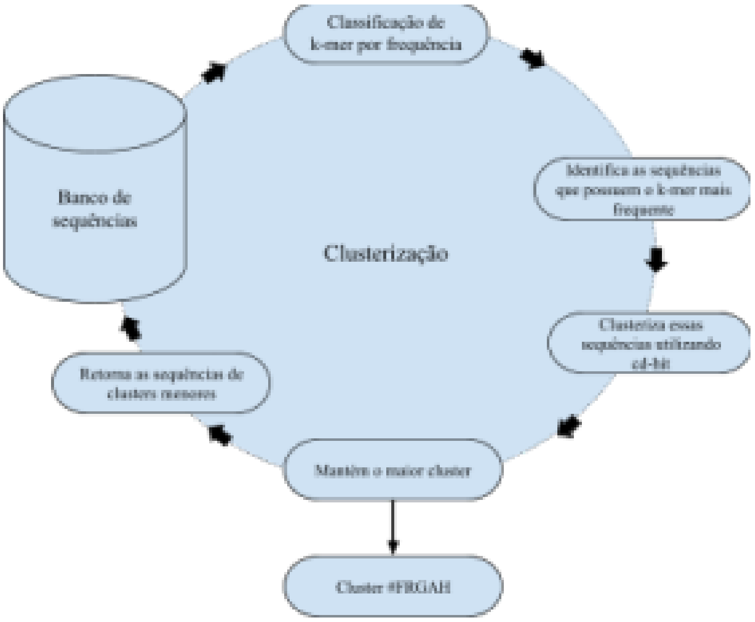
3.5. Clustering

In order to obtain an optimized analysis, it was decided to group the sequences resulting from step "3.3 treatment of the proteome" into clusters. For this, a script in Python was structured[36]which, from the sequence bank, detects subsequences (k-mer) of size eight of each sequence, and orders these k-mers, in descending order, according to their frequency of appearance, from the highest to the lowest. So he uses the cd-hit app[37]~~, which aims to cluster protein sequences with at least 80% identity , using the sequences that have the appearance of this most frequent k-mer.~~

From this, we obtain different clusters arising from the sequences used. The cluster that obtained a greater number of grouped sequences will be considered for final analysis, while the sequences that are part of clusters with a smaller number will return to the sequence bank to be reconsidered by another k-mer, following their frequency order, asfigure 2.

As a result of this script, folders were created for each k-mer sequence, in which there are clusters created by cd-hit, including the cluster with the highest number of sequences. In addition, a log file was created that brings information about the occurrences of the script, a file containing information about the largest clusters of each k-mer.

Figure 2: Clustering process from Python script


Classificação de k-mer por frequência
Identifica as sequências que possuem o k-mer mais frequente
Clusteriza essas sequências utilizando cd-hit
Mantém o maior cluster
Cluster #FRGAH
Retorna as sequências de clusters menores
Banco de sequências
Clusterização

Source: Stéffani Vibanco de Oliveira Neves (2022).

3.6. POSSIBLE EPITOPE SEQUENCE, CONSENSUS SEQUENCE AND VISUALIZATION IN PROTEIN

From the document that contains information about the largest clusters of each k-mer, a cluster was selected in which its possible epitope was already cataloged in the IEDB database.

So, in order to obtain a sequence that possibly represents an epitope of this cluster, the program Clustal Omega was used[38]containing the complete sequences of it. The result was visualized by the other application, Mview[38]. So, in order to verify if this possible epitope is already cataloged in the IEDB, we looked for this sequence in the result of the BlastP with the IEDB (done in step “3.4 - Validation”).

On the other hand, to obtain a consensus sequence of this cluster, the Epitope Cluster Analysis web program was used with the default parameters.[39]~~, containing the fasta with complete sequences of it. The result was the consensus sequence of sequences.~~

Upon obtaining the 100% consensus sequence of this cluster, the BlastP tool was used[31]of the NCBI, using several protein databases, to identify

which protein of the Trypanosoma cruzi organism could correspond to this19 consensus sequence. Then, the protein that contained the greatest correspondence with the possible epitope was selected.

From this, this protein was found in UniProt[40], a protein database that contains, in addition to proteins with validated structures, predictions of non-validated ones. So, we use the PyMol Desktop application[41]for the visualization of this protein, and therefore, the "protein residue sequence selection" function with a single letter code was used to select the part of the protein in which there was an alignment with the consensus sequence, this part was colored for better visualization.

4. RESULTS and DISCUSSION

4.1. SKILLS WITH PROGRAMMING LANGUAGES AND HIGH PERFORMANCE COMPUTING SYSTEMS

Scripts were made in Python that use the argparse modules[42]~~, which make it possible to create user interfaces from the command line. In addition, modules from the biopython library were used.~~[43]~~, which is a tool library for biological data, pandas library modules~~[44]which is used for data manipulation and analysis. The scripts are stored on github by doctoral student Gianluca Major and the author herself, and are being produced throughout doctoral student Gianluca Major's project.

Also, the Google Drive drawing tool was used.[45]to make infographics that facilitated the understanding of the process flows in this project.

4.2. GENOME TREATMENT

Sequences obtained from the library came from patients who tested positive for Chagas disease with mild symptoms. The CCC_mild_A sequence group has 868 thousand sequences, while the CCC_mild_B group has 651 thousand sequences. As the sequences walk through the proposed pipeline, there is a change in the number of sequences in each group, due to filters and alignments that provide sequences that are increasingly close to the epitopes of the Trypanossoma cruzi organism, these changes can be seen in thetable 2.

Initially, the adapters of each sequence were removed, so it was possible to obtain only the nucleotides of interest that represent the possible20 antigens, therefore, sequences that are unique were eliminated, that is, they presented only one recurrence in the entire bank of sequences, and thus errors in the phage display process were reduced, increasing the confidence of the data. At this stage, the number of sequences in group A and B were, respectively, 48435 and 21154 sequences.

Then, after alignment between the sequences of groups A and B, with the sequences of Trypanossoma cruzi already known, provided by the National Center for

Biotechnology Information (NCBI) and filtering by 90% identity, both groups A and group were obtained. B sequences (Table 2) that are more likely to belong to the organism of interest.

4.3. PROTEOME TREATMENT

The sequences, until then of nucleotides, were transformed into Open Reading Frames (ORF), obtaining sequences between the start and stop codons, as the number of sequences in group A and B were, respectively, 414251 and 175888, astable 2. After that, there was a filter in which ORFs that did not start at frame +2 were discarded, which resulted in a large decrease in the number of sequences in groups A and B, respectively. 35231 and 16873.

So, after the alignment between the proteomes of groups A and B, with Trypanosoma cruzi proteomes already known from the NCBI, and a filtering by 70% identity, we obtained 13569 sequences in group A, and 12162 sequences in group B.

Table 2: Sequences of possible Trypanosoma cruzi antigens according to their passage through the pipeline.

Phases	ccc_mild_a (K)	ccc_mild_n (O)
Original Sequences	868305	651503
Deletion of Unique Strings	48435	21154
Sequences with 90% Identity after BlastN	48118	21138
Orfs that start with frame 2	35231	16873
Sequences with 70% Identity after BlastP	13569	13569

Source: Stéffani Vibanco de Oliveira Neves, 2022.

4.4. VALIDATION

The sequence database of groups A and B so far has sequences with possible Trypanosoma cruzi antigens, so an alignment was made with the epitope sequences that are present in the IEDB. As shown in Table 3, we can see that of the sequences in group A that passed through the proposed pipeline, 14.68% have at least one epitope already known by the IEDB, while in group B, there are 77.35%. By excluding repeated epitopes we obtain the amount of unique IEDB epitopes that were found in the sequences of this project. (Table 3)

Table 3: Sequences of possible Trypanosoma cruzi antigens according to their passage through the pipeline.

Description	Total of antigens	Alignment with at least one IEDB epitope with 100% identity	Number of unique epitopes in the IEDB
CCC_mild_a	13569	1993	376
CCC_mild_b	12162	9408	354

Source: Stéffani Vibanco de Oliveira Neves, 2022.

4.5. Clustering

When we passed this project's sequence database through the clustering script explained in Section 3.5, we obtained 637 clusters for group A, and 66 for group B, each with its representative sequence. With this result, we can observe that in group B, there is a greater similarity between the sequences it contains than in group A. As the number of IEDB epitopes that were identified for both groups is similar, after all they are the same biological samples , this result suggests that in group A there are many sequences that are not antigens. Indeed, in the article by Teixeira et al., it is mentioned that this divergence may be due to differences in the phage “input” on the phage display, which was lower for the selection of group B. However, 70% of all epitopes of group B were contained in A,

4.6. POSSIBLE Epitopes

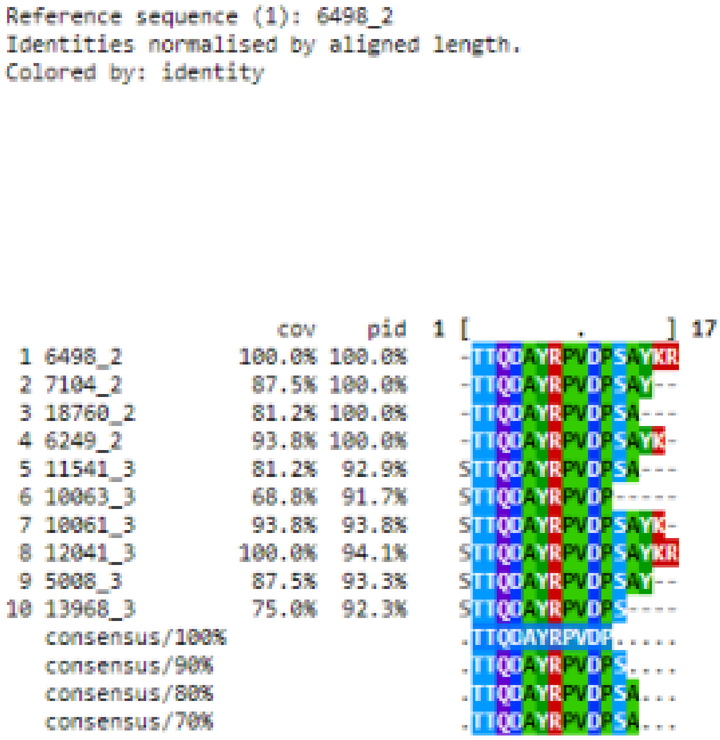
At the end of clustering, we obtained a list of possible epitopes, both from group A and group B. This list is found respectively in appendices A and B of the work.

In these appendices there is a table with several data that are relevant for the selection of clusters and possible epitopes. In the first column of the table found in the appendices, is the K-mer value, which has a chosen size of 8 amino acids that has been listed according to its frequency, the next column is that frequency, then the next column represents the unique sequences that have the k-mer, after that the posterior column represents the amount of sequences that are present in the largest cluster of that k-mer group. The consecutive column has 4 values, represented by the number of sequences with the lowest frequency, the highest frequency, the average size and the standard deviation. In addition, the following columns contain the identification number of the largest Cluster of this k-mer group,

4.7. CONSENSUS SEQUENCE AND VISUALIZATION IN PROTEIN

When obtaining the largest clusters of each k-mer group, a cluster whose possible epitope was already cataloged in the IEDB database was chosen. The selected cluster was the one containing the k-mer “TTQDAYRP”.

In the TTQDAYRP cluster, there are ten sequences that contain your k-mer and ten that have been grouped into a main cluster, that is, all initial sequences. Then, when using Clustal Omega and obtaining a visualization from Mview, we obtained that the closest sequence of an epitope of this cluster is “TTQDAYRPVDP”, asfigure 3.



```
Reference sequence (1): 6498_2
Identities normalised by aligned length.
Colored by: identity



                          cov     pid   1 [        .       ] 17
 1 6498_2            100.0% 100.0%     -TTQDAYRPVDPSAYKR
 2 7104_2             87.5% 100.0%     -TTQDAYRPVDPSAY--
 3 18760_2            81.2% 100.0%     -TTQDAYRPVDPSA---
 4 6249_2             93.8% 100.0%     -TTQDAYRPVDPSAYK-
 5 11541_3            81.2%  92.9%     STTQDAYRPVDPSA---
 6 10063_3            68.8%  91.7%     STTQDAYRPVDP-----
 7 10061_3            93.8%  93.8%     STTQDAYRPVDPSAYK-
 8 12041_3           100.0%  94.1%     STTQDAYRPVDPSAYKR
 9 5008_3             87.5%  93.3%     STTQDAYRPVDPSAY--
10 13968_3            75.0%  92.3%     STTQDAYRPVDPS----
   consensus/100%                      .TTQDAYRPVDP.....
   consensus/90%                       .TTQDAYRPVDPS....
   consensus/80%                       .TTQDAYRPVDPSA...
   consensus/70%                       .TTQDAYRPVDPSA...
```

Figure 3: Result of Multiple Alignment using Clustal Omegaand Mview.

Source: Mview, 2022.

From the comparison made of the closest sequence of an epitope, with the BlastP result of the IEDB epitopes (seen in step 3.4 of Validation), it was possible to identify that this cluster represents an epitope that is already in the IEDB sequence database “TQDAYRPVDPSAYKR” and identification number“397929”, and therefore is valid.

From another perspective, the consensus sequence was found using the Epitope Cluster Analysis web tool.[39]The sequence is configured by “STTQDAYRPVDPSAYKR” as can be seen in thetable 4.

Table 4: Result of Clustering Epitopes of Cluster “TTQDAYRP”

Sequences	Peptide Number	Alignment	Position	Description	Peptide
1.1	**Consensus**	**STTQDAYRPVDPSAYKR**	-	-	-
1.1	1	STTQDAYRPVDPSAYKR	1	12041_3 [2 - 52] | freq=3	STTQDAYRPVDP SA YKR
1.1	two	STTQDAYRPVDPSAYK-	1	10061_3 [2 - 49] | freq=3	STTQDAYRPVDP SA YK
1.1	3	STTQDAYRPVDPSAY--	1	5008_3 [2 - 46] | freq=7	STTQDAYRPVDP SA Y
1.1	4	STTQDAYRPVDPSA---	1	11541_3 [2 - 43] | freq=3	STQDAYRPVDPSA
1.1 1.1 1.1	5 6 7	STTQDAYRPVDPS-- -- STTQDAYRPVDP--- -- -TTQDAYRPVDPSAY KR	1 1 two	13968_3 [2 - 40] | freq=2 10063_3 [2 - 37] | freq=3 6498_2 [2 - 49] | freq=5	STTQDAYRPVDPS STTQDAYRPVDP TTQDAYRPVDPS AY KR

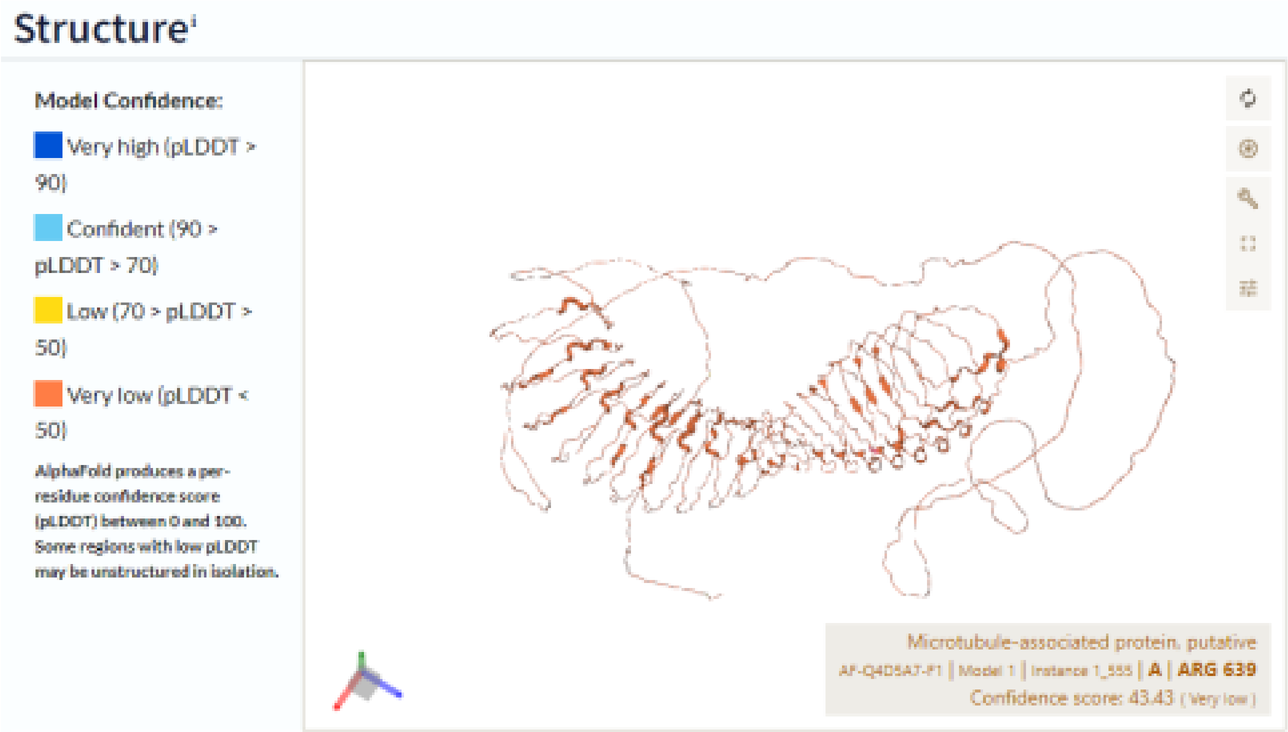
1.1	8	-TTQDAYRPVDPSAYK-	two	6249_2 [2 - 46] | freq=5	TTQDAYRPVDPSAY K
1.1	9	-TTQDAYRPVDPSAY--	two	7104_2 [2 - 43] | freq=5	TTQDAYRPVDPSAY
1.1	10	-TTQDAYRPVDPSA---	two	18760_2 [2 - 40] | freq=2	TTQDAYRPVDPSA

Source: Epitope Clustering, 2022.

From the consensus sequence, the BlastP tool was used[31]from the NCBI, to obtain the proteins that most closely align with this sequence. There were 45 proteins that obtained 100% identity with the sequence, however, from the data obtained from the corresponding epitope found in the IEDB, the protein that best represents this possible epitope was the 6th in the BlastP list, called “microtubule-associated protein” code XP_809567.1, from the strain of T. cruzi CL Brener.

Then the protein was found in the UniProt database, where it was possible observe a structural prediction of the molecule, as it is possible to observe in theFigure 4.

Figure 4: Prediction of the Structure of the Protein “Microtubule Associated Protein”



Structure

Model Confidence:

🟦 Very high (pLDDT > 90)

🟦 Confident (90 > pLDDT > 70)

🟨 Low (70 > pLDDT > 50)

🟧 Very low (pLDDT < 50)

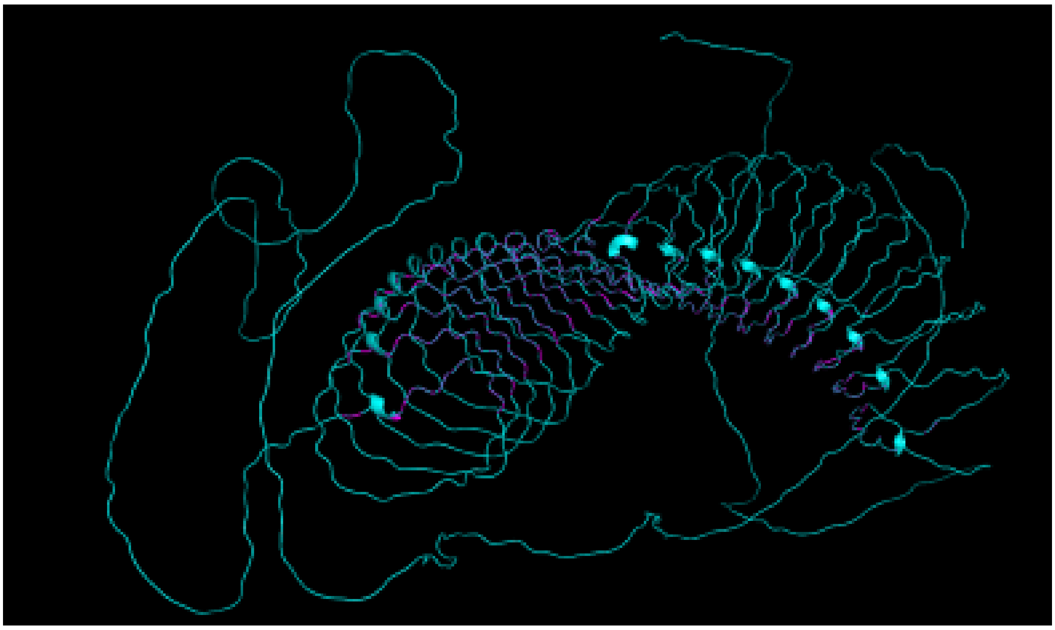
AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.

Microtubule-associated protein, putative
AF-Q4D5A7-F1 | Model 1 | Instance 1_555 | A | ARG 639
Confidence score: 43.43 (Very low)

Source: UniProt 2022 the unreliable in yellow and the very unreliable in orange.

Then, the protein was opened in the PyMol viewer, and the representative part of the “STTQDAYRPVDPSAYKR” consensus sequence that aligns in the protein was colored, as seen in thefigure 5.

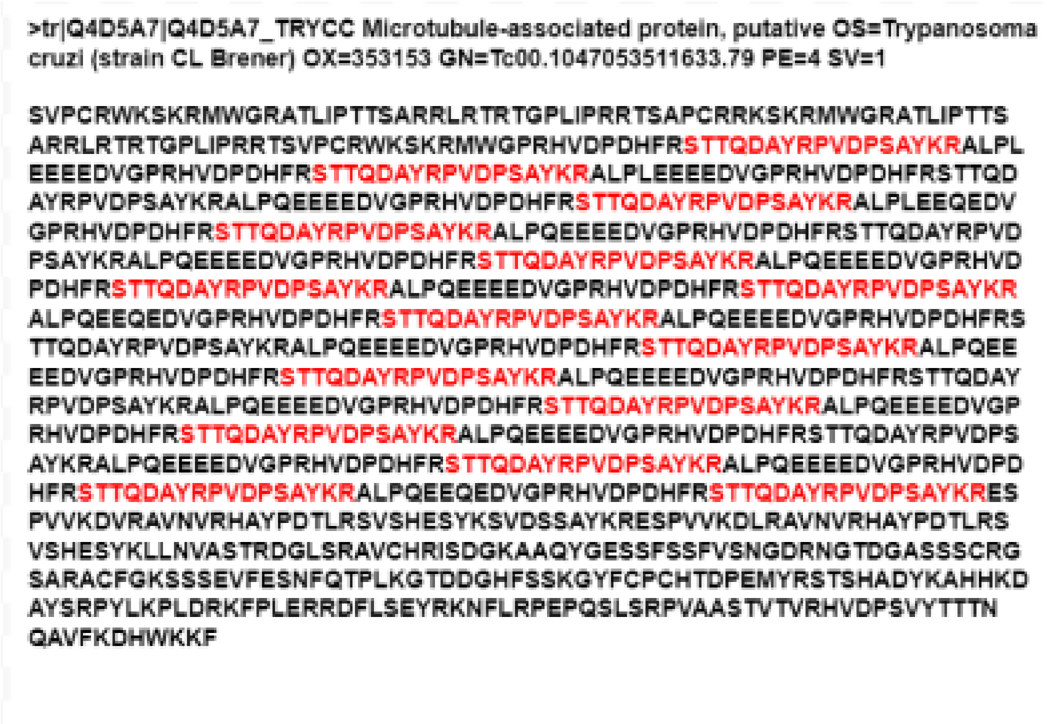
Figure 5: Consensus sequence represented in the molecular structure of its corresponding protein.



Source: Stéffani Vibanco de Oliveira Neves via PyMol, 2022 . Caption: The protein is represented by the cyan color, while the consensus sequence is represented by the magenta color.

Therefore, it is also possible to identify the consensus sequence contained in the complete protein sequence, asFigure 6.

Figure 6 : Complete protein sequence highlighting the location of the consensus sequence



```
>tr|Q4D5A7|Q4D5A7_TRYCC Microtubule-associated protein, putative OS=Trypanosoma
cruzi (strain CL Brener) OX=353153 GN=Tc00.1047053511633.79 PE=4 SV=1

SVPCRWKSKRMWGRATLIPTTSARRLRTRTGPLIPRRTSAPCRRKSKRMWGRATLIPTTS
ARRLRTRTGPLIPRRTSVPCRWKSKRMWGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPL
EEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPLEEEEDVGPRHVDPDHFRSTTQD
AYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPLEEQEDV
GPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVD
PSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVD
PDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKR
ALPQEEQEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRS
TTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEE
EEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAY
RPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGP
RHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPS
AYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPD
HFRSTTQDAYRPVDPSAYKRALPQEEQEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRES
PVVKDVRAVNVRHAYPDTLRSVSHESYKSVDSSAYKRESPVVKDLRAVNVRHAYPDTLRS
VSHESYKLLNVASTRDGLSRAVCHRISDGKAAQYGESSFSSFVSNGDRNGTDGASSSCRG
SARACFGKSSSEVFESNFQTPLKGTDDGHFSSKGYFCPCHTDPEMYRSTSHADYKAHHKD
AYSRPYLKPLDRKFPLERRDFLSEYRKNFLRPEPQSLSRPVAASTVTVRHVDPSVYTTTN
QAVFKDHWKKF
```

Source: Stéffani Vibanco de Oliveira Neves, 2022. Reference: UniProt

5. CONCLUSION

In summary, a bioinformatics pipeline was used in the samples, obtained from the Phage Display in the article by Teixeira et al, in which there was an initial treatment of the genome, from conversions, frequency filter, nucleotide alignment and identity filter; a treatment of the proteome, from obtaining ORFs, frame filter, protein alignment and identity filter; And finally, a clustering and data analysis, from obtaining k-mers, then clustering and identification of potential epitope sequences and a consensus sequence.

Finally, it is possible to observe that bioinformatics is an invaluable tool for the production of new technologies aimed at human health. We also conclude that the flow traversed in this project has great potential for a possible identification of epitopes.

As for the list of epitopes generated by the pipeline, we can see that there is a large amount of sequences that were not found in the IEDB, showing a possibility of identifying new epitopes.

With regard to the cluster chosen for further analysis, it was possible to notice that despite the epitope being available in the database, there are few tests on its biological and structural function. This pipeline has the ability to illuminate the possible epitopes, calling the attention of the scientific community and instigating in-depth, individual research and possible applications in the health area of each epitope.

6. LIMITATIONS OF THIS WORK

As with any new methodology, complete epitope validation should ideally be confirmed experimentally in an independent patient validation cohort.

Within this framework, the provisional antigens presented (Appendix AandB) should only be considered as candidates until they are unequivocally proven experimentally.

7. REFERENCES

1. Fernandes HJ, Barbosa LO, Machado TS, et al. Meningoencephalitis caused by reactivation of chagas disease in patient without known immunosuppression. Am J Trop Med Hyg 2017;96(2):292–4.

2. Perez CJ, Lymbery AJ, Thompson RCA. Reactivation of chagas disease: implications for global health. Trends Parasitol 2015;31(11):595–603.

3. Bernstein R. Darwin's illness: Chagas' disease resurgens. JR Soc Med. 1984;77:608–609.

4. World-Health-Organization. Neglected tropical diseases. In: World-HealthOrganization, ed. Vol http://www.who.int/neglected_diseases/diseases/en/. Geneva; 2018: Accessed March 20, 2022.

5. World-Health-Organization. Chagas disease (American trypanosomiasis). Available in:

http://www.who.int/news-room/fact-sheets/detail/chagas-disease-(americantrypanosomiasis. Accessed on March 20, 2022.

6. Garcia M, Woc-Colburn L, Aguilar D, Hotez P, Murray K. Historical perspectives on the epidemiology of human Chagas disease in Texas and recommendations for enhanced understanding of clinical Chagas disease in the Southern United States. PLoS Negl Trop Dis. 2015;9:e0003981.

7. Garcia M, Aguilar D, Gorchakov R, et al. Evidence of autochthonous Chagas disease in southeastern Texas. Am J Trop Med Hyg. 2015;92:325–330

8. Bocchi EA. Heart failure in South America. Curr Cardiol Rev 2013;9(2):147–56

9. Guarner, Jeannette. "Chagas disease as an example of a reemerging parasite." Seminars in Diagnostic Pathology. Vol. 36. No. 3. WB Saunders, 2019. 10. Echeverria, LE, & Morillo, CA (2019). American trypanosomiasis (Chagas

disease). Infectious Disease Clinics, 33(1), 119-134.

11. Yamagata Y, Nakagawa J. Control of Chagas disease. Adv Parasitol 2006; 61: 129–65.
12. Pérez-Molina, JA, & Molina, I. (2018). Chagas disease. The Lancet, 391(10115), 82-94.
13. Brenière, SF, Waleckx, E., & Barnabas, C. (2016). Over six thousand Trypanosoma cruzi strains classified into discrete typing units (DTUs): attempt at an inventory. PLoS neglected tropical diseases, 10(8), e0004792.
14. Zingales B, Miles MA, Campbell DA, et al. The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol 2012; 12: 240–53.
15. Brenière, SF, Waleckx, E., & Barnabas, C. (2016). Over six thousand Trypanosoma cruzi strains classified into discrete typing units (DTUs): attempt at an inventory. PLoS neglected tropical diseases, 10(8), e0004792.
16.De Bona E, Lidani KCF, Bavia L, Omidian Z, Gremski LH, Sandri, TL, & Messiah Reason, IJD (2018). Autoimmunity in chronic Chagas disease: a road of multiple pathways to cardiomyopathy?. frontiers *in Immunology, 1842.*
17. Smith, GP (2019). Phage display: simple evolution in a petri dish (Nobel Lecture). Angew. chem. Int. Ed. English 58, 14428–14437.

18. Mistry, J., Finn, RD, Eddy, SR, Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 41, e121.

19. Zhang, W., & Reichlin, M. (2005). A peptide DNA surrogate that binds and inhibits anti-dsDNA antibodies. Clinical Immunology, 117(3), 214-220. 20. Ellis, SE, Newlands, GF, Nisbet, AJ, and Matthews, JB (2012). Phage-display library biopanning as a novel approach to identifying nematode vaccine antigens. Parasite Immunol. 34, 285–295.

21. Alvarez, P., Leguizamo´n, MS, Buscaglia, CA, Pitcovsky, TA, and Campetella, O. (2001). Multiple overlapping epitopes in the repetitive unit of the shed acute-phase antigen from Trypanosoma cruzi enhance its immunogenic properties. Infect. Immun. 69, 7946–7949.
22. POSNER, B.; SMILEY, J.; LEE, I.; BENKOVIC, S. Catalytic antibodies:Pusing combinatorial libraries. Trends in Biochemical Sciences. v.19, n.4,P145-150, 1994.

23. PANDE, J.; SZEWCZYK, MM; GROVER, AK Phage display: concept, innovations, applications and future. Biotechnology Advances. v.28, n.6,P.849-858, 2010.

24. Teixeira AAR, Carnero LR, Kuramoto A., Tang FHF, Gomes C. H., Pereira, NB, ... & Giordano, RJ (2021). A refined genome phage display methodology delineates the human antibody response in patients with Chagas disease. Iscience, 24(6), 102540.

25. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update.Noucleic Acids Res. 2018 Oct 24. doi: 10.1093/nar/gky1006. [Epub ahead ofPrint] PubMed PMID: 30357391.

26.LA Rodriguez-Carnero, AAR Teixeira, FHF Tang, A. Kuramoto, MJM Alves, W. Colli, JC Setubal, E. Cunha-Neto, R. Pasqualini, W. Arap, RJ Giordano. Protocol for design, construction, and selection of genome phage (gPhage) display libraries. STAR Protocols Volume 2, Issue 4, 100936, 2021.

27.MARTIN, Marcel. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, [Sl], v. 17, no. 1, p. pp. 10-12, May 2011. ISSN 2226-6089. Available in:

<https://journal.embnet.org/index.php/embnetjournal/article/view/200>. Date Accessed: 12 Oct. 2022. doi:https://doi.org/10.14806/ej.17.1.200. 28. Dias-Neto, E., Nunes, DN, Giordano, RJ, Sun, J., Botz, GH, Yang, K., Setubal, JC, Pasqualini, R., and Arap, W. (2009). Next-generation phage display: integrating and comparing available molecular tools to enable cost-effective high-throughput analysis. PLoS One 4, e8338.

29. Major, Gianluca. Frequency_Fasta.py. 2022. Available at: <https://github.com/gianlucamajor/pknife> Accessed 11 Nov. from 2022. 30. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; (1988) Available from: https://www.ncbi.nlm.nih.gov/.

31. Madden T. The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from:http://www.ncbi.nlm.nih.gov/books/NBK21097/

32. Major, Gianluca. Blast_fmt6_parse.py 2022. Available at: <https://github.com/gianlucamajor/pknife> Accessed 11 Nov. from 2022. 33.

Rice P., Longden I. and Bleasby A. GETORF FROM EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics. 2000 16(6):276-277

34. Neves, Stéffani. Fasta_filter_by_frame.py. 2022. Available at: <https://github.com/svibanco/tcc> Accessed 11 Nov. from 2022. 35. Neves, Stéffani. Filter_by_identity.py. 2022. Available at: <https://github.com/svibanco/tcc> Accessed 11 Nov. from 2022. 36. Major, Gianluca. main.py 2022. Available at: <https://github.com/gianlucamajor/pknife> Accessed 11 Nov. from 2022.37. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012 Dec 1;28(23):3150-2. doi: 10.1093/bioinformatics/bts565. Epub 2012 Oct 11. PMID: 23060610; PMCID: PMC3516142.

38. Madeira F, Pearce M, Tivey RNA, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Research. 2022 Apr:gkac240. DOI: 10.1093/nar/gkac240. PMID: 35412617; PMCID: PMC9252731

39. Sandeep Kumar Dhanda, Kerrie Vaughan, Veronique Schulten, Alba Grifoni, Daniela Weiskopf, John Sidney, Bjoern Peters, Alessandro Sette: Development of a novel clustering tool for linear peptide sequences. Immunology (2018) doi:https://doi.org/10.1111/imm.12984(ahead of prints) PMID: 30014462

40. Wang Y, Wang Q, Huang H, Huang W, Chen Y, McGarvey PB, Wu CH, Arighi CN, UniProt Consortium. A crowdsourcing open platform for literature curation in UniProt. Plos Biology. 19(12):e3001464 (2021)

41. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. 42. Van Rossum, G., & Drake Jr, FL (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.

43. Cock PA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B and de Hoon MJL (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics . Bioinformatics, 25, 1422-1423

44. The pandas development team. (2022). pandas-dev/pandas: Pandas (v1.5.1). Zenodo.https://doi.org/10.5281/zenodo.7223478

45. Google Drawings. Google. Version 8.1. 2022. Available at:https://docs.google.com/drawings/

```
K-mer Frequency Unique Sequences Sequences in TheLlargest Cluster Low Frequency | Bigger Frequency | Medium Size | Standard Deviation Cluster ID Representative Sequence ID Sequence QAAAGDKP 2013 600 582 14|81|51.0|15.76 0 44523_8
SPFGQAAAGDKPSPFGQAAAGDKPSPFGQAAAGDKPSLFGQAAAGDKPPLFGQAAAGDKPSPFGQAAAG KVAEAEKQ 1045 397 359 20|97|55|14.74 0 27645_8
EAAKAVETEKQRAAEAMKVAEAEKQRAAEAMKVAEAEKQKAAEAAKAVETEKQRAAEATKVAEAEKQRAAEAMKVAEAEKQRAAEATKVAEAEKQKA
MRRGQQAQ 945 945 228 21|107|62.0|19.86 2 5433_3 ARESADNMRRGQQAQPHSPHTTRRHPRNEGTQRQQPPTMQEEKTSTTISPSSFPQPQGNGKTPRPPHNAHGASEPSSIHHAAQKFPVIPPSMGHGAQLSAVTRGAPH PQHRPPPT 747 747 745
25|100|71|16.12 0 41354_5 AASTRRPGSPSNKTPPRSPPSTRIRSAQAPQWSPTRSQPHAQTPRACTAGSRCQQTPPQSRRWPQHRPPPTRTAPAPSALTARPSTALRATTAPCRQQTT RGQQAQPH 741 741 174 18|110|57.0|17.7 1 45241_4
KRMEREGEGRSTADTARESADSMRRGQQAQPHSPHTTRRHPRNEGTQRQQPPTMQEEKTSTTISPSSFPRPPHNAHGAGKHSSIHHAAQNFLPSRHQWDTEPNSAQSHAV PGVFETTG 724 724 621 25|67| 48|10.63 0 13669_6
TIQRLVRAMEGATDMPVACTPRVTEGLRLVDGRFSTKMPEERCTPGVFETTGLRLIDDVGSDAVLQW RAQELARE 572 486 339 53|103|76|9.31 1 48266_7
NARAQELAREKKLADRAFLDQKPEGVPLRELPLDDDSDFVAMEQERRQLLEKDPRRNAREIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDD
RGQQAQPH 567 567 126 22|104|55.0|20.89 1 45610_3 RERDGVQRIQRERADNMRRGQQAQPHSSHTTRRHPRNEGTQRQQPPTMQEEKASTTISPSSSLQPQGNGQTPRPPQQPHAPANPHPSMTRQQNLLPSRPQWDTE HSSHTTRR 487 487 130
21|99|79.0|17.63 2 17653_5 RSGVAGVSQLISDTEGKEWNERKRDGAQRIQRESADNMRRGQQAQKHSSHTTRRHPREEGTQRQQPPTMQEEKTSTTISPSSFPQPPHNAHGASEPSSI RETEHSGY 448 448 237 12|46|14|2.51 0 31889_3
TGGRETEHSGYSERERRQHEERAAGPATQSTHHETASTRGGHTTPA AKAAAAPA 441 84 83 18|80| 51|15.0 0 9036_3 AAPAKAATAPAKAATAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAP RGQQAQPH 407
407 91 42|101|80|12.37 1 33559_5 RSGVAGVSQLISDTEGKEWNERERDGAQRIQRESADKMRRGQQAQPHSPHTTRRHPRNEGTQRQQPPTMQEDNTSTTISPSSFPRPPHNAHGAGKPSSIHH
PAAKPAAK 367 198 193 12|70|26|9.33 0 20739_3 GSMPAKSANKPASKPAAKPAAKPAAKAPAPKAEKKGAAKAPAPKAAAPAPKAAAAAPKPAVRDAKQRSDA PAAGGFGS 355 152 138 17|82|52.0|14.66 0 28449_3
ATTTSAPAAGGFGSATTTSAPAAGGFGSAAHTSTPAVGGFGSATTTSAPAAGGFGSATTTSAPAVGGFGSAAHTSTPAAGGG HSSHTTRR 354 354 103 44|89|63|8.68 8 41790_3
RSGVAGVSQLISDTEGKEWNRRERNRAQRIQRERVDNTKRGQQAQPHSSHTTRRHPPPPSPPAHFHNPREMAESPAATTQHTRRKRTLS QPHSSHTT 276 276 74 19|116| 103.5|26.36 1 34204_6
RSGVAGVSQLISDTEGKEWNERERDGAQRIQRERAENMRRRQQAQPHSSHTTRRQPPPSPPAHSRSPGKWPKPRGHHTTHTEQAPTTPATRASEPSSIHDAATKFTAIPPSMGHGA RGQQAQPH 241 241 61 34|99|43|18.56 1 2349_4
VSQKEKEWNRRERDGVQRIQRERADNMRRGQQAQPHSPHTTRGHPRNEGTQRQQPPTMQEEKTSTTISPSSSLQPQGNGQNTAATIPATRASEPSSIND GNAGGLAP 212 212 212 27|75|51.0|5.91 0 44846_5
PQTVPQPAPETKAPPQSPCCDKRAGNAGGLAPHFRFGRPDRKKDTAEETRTGRPRAVCPKHRHGIGESVTLINGC RETEHSGY 211 211 211 15|27|23|1.18 0 34666_1 SVGQKKKNGTGGRETEHSGYSERAQTK EQERRQLL 191 186 146 28|105|59.0|14.62 0
45628_9 QATVNAMPAKKTEESYPYIEANPEGIHKQHLELNKDSKFLALEQERRQLLEKDPRRNAREIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDDS VDPDHFRS 191 162 157 21|82|43|12.49 0 26659_6
PDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEQEDVGPRHVDPDHFRS HSSHTTRR 184 184 92 22|98|63.0|15.79 1 20148_5
RNRAQRIQRERVDNTKRGQQAQPHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSVSNDAATKFTLPSSPHWNTEPNPTQPQEKRRGL LEQKAAEN 170 49 49 36|70|55|6.69 0 43642_2
RLAEELEQKAAENEKLADELEQKAAENERLADELEQKAAENEKLADELEQKAAENERLAEELEQKAAENE ANGIRPQP 149 149 60 37|87|65.5|14.44 4 43425_3
RHGPQPGRQLHTSPQTHTQPAVTHQGRRGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEPHAEAIPHPSTTWRQNFQPFHPK62 0 45628_9
QATVNAMPAKKTEESYPYIEANPEGIHKQHLELNKDSKFLALEQERRQLLEKDPRRNAREIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDDS VDPDHFRS 191 162 157 21|82|43|12.49 0 26659_6
PDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEQEDVGPRHVDPDHFRS HSSHTTRR 184 184 92 22|98|63.0|15.79 1 20148_5
RNRAQRIQRERVDNTKRGQQAQPHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSVSNDAATKFTLPSSPHWNTEPNPTQPQEKRRGL LEQKAAEN 170 49 49 36|70|55|6.69 0 43642_2
RLAEELEQKAAENEKLADELEQKAAENERLADELEQKAAENEKLADELEQKAAENERLAEELEQKAAENE ANGIRPQP 149 149 60 37|87|65.5|14.44 4 43425_3
RHGPQPGRQLHTSPQTHTQPAVTHQGRRGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEPHAEAIPHPSTTWRQNFQPFHPK62 0 45628_9
QATVNAMPAKKTEESYPYIEANPEGIHKQHLELNKDSKFLALEQERRQLLEKDPRRNAREIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDDS VDPDHFRS 191 162 157 21|82|43|12.49 0 26659_6
PDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEQEDVGPRHVDPDHFRS HSSHTTRR 184 184 92 22|98|63.0|15.79 1 20148_5
RNRAQRIQRERVDNTKRGQQAQPHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSVSNDAATKFTLPSSPHWNTEPNPTQPQEKRRGL LEQKAAEN 170 49 49 36|70|55|6.69 0 43642_2
RLAEELEQKAAENEKLADELEQKAAENERLADELEQKAAENEKLADELEQKAAENERLAEELEQKAAENE ANGIRPQP 149 149 60 37|87|65.5|14.44 4 43425_3
RHGPQPGRQLHTSPQTHTQPAVTHQGRRGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEPHAEAIPHPSTTWRQNFQPFHPK79 1 20148_5
RNRAQRIQRERVDNTKRGQQAQPHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSVSNDAATKFTLPSSPHWNTEPNPTQPQEKRRGL LEQKAAEN 170 49 49 36|70|55|6.69 0 43642_2
RLAEELEQKAAENEKLADELEQKAAENERLADELEQKAAENEKLADELEQKAAENERLAEELEQKAAENE ANGIRPQP 149 149 60 37|87|65.5|14.44 4 43425_3
RHGPQPGRQLHTSPQTHTQPAVTHQGRRGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEPHAEAIPHPSTTWRQNFQPFHPK79 1 20148_5
RNRAQRIQRERVDNTKRGQQAQPHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSVSNDAATKFTLPSSPHWNTEPNPTQPQEKRRGL LEQKAAEN 170 49 49 36|70|55|6.69 0 43642_2
RLAEELEQKAAENEKLADELEQKAAENERLADELEQKAAENEKLADELEQKAAENERLAEELEQKAAENE ANGIRPQP 149 149 60 37|87|65.5|14.44 4 43425_3
RHGPQPGRQLHTSPQTHTQPAVTHQGRRGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEPHAEAIPHPSTTWRQNFQPFHPK
TTGLRLID 146 146 103 25|63|40|7.04 0 10847_5 IRRFVRVMEGATDMPVACALRVTDGPRLVDDRLSTKVPGERCAPGVFETTGLRLIDDVGSDAV KRQSVNNY 132 132 98 22|55|25.0|4.78 1 1223_5
SPFEEHQSTGTKTTEDARTPDAAATEKRQSVNNYTTPPGDSDGSTAVSQTTSPLL RGQQAQPH 130 130 39 17|118|31|25.37 0 23470_5
GGTEEKECNERERDGAQRIQRERADKMRRGQQAQPHSSHTTRRHPPPSPPAHSCCPREMAKPRGHHTTHTEQAPTTRGRSGSRASSIHDAATKFPAIPPSMEHGAQLNAVTSGASHPQ QRQQPPTM 125 125 44 27|97| 60.0|17.44 0 44661_4
WSQLISDKEGKQWNERERDGTQRIQRERADNMRRGQQAQPHSSHTTRRHPRNEGTQRQQPPTMQEEKTSTTISPSSFPQPQGNGQNTAATTQRTWSK HRLQFNPT 116 116 32 27|102|46.0|13.24 0 8703_5
TQKEKTPASGKSPPRPSPWPATPHQPSHTLTPSQCRQSAGERIRPSTQQTHTPPAATRHGRGEETQRHHRLQFNPTANGIRPQPRSQYTAHTEQAITQRATR
PKSAEPKS 105 13 9 43|81|58|11.71 0 45139_6 AVPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAVPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSA RGHHTTHT 92 92 20 36|111|58.5|18.76 1 38362_6
VTITRSGVAGVSQLIGETERKRMEREGGGRSTADTARERRQHEERAAGPAARSTHQETASTQRRHTTPAATHHAGEDIHHHLPQLIPAVPGKWPKPRGHHTTHTEQANPHP MHRTPHTT 91 91 21 27|64|46|10.21 2 35563_3
MRRGQQPGMHRTPHTTHGTHSNPRQHAHAHTQPCRSTNTQMEEREANRSGTQGSTACTHGESGP EEVPLTGE 87 87 35 22|84| 28|10.67 0 17763_6
LLTAVLADTETPYFMRLSYTADNKWETISNGKAKSTTEGGTWEAGKEHQVALMLQGNKASVYIDGKSLGEEEVPLTGEKPLELF
RGQQAQPH 85 85 15 33|90|72|18.98 1 17875_4 EHSGYSESADNMRRGQQAQPHSPHTTRRHPRNEGTQRQQPPTMQEKKTSTTISPSSFPQPQGNGRTRGHHTTHTEQANPHPSITRQKNFL RHRLQFNP 85 85 10 31|69|48.0|10.93 7 17925_3
VIPRVRGFGPPHNKRTHRHRLQFNPTANGTRPQPRSQHTAHTEQAITQRDTRGSGSSPIHDATPRPMRS QPHSSHTT 82 82 10 29|90|65.0|18.42 4 4389_5
TRSRVAEFSELICGREEKEWNRRERNRAQRIQRERVDNTKRGQQAQPHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSV AAEATKVA 78 52 26 19|93| 36.0|19.8 0 40983_7
EEEKAKTFQRLITFESENINLKKRPNDAVSNRDKKKNSETAKTDEVEKQRAAEAAKAVETEKQRAAEATKVAEAEKRKAAEAAKAVETEKQRA RHRLQFNP 75 75 9 68|82|82|4.38 2 1656_3
THGRNKRQAGKAPSRPSPWPATPHTQTHTQPAVTHQGRKGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEPHAEAIP HRLQFNPT 74 74 8 26|85|44.5|16.16 3 8538_4
SSCPATPHQPSHTHHRSAVSPRVRVFGPPHNKRTHRQQSSTNGQEQEATHRHRLQFNPTANGIRPQPHSPHTKHSKQSPKEPHVN TKMPEERC 73 73 73 29|59|48|10.16 0 45377_4
VRAMEGATDMPVACTPRVTEGLRLVDGRFSTKMPEERCTPGVFETTSLRLIDDVGSDAV QPHSSHTT 72 72 9 20|96|65|25.41 0 4547_5
AQQIQRERAENMRRRQQAQPHSSHTTRRHPPPSPPAHSCCPREMAKPRGHHTTHTEQAPTTRGRSGSRASSIHDAATKFPAIPPSTGHGAQLNAVT QRQQPPTM 71 71 18 45|92|67.5|13.94 0 14951_3
TRGRGTEHRGYSERKSRQNEEKAAGPVAQFTHHGAASTQRKQRQQPPTMHEEKISTTISPSSFPQPTGNAKSPRPPHNAHGASTHHNSHTRQ MHRTPHTT 70 70 11 31|60|38|7.54 2 39388_3
RDIESNRGDERCSRHTQRAHHKRRGQQPGMHRTPHTTHNTHSNPRQQAHTHTALPQQQHT HRLQFNPT 66 66 8 30|82|51.0|16.14 3 6132_3
TLTGNSTPAITHKHHRSAVSPRPRRQPHNKRTHHQQSPTNGRGREATHRHRLQFNPTANGIRPRPAAGTQRTTSEQSPKEPH94 0 14951_3
TRGRGTEHRGYSERKSRQNEEKAAGPVAQFTHHGAASTQRKQRQQPPTMHEEKISTTISPSSFPQPTGNAKSPRPPHNAHGASTHHNSHTRQ MHRTPHTT 70 70 11 31|60|38|7.54 2 39388_3
RDIESNRGDERCSRHTQRAHHKRRGQQPGMHRTPHTTHNTHSNPRQQAHTHTALPQQQHT HRLQFNPT 66 66 8 30|82|51.0|16.14 3 6132_3
TLTGNSTPAITHKHHRSAVSPRPRRQPHNKRTHHQQSPTNGRGREATHRHRLQFNPTANGIRPRPAAGTQRTTSEQSPKEPH94 0 14951_3
TRGRGTEHRGYSERKSRQNEEKAAGPVAQFTHHGAASTQRKQRQQPPTMHEEKISTTISPSSFPQPTGNAKSPRPPHNAHGASTHHNSHTRQ MHRTPHTT 70 70 11 31|60|38|7.54 2 39388_3
RDIESNRGDERCSRHTQRAHHKRRGQQPGMHRTPHTTHNTHSNPRQQAHTHTALPQQQHT HRLQFNPT 66 66 8 30|82|51.0|16.14 3 6132_3
TLTGNSTPAITHKHHRSAVSPRPRRQPHNKRTHHQQSPTNGRGREATHRHRLQFNPTANGIRPRPAAGTQRTTSEQSPKEPH
DNMRRGQQ 65 65 16 40|46|44.5|1.94 9 34172_4 SVRREEKEWNERKRDGAQRIQRESADNMRRGQQAQKHSSHTTRLGM EDLQRQLE 62 42 32 34|70|54.0|7.09 0 29003_6
EHEHKIRGLQEVSEQAEDLQRQLEELRVENEELRAEGEDKTRGLQEVSEQAEDLQRQLEELRAENEELRG ARRLAEEA 61 15 6 32|72|53.5|13.43 0 7841_2
LVKEAEARRLAEEAEARRLAEEAEARRLAEEAEEARRLAEEAEEARRLAEEAEARRLAEEAEEARRLAEEAE RGHHTTHT 60 60 7 22|116| 66|28.46 0 36531_6
RSGVAGVSQLISDTEGKEWNERERDGAQRIQRERAENMRRRQQAQPHGSHTTRRQPPPSPPAHSHSPGKWPKPRGHHTTHTEQAPTTPATRASEPSSIHDAATKFTAIPPSMGHGA MHRTPHTT 59 59 9 21|79|36|19.81 0 26227_4
DLHSEGHEEQQGRREVRQTHTEGTHMRRGQQPGMHRTPHTTHTPIHGSRHTHTQPCRNTNKQMEEGEANRSGTQGSTAC HRLQFNPT 58 58 8 31|79|56.5|15.36 4 1942_5
LNPNQDTHTHNRSAVSPRVRGFGPPHKGTHRQQPHTNGREVTHRHRLQFNPTANGIRPQPRSQHTTHTEQTTTQRATRG KAHRWPLP 58 58 57 20|24|22|0.53 0 39008_3 QAVERKAHRWPLPRPSPPPTHPPD RAPEPQVK 57 57 29 35|86|53|14.73 0 45725_4
PLNSTERTAIKDRKPFPKRAPEPQVKIAPKPVAPAAPAAPGPREVPAALGRTTVGRTANTQHAPAGRLASAGNEGTAREKGDGGAN RGQQAQPH 56 56 7 52|77|64|9.79 8 9692_5
GVAGVSQLIGETEMKRMEREGGGRSTADTARESADNMRRGQQAQPHSPHTTRRHPRNEGTQRQQPPTMQEEKTSTTI SLEVERKK 55 55 55 25|99|67|11.62 0 42919_6
IPEALLHPPHNDVITPTKTDESLEVERKKRSRSPREHRDGGASYGASDDVYSDNGGDERSEEGFMELLNLEGFPSTSLYEQSYCHGHLLQDIGVVQDDI HHRPPTSH 55 55 51 19|56|38|9.36 0 21891_3
QRIFATEHSRVHRDTATDKLQHGAVHHHRPPTSHHRHEGHAEHNQHGPHKPQARQH KVAEAEKQ 54 25 24 30|88|42.5|14.08 0 40224_9
AEAMKVAEAEKQKAAEAAKAVETEKQRAAEATKVAEAEKQRAAEAMKVAEAEKQKAAEAMKVAEAEKQKAAEAAKAVETEKQRAAEAT QPHSSHTT 54 54 7 47|88|61|14.4 3 9353_4
VRQKKNNTTRGRGTEHSGYSERERADKTRRGQQAQPHSSHTTKRHPRNEGTQRQQPPTMQEEKTSTTISPSSSLQPQGNGRNPRPPQN RGHHTTHT 53 53 6 27|79|41.5|17.45 3 20897_3
EGRRPSRTVHTPRDGITQRRHTTPAVTHHAGGQHIHHHLLPAHRHSPREMAKPRGHHTTHTEQAITTPATRASGPSSIH45 3 20897_3 EGRRPSRTVHTPRDGITQRRHTTPAVTHHAGGQHIHHHLPAHRHSPREMAKPRGHHTTHTEQAITTPATRASGPSSIH
WRRDRCWE 53 53 30 31|79|59.5|10.88 0 42038_5 DGGPRGGVLFDGEPGRWVDSAGEGQRARRRLEGDALRRWRRDRCWEHAELCAQPSAAAADAVAEPVACGWGACEGGVQR TPPPTGLE 52 52 51 33|48|33|4.49 0 37476_5
AVAELEAEIHTNKKELLMRLYSDLSTPPPTGLEDKDAGGAGVEVDQML EEVPLTGE 52 52 35 22|61|38|8.8 0 32964_6 PKKEHRVTLMLQGNNASVDVDGESLGKEEVPLTGERPPEVLRLCFGACGGHESHVTVKNVF MHRTPHTT 50 50 8 23|69| 26.5|14.44 0
29830_3 GMESNRETRGAADTHGGHTHERRGQQPGMHRTPHTTHDTHSPAATPTHRWRWEKQTGAAHKAPPHAHTA HRLQFNPT 50 50 5 44|63|47|7.52 10 9055_3
KRTHRQQSSTNGQEQEATHRHRLQFNPTANGIRPQPRCQQTQRTPSEQSPKEPHAETIPYPSM GPARVRDA 50 50 50 34|58|47.5|6.27 0 16683_6 PQFGKRWEKTKPRGPARVRDAAATRDEIPQFPVSLTTKTREDAIPDSAHTEVLCFTPP YYTNPNRT 50 50 7
29|59|33|9.66 4 25134_5 IKYTNQNKHIIKINYKSYPYTFTQHLYYTNPNRTPPPRDNPLKSRNNVRGRCMNFPPKI PPPSPPAH 50 50 16 43|93|55.0|13.33 2 42399_3
EGSRPSRTVHTPRDGIHATKAHNASSHPPCRRRRHPPPSPPAHSRSPREMAKTPRPPHNAHGAGEPSSIHHAAQNFLPSRHQWDTEPNSAQSH QHSLPRRH 49 45 43 13|73|19|8.97 0 36907_3
QHSLPRRHPSPSAAQHSLPRHHPSPSAAQHSLPRRHPSPSAAQHSLPRRHLSPSAAQHSLPRRHPSPSAAQHS RRERPRRP 49 49 37 23|51|33|7.31 0 10819_3 EHDHHHQGTDSWRRERPRRPFVGSGAAAAACQCRCGDGRCCVLRTGRGACA LRELPLDD
48 47 31 20|99|50|21.82 0 40929_6 LPLDDDSDFVAMEQERRQLLEKDPRRNAREIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDDSDFVAMEQERRQLLEKDPRRNAKEI SRCQQTPP 47 47 45 25|85|48|16.84 0 41286_3
TCRPGSPSNKTSPRSPPSTRIRSAQAPQRSPTRSQPHAQTTRACTAGSRCQQTPPRSRRWPQRRPPPTRTAPAPSAPTAHRSTAH QPHSSHTT 47 47 7 34|64|52|11.79 12 38418_3
GRGTEHSGYSESADNMRRGQQAQPHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASE RGHHTTHT 46 46 6 36|76|50.0|13.35 4 37971_4
TTRREGSRPSRTVHTPRDGTRHHHHTQLISTTPGKWPKPRGHHTTHTAQANPQLVMTRPQNSHCHPVLTGTRSPTQ FGQAAAGD 46 19 16 23|79|49.0|16.86 0 47235_6
TLEKTKTEQKTAPFVQGAAGDKPSPFGQAAAGDKPSPFGQAAAGDKPSPFEQAAAGDRPSPFEQAAAGDKPSPFGQAAA GTQRQQPP 46 46 10 53|89|67.5|9.78 3 2982_3
GYSERADNMRRQQAQPHSSHTTRRHPRNEGTQRQQPPTMQEEKTSTTISPSSFPQPQGNGKTPRPPHNAHGAGEPSSIHHAAPRFPAIP HRLQFNPT 45 45 4 43|66|44.5|9.6 8 9189_3
RQSAGEKIQPSTQQYRTRNKQSSTNGQEQEATHRHRLQFNPTDNGIRPQPAANKHNAHQASNHPKS
YYTNPNRT 43 43 4 30|83|33.5|22.06 0 43062_5 LTTSKQKFYTHFFNYISVVIPTPYTPYLTNLITNFITNLNQHYKLSISSLYYTNPNRTPPPRDNPQFYRNNVRGRCMIFRPQI MHRTPHTT 42 42 5 26|56|47|10.35 1 24636_1
NHMRRKQQPGMHRTPHTTHDTHSNPRQQAHTHTQRWKREKQTGAAHKAPPHAHTAR HRLQFNPT 41 41 3 49|69|51|8.99 6 44584_4 RGFGPPHNKRTHRQQSSTNEREREETHRHHRLQFNPTANGIRPQPRSQHTAHTEQAITQRATRRSDSLS
TPTPIEPH 41 41 7 15|49|27|11.04 2 13518_3 LRSIPRHHHQNPHNYPHYQTNAYITPTPIEPHLPVTTPKIGQIMYGGDA PSHAEHST 41 41 38 29|103|51.0|14.63 0 23639_3
RPDRTTSAPSPSLCSSSLRRVLSSCSSPKFRLSRSSSPSRSDISLPSPEATHNNTHRAKRGKYHTSYVRPWPSHAEHSTTGAADRHFPAHRHHTHAQQITGTH SHTTRRHP 41 41 5 40|84|81|17.59 1 12705_4
IGEPEGKGMERKGEGRSTADTARKRADNMRRGQQAQPHSSHTTRRHPRNEGTQRQQPPTMQEKKTSTTISPSSFPQPQGNGKNT RGHHTTHT 40 40 5 29|108| 34|30.03 0 46783_5
EGKGMQQEGEGRSAADRARESRQHEKRAAVPVAKVTHHETTSTQGRNTTPAATHHAGGEDIHHHLPQLIPTANEKWQKHRGHHTTHTERAPTTPATRGNKISSHPALN
GFGQQAGG 40 24 18 24|64|43.5|10.75 0 42172_3 AAGRGTAGGFGQNTGTAGGGFGQTAATGGFGQQAGGFGQQAGGFGQQTGGFGQAAGRGTAGGFG NGIRPQPR 40 40 13 38|93|57|14.53 0 41181_4
QGRNKRQAGKAPPRASPWPATPHQPSHIHTPSQCCQSAGERIWSSTQQTHTLPAINHRGQGQEATHRHRLQLNPTDNGIRPQPRSQHTTHTKQ FPPNIDSP 39 39 32 28|100|65.0|18.42 0 21144_8
HSPAQPSESESGPVIFKQSSSDVIEPSTSAGVGMAEEESPGSGALAPASSQTQNAGSHELLGTEMPVSGEHFPPNIDSPLMGQVDTADEESPRIGNTDDQ YYTNPNRT 39 39 4 31|51| 36.5|8.07 4 20502_3
ILIDNHIIIIKPTVSLPLYYTNPNRTPPPRDNPQKCRNNVRGRCMNFQPQI HRLQFNPT 38 38 3 45|59|59|6.6 9 31858_3 VTHQGRRGRENHRHRLQFNPTANGIRPQPRSQHTTHTEQTTTQRATRGSDSSSTHDAAP AATEKRQS 38 38 31 20|71|39|13.02 0 1500_6
YPATTEGTAQSTSTGSQEQEAEPSTSEEPSPFEEHQSTGTKTTEDARTPDAAATEKRQSVNNYTTTPGDSD MHRTPHTT 37 37 5 30|46|34|5.46 7 18844_3 QRANTQEKRGQQPRMHRTPHTTHNTHSNPRQQAHTHSPAATPTTHR HTHTQPCR 37 37 29
25|63|40|9.14 1 17607_3 EGSSQACTVRRTPRTTHTPIHGSRHTHTQPCRNNNTQMEEGEANRSGTQGSTACTYRESGPSC SHTTRRHP 36 36 5 41|72|49|10.54 5 19427_4
GKRMQQEGERRSTADTARERADKMRRQQAQPHSSHTTRRHPPPSPPSSFLLPQGNGKTPRPPHNAHGASTHH RGHHTTHT 35 35 5 27|67|32|14.99 7 27487_4
HHHLPQLIPAAPGKWPKPRGHHTTHTEQAPTTRGRSGSRASSIHDAAQNFLPSRPQWNTEPNSTQSH THRHRLQF 35 35 5 30|65|35|12.53 4 5643_3 VSPRVRGFGPPHNKRTHRQQSSTNGQGREATHRHRLQFNPTANGIRPQPAANKHNANQASNHLKS
EARRLAEE 35 9 5 40|64|43|9.11 1 34388_2 ARRLAEEAEEARRLAEEAAEARRLAEEAEARRLAEEAEARRLAEEAEEARRLAEEAAEARRLAE RWFNLTGR 35 35 35 30|49|39|4.49 0 37626_5
LRRGSSSSRWFNLTGRGSERPSCGSGIPLRRCEGPSGRSGVVSGGSESI APYAAHHA 35 35 11 32|65|49|8.71 1 27097_3 TGETRGAADTHTERTSHEERAAARHAPYAAHHAQHTLQSTAEGTHTHTHTHSPAATTTHRWKREK YYTNPNRT 35 35 3
37|69|49|13.2 0 19414_5 NLSTSNNSTNTSFIHLNKSKLYTSRYNINKLPLNTIRYCTLYYTNPNRTPPPRDNPHFHQNNVRGRCMK HRLQFNPT 34 34 3 37|54|52|7.59 12 6493_3 DEKIRPSTQRRHTLPADIHRGEEQEATQRHRLQFNPTANGIRPQPAANKHNAHQ
TPTPIEPH 34 34 5 18|37|20|7.07 6 712_2 TNSYLHHSALYITPTPIEPHLPVKIAKSVEIMYGGDA PPPSPPAH 34 34 4 35|72|72.0|16.02 6 37349_3 RRTTHPPPSPPAHLYSPREMAKTPRSPHQPHGASTHHPSTRGNKISCHPVLNGTRGPTQRSHTRCTTPTGKE
RGQQAQPH 34 34 5 38|61|55|9.87 8 6291_5 QQEGEGWGTADTVSESADNRRGQQAQPHTSHTTRRHPRKDGTQRQQPPTMQEEKISTTISS DSSAHSTP 34 12 10 43|73|56.0|9.72 0 47656_8
DSSAHSTPSTPVDSSAHSTPSTPADSSAHSTPSTPADSSAHGTPSTPADSSAHSTPSTPVDSSAHSTPSTPAD MHRTPHTT 32 32 5 23|41|27|6.22 10 32579_4 TGETRGAADTQRAHHMRRRQQPAMHRTPHTTHDTHSNPRQH YYTNPNRT 32 32 3 25|48|
35|9.42 4 22265_5 NYISNNHNILYTAPLYYTNPNRTPPPRDNPLKCQNNVRGRCMKFRPKI SKQSPKEP 32 32 5 50|96|68|15.55 0 25931_4
QRCQSAGERIWSSRQQTNTSPVITHRGRGGTTHRHCLQFNPTTNGIQPQPLTQRTTSKQSPKEPHAKAIPHPSMTRPQHFQSFHPQWNLMPDSKRS NMRRGQQA 32 32 13 16|81|26|23.45 2 13411_4
KRERDGAPRIQRESADNMRRGQQAQPHSPHTTRRRPRKDGTQRQQPPTMQEETTSTTISPSSFPRPPHQPHNEQADPHPPT HHSLRPHG 31 31 19 32|62|48|7.04 0 12044_4
AHPLLQGLPGGACCQVHPVTQRREEQGKPVGDAAGLRRRDEGRHHSLRPHGRGPGGRAEVGQ HVELQRRN 31 31 11 24|58|30|10.31 0 34881_3 INGRPAANTIPKEIVQPHVGAAHWSTAATLRSPPHVELQRRNHHKSHGEVPSQNPCGE
QQQQQQQQ 31 13 4 33|46|41.5|5.68 5 3442_4 VFHRFRMHGDSEMRDGVRQRRQPQQQQQQQQHGRCGFEEDVEEEGG RGHHTTHT 30 30 4 51|65|55.5|5.21 8 28141_1
PPSPPAHSRSPREMPKPRGHHTTHTEQANPHPSITRHKITCHPALNGTRSPTQRSQMRCTTPAGK PRPPHNTH 30 30 11 46|55|50|3.32 5 27213_3 TRREGSRPSSAVHTPRDGTRHHHHTPLISTTPGKWPKAPRPPHNTHGASEPSVGN HRLQFNPT 30 30 3
38|49| 38|5.19 13 6653_4 YTPPAVTHNAREEVTHRHRLQFNPTANTQRKWSKQSPKEPHEEADPHPS LTVRHGAH 30 30 8 39|71|48.5|11.86 0 29949_4
ALSTHTAATKETGGSNSTRQSFTSLRTHGLAERLTVRHGAHHPLSSECTQQASAQSPSCHRNGNKQSSSQT TAHNRVHP 30 30 30 19|80|34.0|11.87 0 36679_3
MAARQRTPRESIAPPPSSFPGSPARTATAHNRVHPHPSPRRPQRPNSQPPKNRCGNHTPATKHATQHRVTWTPPSEECHV TPTPIEPH 29 29 4 21|47|29.5|9.47 4 39868_4 IHHDILKHTVNILNYNTHYAYITPTPIEPHLPVTTPDFGEIMYGGDA YYTNPNRT 29 29 3
```

```
33|39|35|2.49 8 26952_4 SQTYDTLNKKLKSQIQRNNYSLYYTNPNRTPPPRDNPHF PAAGGFGS 29 14 14 26|76|37.5|14.64 0 36001_3 HTSAPAAGGFGSAAHTSTPAVGGFGSATTTSAPAAGGFGSAAHTSAPAAGGFGSAAHTSAPAVGGFGSATTTSAPA
PPPSPPAH 29 29 3 89|97|93|3.27 0 35758_5 TTGQHPRNANNASSHPPCTRRRYPPPSPPAHSHSQREMPKARGHHTTHTERAPTITATRASGTSSIHDAAKKLPAIPPSMGHGAQLNAVTRGAPHPQ RAPEPQVK 28 28 17 33|86|57|13.82 0 18002_7
SQEESSQAASPVKPSPEEIGKKSQVTVKNVFLYNRPLNSTERTAIKDRKPVPKRAPEPQVKIAPKPVAPAAPAAPGPREVPAALGR PQATQQRG 28 28 28 34|88|63.0|12.76 0 11556_4
PRKERTETRRSQTSPPTASQARASLWLQSSQKETHTRRTTPSTSPERSLPQATQQRGGSCGSLTRPAPRQSPRTCREQTRMGSSADTA TANGIRPQ 28 28 9 29|88|43|19.73 2 23768_4
ALTGNSTPALKHTHHRSAASPRVRGFGPPHNKRTHRQQSSTNGQGREATHRHRHLHFNPTANGIRPQLAAGTQRTRSEQSPKEPHAEAD82 0 18002_7
SQEESSQAASPVKPSPEEIGKKSQVTVKNVFLYNRPLNSTERTAIKDRKPVPKRAPEPQVKIAPKPVAPAAPAAPGPREVPAALGR PQATQQRG 28 28 28 34|88|63.0|12.76 0 11556_4
PRKERTETRRSQTSPPTASQARASLWLQSSQKETHTRRTTPSTSPERSLPQATQQRGGSCGSLTRPAPRQSPRTCREQTRMGSSADTA TANGIRPQ 28 28 9 29|88|43|19.73 2 23768_4
ALTGNSTPALKHTHHRSAASPRVRGFGPPHNKRTHRQQSSTNGQGREATHRHRLHFNPTANGIRPQLAAGTQRTRSEQSPKEPHAEAD82 0 18002_7
SQEESSQAASPVKPSPEEIGKKSQVTVKNVFLYNRPLNSTERTAIKDRKPVPKRAPEPQVKIAPKPVAPAAPAAPGPREVPAALGR PQATQQRG 28 28 28 34|88|63.0|12.76 0 11556_4
PRKERTETRRSQTSPPTASQARASLWLQSSQKETHTRRTTPSTSPERSLPQATQQRGGSCGSLTRPAPRQSPRTCREQTRMGSSADTA TANGIRPQ 28 28 9 29|88|43|19.73 2 23768_4
ALTGNSTPALKHTHHRSAASPRVRGFGPPHNKRTHRQQSSTNGQGREATHRHRLHFNPTANGIRPQLAAGTQRTRSEQSPKEPHAEAD
HHLPQLIP 28 28 11 36|95|54|16.88 0 44503_4 NEKNGTRGRGTEHSGYSERERRQHEERAAGPAAQSTHHETASTQRRHTTPAATHHAGGEDIHHHLPQLIPAAPGKWQKPRGHHTTYTEQANPHPS MHRTPHTT 27 27 4 46|53|50.0|2.86 2 24828_3
EGHGEQQGRREVRQTHTQRANHAGRGQQPGMHRTPHTTQGSLKAKQNNEKISQ VDAAGEGQ 27 27 25 37|72|55|7.42 0 29473_5 ERLFVDGGLRGGVPFDGEPGRWVDAAGEGQRARWRLEGDDLRRWRRDRWRRDRCWEHAELCAQPSAAAADAV
YGPLRPTG 27 27 26 44|91| 48.0|10.83 0 10980_6 FSKSHIVLRSTTGGDAAAGTPQEVSGSVTSTEPTDGPMEPDYGPLRPTGMWNVEEVVDVKNSTVDFRRIDDVESEVIEALSQPDDAVVPYE HRLQFNPT 27 27 2 69|88|78.5|9.5 2 28949_5
EETNDRREKHRQRPHPDRQLHTCPHTHKHTPSRCRQSPGERICSSTQQTHTPPAVTHNARGEETHSHHRLQFNPTANGIRPQPAANNI QREAEERA 27 11 8 25|65|54.5|12.22 0 30426_2
EERAQREAEERAQREAEKRAQREAEKRAQREAEERAQREAEKRAQREAEERAQREAEERAQREAE KGQTSLED 27 27 27 44|99|68|13.73 0 12764_5
TEGDKPPVNRLYGRPSSLTVSPKKSDATKKGQTSLEDRHSPGTRASRSENRVPTTKRPTATRNTTTKEGKSRPGGDSLPSGINADFAERIESEIIERSP
IRNEGSVS 27 27 26 48|131|87.5|29.07 0 2061_7ERQNNGTRNDGYGNNASNRRDRVEWNDDRRGDRTDMRTLIRNEGSVSRNNRPERAERNNGRYDERNERNGERRWERHDGMRGERNDNVRARDERGVLEEGPAPQSTRPLRSREQPSTAASKSENGGSKPSI
QQQQQQQQ 27 9 3 57|61|58|1.7 2 28495_3 NIKQQQQQQQQQQPLQQKKKRKGLRSRLRLPKRKWGRNKHPHTHTHAKHTHTHTQKEREGE HRTPHTTH 26 26 7 27|56|37|9.2 0 13638_4
DMESNRGDERCGRHTQRAHRMRRGQQPGMHRTPHTTHGTHSNPRQHAHSPAATPTH GTQRQQPP 26 26 4 55|91|62.5|14.31 0 29199_4
QAQPHSSHTTRRHLRKEGTQRQQPPTMQEKKTSTTIFPSSFPQPQGSGKTPRPPHNAHGAGEPSSMTRHQNFLPSRTQWETEPNSAQSHAV YYTNPNRT 26 26 3 21|34 |27|5.31 11 8843_3 SLISTQLLYYTNPNRTPPPRKFRQIDKIMYGGDA SKQSPKEP 26 26
4 47|65|56.0|6.71 5 11612_3 HNKRTHCQQSTTVDKDKKRHTATASSSIPRTTGYGHSPAVNTQRTPSKQSPKEPHAEAIPHPSMT PPPSPPAH 26 26 3 72|94|76|9.57 0 9491_3
DTAREREQTTRGEGSRPSHTVHTPRNGIHERKEHNASSHPPCRRRRHPPPSPPAHSHSQREMAENRGHHTTHTERAPTTPVTRASEPSSIHHAA RGQQAQPH 26 26 4 41|91|68.0|23.57 0 3868_4
EEKECNKRERDGAQRIQRERADKTRRGQQAQPHSLHTTRRHPRNEGTQRQQPPTMQKEQTSTTISQLIPQPTGTGHSPRSPHNAHGASEPS TPTPIEPH 25 25 3 18|52|22|15.17 1 43964_3
TQSNLIYTSQHHSPSNITQSNINTVTINHNLCYITPTPIEPHLPENSANSTK THRHRLQF 25 25 4 54|86|74.5|13.01 1 42220_5 KRHPHSKHTTPVVTHQQGEEVTHRHRLQFSPTDNGIRPQPRSQHTTQTEQAISQRATRRSDSSSIRDEAPHSMRSHGAHHFLNRTH
HRLQFNPT 25 25 2 72|73|72.5|0.5 3 4982_3 TQQTHTPPAATRHGRGEETQRHHRLQFNPTANGIRPQPRSQHTAHTEQAITQRATHAEAIPHPSTTQRPTPSG THTHSPAA 25 25 7 28|74|47|14.7 0 33225_5
DLHSEGHGEQQGRREVRQTHTEGTPHEERVAARHAPYAAHHAQHTLHSTAAGTHTHSPAATPTHTWKREKQTGA VQQARRTG 25 25 19 12|58|16|12.61 0 42724_1 LRVSASPRRKENKTRGKEMCAPPHRPSSHAHTSRLPQRAVRRGTSSCVQQARRTGPSS
TTTTTTTT 25 12 3 41|54|49|5.35 1 27931_3 TTKPPTTTTTTTTKPPTTTTTTTTTAPEAPSITTTETPNTTTTRAPSSIRRIDG PKSAEPKS 24 4 4 48|77|61.5|11.9 0 47942_5
PFTEKKPAKASTATSPSVEHVTTPVATEPKSAEPKSAEPKSAVPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSA FHPSTPLP 24 24 17 27|73|42|12.14 0 10269_5
EASATRDVANDGVSTLPHQTDILMAGDVSSSAARRSSAAHRPPHGNAARPRFHPSTPLPSRTSPLDARLPMEE ERREARER 24 14 14 35|47|42.0|3.36 0 40736_4 REERRGVEERREARERARHEAKERREARERARLQQKLELTVTTTIKD GGRKCPTL 24 24 24
21|81|51.5|15.94 0 36331_5 KQRKKGIAESNSLPDVRRCEGGRKCPTLRPHTPHAFARPHSPSRSSSGSSSLPYAEPASTFAPPSPASRAVPSFPEEVKLP WRWAWHSH 24 24 13 17|46|29|9.16 0 33917_4
SQWRWAWHSHGGAVGAPCSRTATPVGEAAEAHRPRSNATCDRQKHR SHTTRRHP 24 24 4 52|83|70.5|11.51 1 25333_2 ADNMRRGQQAQKHSSHTTRRHPRKEGTQRQQPPCRRRRHPPPSPPAHSRSPREMAKTPRPPHNTHGASKPSSIHHAAQKFPAI
QQITGTHV 23 23 23 31|63|40|7.81 0 42302_4 SYVRPWPSHAEHSTTGAADRHFPAHRHHKHAQQITGTHVRTPNKGGQRQRQTEQRTRTKPTTS YYTNPNRT 23 23 2 36|61|48.5|12.5 1 47345_4
RLKVYHSIHHYIQPYQNHNHKLISYTILYYTNPNRTPPPRDNPHFQPNNVRGRCMNFPPQI HRLQFNPT 23 23 2 30|59|44.5|14.5 4 2506_3 VTHQGRRGRENHRHRLQFNPTANGIQPQPRSQHTKHSKQSPKEPHAEAIPHSTWR
TPTPIEPH 22 223 25|48|31|9.74 1 19664_3 KSINKQNHTKDFITLHPPTTAYITPTPIEPHLPVKSANFQKIMYGGDA SPPESFTA 22 22 15 22|22|22|0.0 1 14447_3 FLFTALRRIDASPPESFTAAPR LTVRHGAH 22 225 24|61|43|13.020 43962_4
TKETDTNSRRRALLLSSPADWRPTLTVRHGAHHPLPSECTQQAGAQSPSCHRKGKNKQSSS HESFLSLE 22 22 15 42|102|85|16.48 0 21607_7
NQFGQPYMSSNSKTADVSSIFSTEHGKGDPHKEATPTLTEIQPPQQLRQKQRLHESFLSLEKTGTEVERQLWSSPRRREKTEPRRPQAEGSPVFELPPSEVA RWHTLRSE 22 22 22 23|57|40.5|8.17 0 12570_3
LDRAIAAAPDAEDGRRLDLRRREALREASKSRWHTLRSELEGTDGNRQHVVRPCLSP NPRQQAHT 22 22 6 30|44|35.0|5.25 6 36711_3 PGMHRTPHTTHDTHSNPRQQAHTHTHSPAATPTHRWRWEKQTGA
QALGRHSH 21 21 17 26|41|32|3.79 1 33861_3 VSTTLPPRHAQGRSQALGRHSHGGRADGYRQQGPLSTVTDA VYEKPHPQ 21 21 21 40|99|65|12.120 19240_7
KGESGNKIGPAAHATVATHSDNKREHKTTKAVYEKPHPQKKGPEKPGKSKNKNKDAGKSAENLNESLPLPLPSPRSNGVVEEASVAPVNVTNTPIVEDL YYTNPNRT 21 21 2 32|46|39.0|7.0 4 28622_3
YNTKINKYHKPTVYYTNPNRTPPPRKIHGFGRNNVRGRCMNFRPQI THRHRLQF 21 213 25|72|66| 20.892 24074_4 TPSHSAHTAKRHPHSKHTTSVVTHQQGEVTHRHRLQFSPTDNGIRPQPRSQHTTHTEQSPKEPHAEAIPHPS SKQSPKEP 21 214
45|62|54.0|7.795 44733_3 IASSSIPRTTGYGHSPAANTKHSKQSPKEPLAEAIPHPSMTRRQNFQPFQPKWNLMPDSKRS PPPSPPAH 21 213 52|77|63|10.23 1 21141_4
ADRWDRRKRMQREGERRSAADTARERADKMRRGQQAQPHSSHTTRRHPPPSPPAHSCCPREMAKPRGHHTTHTEQAP ELQRRNHH 21 21 12 18|56|36.0|9.750 34114_3 EEITQPHVGAAHWSTAATLYSPPHVELQRRNHHKSHGEVPSQHPCGEESHSHRPTA
SAIGPAPE 21 21 21 32|68|56 |8.570 29131_4 ATAEFAARAPNSAIGPAPEERSRRMEASGRRHPAGAGKSEKREEEAPPRGGDCGEPLDVTANLSPLLT TRAPSRLR 21 215 27|38|35|3.93 48354_3 AKTTTTTTTAPEAPSTTTTEASTTTTTRAPSRLREIDG
PGRQLHTS 20 209 19|82|35|16.750 46381_3 KDRQTGGEMNGRRQDALTINMTREHAATPLRNTGREGEKSKKQHNTHGRNKLREKHRHSPRPGRQLHTSPHTHIIAVPSVPG APYAAHHA 20 208 45|61|49.5|6.340 45153_3
HEERAAATHAPYAAHHAQHTLQSTAAGTHTHSPAATPTRRWNGEKQTGAAHKAPPHAHTAR PTASTHRE 20 20 10 33|48|39.5|4.32 23265_4 TRTQRMHPQRLPTASTHREVRALEWEGRSRQRTQKEEKHAEPHPSTNS HRLQFNPT 20 20 2 50|57|53.5|3.5 5
47004_3 QQLWRTACVNCETAPTQQTHTPPAATHNGQEATHRHRLQFNPTANGIRPQPAAGTQR AVHHHRPS 20 20 11 27|70|40|11.620 39947_4 HLAPRHAQHSALQRNHARTRVKRNVATEHSRVQRDTATDKLQHSAVHHHRPSTSHHRHEGNAGHNQHGPH
KTFRGTRW 20 20 19 22|83|48|16.230 16114_3 GRSEHPHVLKSHRRSQPRSLAASGMPPSRDGHKDVASNKTFRGTRWANDDRHCGMCGPSAAHLKSHFYASSSSSSSSINTPVH CSRVSRPP 19 196 24|63|40.5 |15.620 10118_4
PSNARTRQSKNKREEKNDPKNRQHLLAWRASCSRVSRPPAQCKRRPTHPSNGTVIPTAGTHAA LPVPSTGE 19 19 12 31|54|42.5|7.03 1 18188_3 NGAIEGHAGESDVHAKDSSSVSSAESRPTGKPSLPVPSTGESPPTTAGDTRNST VKAERKGK 19 19 11
37|65|53|7.31 0 11676_3 FGPVKAERKGKDAAAPARAEKKPKAAAAAAGGAEEEDEAPREKKKPNPLDELPPSPFVLDAFKRE TPTPIEPH 19 193 24|34|27|4.193 8060_2 KNSKMYRPSYITPTPIEPHLPEKSLFLGQISSWF
AQHTHGRN 19 19 18 29|79|47.5|13.28 0 25067_6 QPAQVPRTHTHTHTDPLSIHSTTHKGEDISRKIPPHRNEGPWAQHTHGRNKRDGTHPSTRRSPLLLASPTDRHHASQQAVA RHHPHSWG 19 19 19 17|48|24|8.36 0 11424_3
ILFVPDSSGTKLAAAVEPATGTHAHRHHPHSWGGGRTGTPSHQGYAQI NKSHEHTR 19 19 11 33|84|51|15.32 0 46010_6 DSRVVNKQGSRVTGVVLPTGHTSGGASVTAHTLPTPGATPPALTAVAPWDAPFTNKSHEHTRIYPNRKHPPRGEGSGVGEPKQA
PAHFHNPR 19 199 33|83|58 |12.620 39812_5 RNRAQRIQRERVDNTKRGQQAQPHGSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSVSNDAATKFTLPSSPHWN YYTNPNRT 19 192 30|32|31.0|1.09 45840_3
TLPPRLYYTNPNRTPPPRDNPQNPKIMYGGDA EEGEANRS 19 195 38|57|40|7.14 1 33209_2 RTTHTPIHGSRHTHTHTQMEEGEANRSGTQGSTACTYGESGPSCLHPSRRARTHTRE YQHPPRHR 19 19 16 21|47|35.5|6.630 17055_3
VQGEAPATRLYQHPPRHRPPAPPSPPRHTSDARRPSGEDAPAVARQR RGHHTTHT 18 183 46|72|70 |11.812 48259_2 ARESRQHEERAAGPAAQFTHHETAPATTITPRSFPQLQGNGRKPRGHHTTHTAQANPQLVMARPQNSHCHPV FLDQKPEG 18 18 17
22|72|65|10.890 23284_5 AQELAREKKLADRAFLDQKPEGVPLRELPLDDDSDFVAMEQERRQLLERDPRRNAREIAALEESMNARAQEL HRLQFNPT 18 18 2 26|27|26.5|0.5 15 8767_3 HTNARGGETHRRHRLQFNPTTNGTRLQ PPPSPPAH 18 183
32|64|45|13.143 7685_3 HATKAHNASSHPPCRRRRHPPPSPPAHRYSPREMAKTPRPPYQPHAPANPHPLMTRHKIPCHPA
TTTTTTTT 1892 37|44|40.5|3.5 1 26042_3 TTTTTTKPPTTTTTTTTQTPSTTTTEVPTVSTTRAPSRLREIDG QQQQQQQQ 1863 44|51|44|3.33 20956_4 VAALARSQSGSGSGRYQQREQQQQQQQQQRFGGPRDGRPLRGRPPQNFQQG TPAATHHA 18 186
55|86|72.0|11.60 21435_4 RQHEERAAGPAAQFTHHETASTQRRHTTPAATHHAGEEDIHHHLPQLIPAAPGKWQKHRGHHTAHGASEPSSIHDAAPKFPAIPPS ASPSGRPS 17 17 10 29|42|39.0|4.68 1 24371_1
RTHSHRCRSAQDGWPRSPRTAPQGAQSSQCASPSGRPSASCS HTTHTEQA 17 17 6 68|82|75.0|4.04 0 42143_3 ESKKKQNTHEGNKRQAGKAPPRPHTAGNSTPALSHTHSRHRLQFSPTGNGVRPQPRSQHTTHTEQAITQRATHGGDSSSTHD DEAWRRIQ
17 17 17 26|38|26|2.820 10549_3 AEAMKRFATQKEKSEKFIQENLDRQDEAWRRIQELERV LTVRHGAH 17 175 26|44|29|6.692 2951_4 HSRSWQWRPTLTVRHGAHHPLPSERTHQGAQSPSHHRNGETQTI RHPRATRW 17 17 9 25|71|40|12.18 0 27447_4
KHKTAEAATTAKLPPCCADATARKGAPATESIAQETEKSPTRHPRATRWQSCRGCLNATTTNSAATTEPCT SKQSPKEP 17 17 4 40|58|49.5|6.46 5 5141_3 TTSVVTHQQGEEVTHRHRLQLSPTDNGIRPQPRSQQTKRRPSKQSPKEPHAEAIPHPS EEVPLTGE 17
17 12 20|39|32.5|5.572 44900_4 SLGEEEVPLTGEAPLGLVHFCFGACGEDAGQKTKVTVKN QPHSSHTT 17 17 4 20|88|43.5|29.69 1 9826_4
NMRRGQQAQPHSSHTTRRHPRNEGTQRQQPPTMQEKKTSTTISPSSFPQPQGNGKNTAATTPHTEQANPHPSMTRHQNFLPSRPQWDT HAMRPLTK 17 17 16 25|43|29.0|4.64 0 2955_5 NLYFPHRTAVPHAMRPLTKKPRLHDPSSNEASQSGTWNSLPPS
MHRTPHTT 17 17 4 41 |52|50.5|4.5 1 9201_3 HMRRGQQPGMHRTPHTTHNTHSNPRQHTHTHSPAATPTTHMEEGEANRSGTQ YYTNPNRT 17 172 20|30|25.0|5.0 10 24336_1 GLVTDPSPILYYTNPNRTPPPRKIAVFGPN TPTPIEPH 16 163 25|30|27|2.053
2887_3 ISARPYITPTPIEPHLPVTTPNFTEIMYGG LQHYHRGG 16 16 16 37|98|68.0|16.290 40585_3
NRGYSGGFQHDRRHFDDYRDGPGYSGNGDGVPHLQHYHRGGGSGGGRSRGGGGSDVPQPTSHRSNENHQEKRRWTQNAKDTAGATTDGEKEQQRKPGN AGWRRQGS 16 16 10 29|45|41.0|4.25 1 38469_5
TAREKERYAENAMPLHNAAGWRRQGSESLYPTLRRGQRSDDPVCS THRHRLQF 16 162 74|92|83.0|9.00 30246_3 HAESGTQENPEGTARSTTAHTSPHTQTHAHKSQCCQTARGRIRPSTQKTRAPPAFTHNGRGKETHRHRLQFSLVANGIRPRPRSQHTTLTVP
HRLQFNPT 16 16 1 92|92|92|0.0 0 23940_5 KNTSLGKSTATALTLAGNSTPALTHTLTPSQCCQSAGERIRPSTQQTHTPPAATRHGRGEETQRHHRLQFNPTANGIRPQPRSQHTAHTEQA PPHNAHGA 16 163 24|49|45|10.968 21813_1
RRTTHPPPSPPAHLYSPREMAKTPRPPHNAHGASTHHTSHTRQRTLIHP SSDVIEPF 16 165 51|76|55 |8.91 1 45031_7 SPRIQHSPAQTSESESGPVISKQSSSDVIEPFTSADVGMAKEESPGSGALAPASSQTQNAGSHELLGTEMPVSGDH HDTHRCRP 16 16 16
19|30|26.0|3.060 46108_3 TPMQARWCSILHDTHRCRPAWLHGRSTVTV SSEGRCSR 16 16 11 26|40|31|4.720 43089_3 TWTSRQPSLAPASRSSAATSSEGRCSRWSVCSRTPRWTSV RGHHTTHT 15 15 3 45|71|64|10.98 2 3487_4
EGSRPSRTAHTPRDDIHHHHPPLISTTPGKWPKPRGHHTTHTEQAPTTRGRSGSRASSIHDAAQNFQPSRP RRSAQGGW 15 15 15 12 |43|16|12.85 0 1322_1 PRTPSHHRRSAQGGWPRWSRTAPQAAQSSPCASPSGRPSASCS GSQGVRRG 15 15 15
46|80|62|9.96 0 43818_3 RSQSSLERNNTPAQAPRRTQQEFRSGSQGVRRGASHEQRGPQSSTATPSLPGRREVAQGLPHAGFRGQLTRPKNPYAAIA HRHRLQFN 15 15 3 58|85|81|11.9 0 2292_4
LNLTGNSTPALTHTPSQCRQSAGERIRPSTQQTHTQPAVTHQGRRGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEPHAE EARRLAEE 1544 47|72|62.0|9.360 35245_3
IRKTTNSKTNSTNTKEEANILVKEAEARRLAEEAEARRLAEEAEARRLAEEAEEARRLAEEAEEARRLAEEA ECWHAQPA 15 156 36|51| 41.5|5.27 1 16897_4 LLYTSDHSRRTHVPLTECWHAQPAHTSSAQSGIRQRGCGEQQRGTHPRRDA TFGPSHAK 15 15 12
27|44|35.0|5.450 40702_3 TRTRPHHTPVKTFGPSHAKDWSGRTAGKTTPAAMSSPHTPTHRW HLPHHHPF 15 15 10 37|59|45.5|6.930 41381_4 TLLTQSQDPTVPAEAAAAAAPSRRRKQHHHLPHHHPFLVPSLHSQQKSSVPLAHVGYWR THSNPRQQ 15 155
30|56|44|8.330 26490_3 QRANTQEKRGQQPGMHRTPQTTHNTHSNPRQQAHTHTALPQQQQHTDGGGRSKQER DEYSPEKA 15 15 15 38| 75|49|10.740 14504_5
RTPDHPSDRHEGHHQRHPQPKHREAPFQWRHQTPHETTVDEYSPEKATLKQPQTLLLPPLARHPVHPPTHSERQD TTTTTTTT 1572 35|36|35.5|0.52 1084_3 IQRRGWQLDRGRYHPSPPPPPPPPTTTTTTTTHSHD YYTNPNRT 15 15 1 66|66|66|0.00 37653_6
QTHHHFNQLSIKYHNNNLYYTTKLNHNDINYNNHIHHPYPKMLYYTNPNRTPPPRKNPKIPEIMYG HRQRRPSW 14 146 21|68|39.5|17.530 17755_3 SYSSFGASSSSSAAAVTALHHAPAAITLTEREECVGVCRQHSDRHRQRRPSWHRHPFSTERQPTSLVT
TTPGDSDG 14 14 14 21|25|25.0|1.030 23583_3 AATEKRQNVNNYTTPGDSDGSTAVS TQRQQPPT 14 143 37|70|57|13.574 34008_3 HSPHTTRRHPRNEGTQRQQPPTMQEEKASTTISPSSSLQPQGNGQTPRPPQQPHAPANPRPSITRQQNSL RPASAPCG
14 14 9 35|52|49|6.08 1 10880_5 PSPSSPSCSPVQRTRPASAPCGRRRPQSVTSAHPSRGAAVSATPASSRPHRH THRHRLQF 14 142 30|54|42.0|12.03 20835_3 STQQAHTPPAATHRGQGQEATHRHRLQFNPTANGIRPQPRCQQTQRTPSEQSPK52 1084_3
IQRRGWQLDRGRYHPSPPPPPPPPTTTTTTTTHSHD YYTNPNRT 15 15 1 66|66|66|0.00 37653_6 QTHHHFNQLSIKYHNNNLYYTTKLNHNDINYNNHIHHPYPKMLYYTNPNRTPPPRKNPKIPEIMYG HRQRRPSW 14 146 21|68|39.5|17.530 17755_3
SYSSFGASSSSSAAAVTALHHAPAAITLTEREECVGVCRQHSDRHRQRRPSWHRHPFSTERQPTSLVT TTPGDSDG 14 14 14 21|25|25.0|1.030 23583_3 AATEKRQNVNNYTTPGDSDGSTAVS TQRQQPPT 14 143 37|70|57| 13.574 34008_3
HSPHTTRRHPRNEGTQRQQPPTMQEEKASTTISPSSSLQPQGNGQTPRPPQQPHAPANPRPSITRQQNSL RPASAPCG 14 14 9 35|52|49|6.08 1 10880_5 PSPSSPSCSPVQRTRPASAPCGRRRPQSVTSAHPSRGAAVSATPASSRPHRH THRHRLQF 14 142
30|54|42.0|12.03 20835_3 STQQAHTPPAATHRGQGQEATHRHRLQFNPTANGIRPQPRCQQTQRTPSEQSPK52 1084_3 IQRRGWQLDRGRYHPSPPPPPPPPTTTTTTTTHSHD YYTNPNRT 15 15 1 66|66|66|0.00 37653_6
QTHHHFNQLSIKYHNNNLYYTTKLNHNDINYNNHIHHPYPKMLYYTNPNRTPPPRKNPKIPEIMYG HRQRRPSW 14 146 21|68|39.5|17.530 17755_3 SYSSFGASSSSSAAAVTALHHAPAAITLTEREECVGVCRQHSDRHRQRRPSWHRHPFSTERQPTSLVT
TTPGDSDG 14 14 14 21|25|25.0|1.030 23583_3 AATEKRQNVNNYTTPGDSDGSTAVS TQRQQPPT 14 143 37|70|57| 13.574 34008_3 HSPHTTRRHPRNEGTQRQQPPTMQEEKASTTISPSSSLQPQGNGQTPRPPQQPHAPANPRPSITRQQNSL RPASAPCG
14 14 9 35|52|49|6.08 1 10880_5 PSPSSPSCSPVQRTRPASAPCGRRRPQSVTSAHPSRGAAVSATPASSRPHRH THRHRLQF 14 142 30|54|42.0|12.03 20835_3 STQQAHTPPAATHRGQGQEATHRHRLQFNPTANGIRPQPRCQQTQRTPSEQSPK530
17755_3 SYSSFGASSSSSAAAVTALHHAPAAITLTEREECVGVCRQHSDRHRQRRPSWHRHPFSTERQPTSLVT TTPGDSDG 14 14 14 21|25|25.0|1.030 23583_3 AATEKRQNVNNYTTPGDSDGSTAVS TQRQQPPT 14 143 37|70|57|13.574 34008_3
HSPHTTRRHPRNEGTQRQQPPTMQEEKASTTISPSSSLQPQGNGQTPRPPQQPHAPANPRPSITRQQNSL RPASAPCG 14 14 9 35|52|49|6.08 1 10880_5 PSPSSPSCSPVQRTRPASAPCGRRRPQSVTSAHPSRGAAVSATPASSRPHRH THRHRLQF 14 142
30|54|42.0 |12.03 20835_3 STQQAHTPPAATHRGQGQEATHRHRLQFNPTANGIRPQPRCQQTQRTPSEQSPK530 17755_3 SYSSFGASSSSSAAAVTALHHAPAAITLTEREECVGVCRQHSDRHRQRRPSWHRHPFSTERQPTSLVT TTPGDSDG 14 14 14
21|25|25.0|1.030 23583_3 AATEKRQNVNNYTTPGDSDGSTAVS TQRQQPPT 14 143 37|70|57|13.574 34008_3 HSPHTTRRHPRNEGTQRQQPPTMQEEKASTTISPSSSLQPQGNGQTPRPPQQPHAPANPRPSITRQQNSL RPASAPCG 14 14 9 35|52|49|6.08 1
10880_5 PSPSSPSCSPVQRTRPASAPCGRRRPQSVTSAHPSRGAAVSATPASSRPHRH THRHRLQF 14 142 30|54|42.0 |12.03 20835_3 STQQAHTPPAATHRGQGQEATHRHRLQFNPTANGIRPQPRCQQTQRTPSEQSPK
PPPSPPAH 14 143 51|55|54|1.74 22704_1 ADNMRRGQQAQPHSPHTTRRHPPPPSPPAHFHNPREMAESPAATTQHTRRKRTLS PGVFETTG 14 147 40|44|40|1.40 2254_4 VACTPRVTDGPRLVDDRLSTKVPGERCAPGVFETTGLRLIDDGV YYTNPNRT 14 14
1 58|58|58|0.00 25282_2 HHTISYCRHLAINTPQQQLNKHLSILTRNHTLYYTNPNRTPPPRKFHKSGKIMYGGDA TRAPSRLR 14 145 26|39|34|4.17 1 20766_3 TSSATTTTTTTKAPITTTTEAPTTTTTTTRAPSRLREID PTQRSHTR 13 132 47|74|60.5|13.50
28075_4 LGAHDTSITNPPAQPPHHQIHVQHSQQSSHTRSKQQHHNAHAQANPHPSMTRHQNFSHPVLSGTRSPTQRSHTR CSRVSRPP 13 136 28|62|42.5|10.870 47271_4
KNGPKSRQHPHAWRAPCSRVSRPPAECKERPTRIHSNGTVIPTADRHAATSPLRHGTQANVV LGSRGHQA 13 134 33|57|41.5|8.732 44466_3 QQHRSTADPKGQERGAAMPRGITGLGSRGHQAVSHGPLAAWRSHGGGHDCAPPVCGA TPTPIEPH 13 133
18|24|18|2.838 34213_1 ISIYYITPTPIEPHLPENSAKIRK
VRHGAHHP 13 133 48|52|49|1.70 14609_2 TAGRGGWRPTLTVRHGAHHPLPSKCTQQGAQSPSHHRNGGTQTIVVADYRCL HAPYAAHH 13 13 3 45|51|46|2.62 1 40850_3 EGTPHEKRAAATHAPYAAHHAQTHSNPRQQAHTHTALPQHQHTDGRGRSKQ
KRWWIFGK 13 138 55|81|69.0|9.14 1 47226_7 GAFATGWSASSRQHENAGDAAAKSAPAPLAESREKRWWIFGKNGSGDKNAGFSSTAGTPSKVNSSGGGGNSGGMNGSVDDD
RWNQCDCR 13 139 22|37|29|4.37 1 11514_3 SLTLTLTLTRPPPQPQPHKPRNRWNQCDCRTRTPSCKPE HRLQFNPT 13 13 1 89|89|89|0.0 0 28618_3
HGRGEETQRHHRLQFNPTANGIRPQPRSQHTAHTEQAITQRATRRSDSSPIHDAAPHSMRSQRSAPHHQEETLVLRAPSLLPSLHAHGK RDGAQRIQ 13 13 5 13|85|41|23.09 0 31500_4
EWNERERDGAQRIQRESADNMRRQQAQPHSSHTTKRHPRNEGTQRQQPPTMQEDNTSTTISPSSFPQPQGNGKTPRPPHNAHGAG TNDRREKH 13 133 41|56|42| 6.853 18200_5
RQEALTIKMSREHAATPRRSTGGEDNRRKKQEAAQHTEETNDRREKHRHSPRTGRQ SSHPPCRR 13 133 30|59|37|12.364 40395_4 ARKAHNASSHPPCRRRRHPPPSPSAHSRSPSEMAKTPRPPHNAHGAGKPSSIHHAAQNY APHRHHRS 13 136
13|49|28.0|11.780 38438_4 TSTGTTTAYKSPNNSPGLPAAVVTHTDEQQQHKSARNNTAPHRHHRSHH
TTTTTTTT 135 1 56|56|56|0.00 27238_3 SDKRLSEEQANSESEDSTEETTTTTTTTTTTTQAPSTTTTTQAPSTTTTEAPAVST HTPIHGSR 13 13 5 31|50|45|6.43 0 16108_3 GEGSSQACTVRRAPRTTHTPIHGSRHTHTQPCHNTNTQMEEGEANRSGTQ
YYTNPNRT 13 13 1 50|50|50|0.00 41453_3 TLHINYQIFKLPYTLPLYYTNPNRTPPPRDNPQFWRNNVRGRCMNFPVQI TPSPKHLV 13 13 11 17|49|21|8.960 38989_2 ERINRKKGSKPSRSMSPDRDNRIKIPWTTRAATPSPKHLVKKPQDLRKG PALKKDEK 1333
69|80|69|5.190 39699_9 KKYEKAISPVPKKDEKVISPALKKDEKVPTPALKKDEKAPTPALKKDEKAPTPALKKDEKAPTPALKKDEKAPTPALKKD TVHTPRDG 12 127 14|54|20|14.210 47468_2
TTRREGSRPSSTVHTPRDGTRHHHHPAHFHNPREVAESPAATTQHTRRKRTLSW
HRPRCGQY 12 127 25|44|30|6.47 1 14796_4 DACGEVCSASLRARAAHRPRCGQYTCRKENMATQREEYRASVTG RGHHTTHT 12 122 48|67|57.5|9.52 40936_4
PREMAKPRGHHTTHTEQAPTTRGRSGSRASSIHDAATKFPAIPPSTGHGAQLNAVTRGASHPQERRR HHSLRPHG 12 12 12 31|58|47.0|9.750 44838_4 PVTQRREEQGKPVGDAAGLRRRDEGRHHSLRPHGRGPGGRAEVGQRSVYADGAGWETV
RWAWHSHG 12 12 5 26|41|29|5.46 1 11506_4 VPQSQWRWAWHSHGSLVGAPRSRTATPVGEAAEAHHPLQQR RLHQRRHT 12 12 10 48|77|64.5|9.160 7414_3
AHPGMPKSDVELSASNATIPHTCRLHQRRHTNHKKHPPHVIQTPFRHTATRQTEVFVTPMHLPTQSNSPCTTHGSSS PPHACSRR 12 125 38|60|46|8.450 42468_4 PTGPRRTPTRHGSLSAPFVAAPPHACSRRGTQGRHERVGGCASGAVGREHPAPLAGRVQQ
DPHRRRMP 12 123 27|41|30|6.02 1 44229_4 KLSPDPHRRRMPWTAAKEYVPGVVLNAKEKMVLDGVQLVDV KECVPGVV 12 126 24|40|33.0|5.19 1 26822_4 SPDPHRRRMPWTAEKECVPGVVRSSKEKLVLDGARRVDVE RSRGDGCV 12 12 9 21|32|29|4.07
1 41207_3 SAAHSLLLPSSFRWIRSRGDGCVCGVPSPQQH THRHRLQF 12 12 1 79|79|79|0.00 37728_3 THRHRLQFNPTANEIQPQPRSQQTQRTRSKQSPKEPHAEAIPHPPTTRRPTPSGRDTRPARCLSPPFASCPREKAAPTR AVRDGRWS 12 127 26|50|36
|8.530 1666_4 LRAVDGAGRRADCCAERVAAERCRAVRDGRWSVARCGDGGERRERAGGAV LSPRGKNQ 12 127 28|69|39|11.970 27355_6 CKTLNSFHVNKQCGTMPELHNNVGWRFPERPPINTARGHSKEKATPPFWLSPRGKNQSTIQQNAANHHI
QGTDSWRR 12 12 11 27|45|37|5.51 0 25928_3 DDDHYQCTRQEHNHLQGTDSWRRERPRRPFVGSGAAAAACQCRCG
PPPPPPTT 12 125 36|77|40|15.440 13382_7 TRQFIADNFAFVPCNEFLPAPTCSDAEQKTQGVEEIEKKRKIQRREWQLDRGRYHPSSPPPPPPPPTTTTTHSHDEE SSTHHHCR 12 127 30|81|49|16.850 8940_6
PSPTQKISAHLHKREGTKRNEGTIPPSRRGPTLSSTHHHCRAGNRSHIAESAPQCEPHHLVVNNGRDAARHRPAPFSHHSS TAAKGGRS 12 12 11 40|75|55|12.840 11913_6
ESESTASRVQSASNSAVRSATAAKGGRSSAEVRQPTGKKGTAKGNEGTSPVKRPSTRSPVTTPPVSAAAASPARH
HHLPQLIP 12 124 74|78|77.0|1.50 26470_4 KEYKKRERDGVQRIQRESRQHEERAAGPATQFTHHETTSTQGRHTTPAATHHAGEEDIHHHLPQLIPAAPGKWQKHRG GRGTEHSG 12 125 17|26|18|3.663 41465_1 VGQKKKNGTGGRGTEHSGYSERAQTT
YYTNPNRT 12 12 1 46|46|46|0.0 0 19106_5 QSCNTIKELPHRYIYYTNPNRTPPPRDNPLKCQNNVRGRCMKFPAK TNADNVRS 1274 55|88|73.0|11.780 11995_9
TPARSTAANPTTPQFNRHSTNADNVRSPLKRQSTNADNVRSPLKRHSANADNARTGFSSSKISPKYSKKDSIITSPVDRAKPLASNVP RLEAEEKE 12 4 4 56|85|71.5|11.25 0 36997_3
```

VGSSMQQQQSNDILWEAKQRRMAAEKERKRLEAEEKERRRLEAEEKERRRLEAEKERKRLEAEEKERRRLEAEEKERRRLEAEKE PTQRSHTR 11 112 19|64|41.5|22.5 1 35871_3
HHAGGEDIHHHFPQLIPAATTQRTRSKQTLIHPSRGTKISCHPAINGTRSPTQRSHTRCITPAG TSSRADGN 11 11 11 41|71|57|8.370 3601_4 FATGKRGSVMREESNLSSGHPQRAAVTSSRADGNNQRQRHRHSPSAPFSQNFTPVQNSNSNNAAVAANTAV
YITPTPIE 11 112 36|47|41.5|5.5 1 25378_2 NNYKHNNIINHPNILMRLQPHHLHHLCYITPTPIEPHLPVKPANQTK RPPQRGAR 11 11 10 12|17|12.5|1.69 1 33730_1 TATTPWLSRPPQRGARV GTHRHKPP 11 11 5 31|63|43|10.46 2 776_4
QERKPSGNSAPSIIPASNAITSKHPPPQMGTHRHKPPKTPATCTAANNTAASRWNATSPRCLH QQHSSHTT 11 11 11 12|76|30 |17.73 0 37096_4
SRVAEFSELICGTEEKEWNRRERNRAQRIQREGVDNTKRGQQAQQHSSHTTRRHPPPPSPPAHFHNPREMAESPAA EAWDDQRT 11 114 25|50|36.0|8.984 28423_4 RKTKQKAEAWDDQRTHSHYSRKAKHTHTGRERERERSPTQRNKNKTETVT TKQKQLRC
11 116 31|68|52.0|12.080 11107_4 NEKKKQKAEAWDYQRTHTLITPVKQNTHTEGEREKSRPLNAMETKQKQLRCSAHAARQDGSGACRAKE AAARHAPY 11 113 35|47|41|4.92 29716_3 HEERAAARHAPYAAHHARHTLQSTATRTHIHTEPCRNANTQMEEGEA
GTPTQDRS 11 11 11 14|41|32 |8.66 0 9639_1 VAAANLPSTAPPRRTHLPPQWWRPGTAGTPTQDRSCWTPPR THRHRLQF 11 11 1 57|57|57|0.00 17827_3 NGRGREETHRHRLQFSPTANGIRPQPRSQQTQRMEQAITQRATRGSDSSSIHDAAPH HRLQFNPT
11 11 1 71|71|71|0.00 45637_4 TTNSHRHRLQFNPTANGIQPQPAAGTQRTRSQQSPKEPHAEAIPHPSSHDAAPHSTRSHGERTVSSTELIR QRQLEELR 11 9 6 37|62|52.0|8.79 2 30636_5
LRAENEQLRVENEELRGEHEHKTRGLQEVSEQAEDLQRQLEELRAENEQLRVENEELRAEGE ADYVGAML 11 11 11 28|72|36|16.510 4991_5 AESLYQEGFISYPRTETDSFSFTDDELREIAGVQRDNPEVADYVGAMLDDSSHKFRRPLKGGHDDKAHPPIY
PWQGLKAH 11 11 11 24|38 |31|3.840 40200_3 GSLPWQGLKAHTKQEKYSRISERKQTTPVETLCKGFPA TATDKLQH 11 119 32|57|40|6.870 3418_3 VQRIFATEHSRVHRDTATDKLQHSAVHHHRPSTSHHRHEGHAEHNQHGPHEPQTRQH LHGWRNWT 11 11
11 29|43|34|4.570 24798_3 GQRSRHGITSRPHSGLHGWRNWTAPSPPHPTQKAVSGWTSVAG AQPHSSHT 11 113 21|91|26|31.890 43806_4
ERDGVQRTQRERADNMRRGQEAQPHSSHTTRRHHTTKTHNASSHPPCRRTTHPPPSPPSSSPQPQGNGKTPRPPHNTHGAGNHHTSHTRQW GRHPQRHT 11 117 17|19| 18|0.83 1 11848_3 AARAGRHPQRHTHSHTLPP PRPPHNAH 11 112
42|68|55.0|13.03 16932_4 HNASSHPPCRRRHPPPSPPAHFCSPREMVKTPRPPHNAHGASEPSSIHHAAPKLPAIPPSMGHGAQLS TVPQPAPE 11 11 10 49|51|51.0|0.920 25717_4
PQTVPQPAPETKAPPQSPCCDKRAGNARGLAPHFRFGRPDRKKDTAEETRT PCGEASHS 11 11 8 36|47|40.5|3.9 0 23251_4 LFPPHVELQRRNHHESHGEVPSQNPCGEASHSHRPTAASRCRWLQST SGVRQRGC 11 11 3 48|57 |56|4.03 0 41401_3
TSDRGRRAQIPAVECWHAQPAYTSSAHSGVRQRGCGSWHCCSTGTAAAPKETNVRSS GPRPGRQL 11 11 2 45|80|62.5|17.5 0 34762_5
EETNDRREKHRHGPRPGRQLHTSPRTHAPSQCCQSAGEKIRPSIQHTTHMEQAITQRATSGSGSSPIHDAAPQSMRSHGV RAPEPQVK 11 11 6 26|75|42.0|15.95 0 28260_6
PLNSTEMTAIKDRKPVSTRAPEPQVKIAPKPVAPAAPAAPAAPVAPAAPVAPAAPAVPAGNEGTAREKGYVGTNG SSDVIEPF 11 114 36|58|47.5|8.173 15254_4 VISKQSSSDVIEPFTSADVGMAEEDSPQNGNTDDPAPQGTSNDVLESVHDEPSNASTL
MHRTPHTT 11 11 2 31|55|43.0|12.0 0 28614_2 NHMRRKQQPGMHRTPHTTHDTHSNPRQQAHTHRDGRGRSKQERHTRLHRMHIRRE NSPSRAHA 11 11 8 40|63|50.0|7.22 0 38829_5
VRQLKPEAHRSSVRGDWEPTHTDSTLNKGKRPHWSASAAGFRTNTDGCYDHCNSPSRAHAVGA RCSQRHSC 11 11 8 38|49|39.5|4.79 1 6612_4 SYKPVSETDPRPSRRCSQRHSCSVSSAPTPCCIVARSRPSPSHAQGSQT DGQSRCSF 11 11 10
50|58|50.0|2.4 0 1254_4 RDMNEAQLLTRGQKNGIVRRLFGDDGQSRCSFSQIAETVDALNEKVWTAEFRQIDTEH
QQQQQQQQ 11 3 1 63|63|63|0.0 0 9422_3 LPPPPPPQQQQQQQQQQQGGTHHSKKARYEREEGGGQTRQYLHYQRHQGQHTQEREGMRRGGGR RNGKRDTG 11 118 32|52|33.0|7.340 15179_3
VPMPFDFRNGKRDTGGEKGWQRQQPQRQSTPFVSPRGPMVHRGNLDPTSFKN
RRPDAADG 11 116 26|70|56.0|14.720 45371_5 GEQHAVVERGGVPDARRDDDGGWRRPDAERHPWRRGRRPDAADGGAAAAVSVGTRPATRHAPELRAGVDG GRGRGRGE 1166 34|60|50.5|9.180 29039_4
FLNDEGVEYLRKYLFLPHDAVPNTHKAEYKVLEREGGRGRGRGEGRGRGRGEGRGRGRGE LYYTNPNR 11 11 1 43|43|43|0.00 14807_3 LSTPFLPTGLYYTNPNRTPPPRKIRDFEPNNVRGRCMKSGPKI PRRQGGRL 11 11 9 23|39|33|5.29 0 28284_4
GPRRQGGRLHHWGADRHDCCDCSQGHLPLLGACAGQGGA RGHHTTHT 10 10 2 36|55|45.5|9.5 4 43000_3 ETSTTISPSSFSAAPGKWPKPRGHHTTHTEQANPHPSITRHKITCHPAINGTRSP VRHGAHHP 10 102 37|49|43.0|6.00 31916_3
WRPTLTVRHGAHHPLFSKCTQQAGAQSPSCHRNGENKQSSSRTIGASSR RRYQPRVH 10 10 10 25|50|29.5|9.10 18483_3 AEGGPALPEQRDSCGAGGDGVLRCVASRRYQPRVHPQRPRDDPAKGHPPR TQRQQPPT 10 102 66|68|67.0|1.05 16583_3
GQQAQPHTSHTRRHPRKDGTQRQQPPTMQEEKISTTISSQLIPTANGKWPKPAATTQRTRSEHPPHQ SHGKVWRT 10 105 37|47|44|3.970 16001_3 GEAAEVHHVHAGAGRWTHAEGSGLHVGSSSDRLSHGKVWRTTKVTS
RLMVLTSD 10 10 8 25|64|45.0|14.33 0 25453_5 EGMGLKVEKGKPPQSWTYKAVGDSLEKDDGVGQSGAPRPRLMVLTSDKGWPYSWKWKENKSTRD PCSRTATP 10 103 26|45|28|8.520 38620_4
SQRRWAWHSHGGAVGAPCSRTATPVGEATEVHRPQQRNVQPTTTP THRHRLQF 10 10 1 56|56|56|0.00 4949_3 HTPPAATHNGRREETHRHRLQFNPTANGIRSQPAAGTQRTPSKQSPKEPHAEADPH ANGIRPQP 10 104 22|73|58.0|20.610 11440_3
QSAGERIRPSTQQYRTRNKQSSTNGQEQEATHSHHRLQFHPTANGIRPQPAAGTQRTPSKQSPKEPHAEAIPH
MEEGEANR 10 10 2 41|53|47.0|6.0 1 11391_3 NNTQMEEGEANRSGTQGSTACTYRESGPSCRHPAWRVRTHTRGGAAHTATQKH NGTRRSAP 10 10 10 38|67|52.5|8.610 1195_1
IQDATRRDIVHTAGNGTRRSAPHGGGGGAGGLLPRAGTKPGRPLPTGGHKPVPRACRQCTPGLFHSL LVHGKCSQ 10 108 27|79|40.0|15.020 9051_2
EEGEANRSGTQGSTACTYRESGPSCRHPSRRIHTHTRGGAAHTATQKHSNNEKGYPRTPCGLLVHGKCSQYSPQALHLQ RGQQAQPH 10 104 34|55|37.0|8.582 27640_4
EKNGTRGRGTGHSGYSGYNADNMRRGQQAQPHSPHTTRRHPRKEGTQRQQPPTMQ SSSSTRAQ 10 107 45|96|81|18.670 29592_5
YNSHTEPQHRQQNTSSTPTHPHARDSSGCRHSEHTTHPPAQPPHPHHQTHVQHSQPPSRTRTSSSSSTRAQGITRRTPHGAAPSPPSSSPVVESLN ARRKEFHQ 10 108 35|88|61.5|16.840 18133_5
LRLQQLEEAARRKEFHQTRGEQEPRHGRHERQNNGTRNDGYGNNASNRRDRVEWNDDRRGDRTDMRTLIRNEGSVSRNNRPERAERNN GGFGQQAG 10 6 6 42|60|51.5|5.71 0 45025_3
GGGFGQTAATGGFGQQAGGFGQQAGGFGQQTGGFGQAAGRGTAGGFGQNTGTTGGGFGQT VKGRESVS 10 10 10 37|96|67.5|18 .42 0 8018_6
QHTPTNFADATSTTTTTTKKSSSANKVKGRESVSGSVVHRMPTKPTSVVRNPRPRNDASQGERITLREHASSLVVSKDRSRDIEGHAAAEGLMAS AHRTLRAW 10 109 14|24|15|3.560 27125_3 ASTCGCLERPAHRTLRAWHGEGRA
LYYTNPNR 10 10 1 40|40|40|0.00 15614_4 KQRINLTYPSMLYYTNPNRTPPPRDNPLKSQNNVRGRCMN PAVGGFGS 1066 20|45|27.5|8.550 5824_3 HTSTPAVGGFGSATTTSTPAVGGFGSAAHTSTPAVGGFGSAAHTS QGRRHERV 10 10 4 33|38|33.5|2.06 1
24699_2 AQGRRHERVGGVRVEQLGGKHTAPLTGRVQKKQLKMEE ARRAQPAL 10 108 23|62|39.0|10.990 43321_4 EHSPQRSRLAGDCSTSVGRDGTGTGARRAQPALSHRQRCLTQSRCRSERRESGDGDGAARDD PFGQAAAG 1033 50|76|55|11.260
36781_8 DKPPPFGQAAAGDKPSPFGQAAAGDKPSPFGQAAAGDKPSPFGQAAAGDKPPLFGQGTVFDASRSTVFANAPGVAQ PAANTQRT 10 103 30|70|60|17.0 1 18696_3
KSHGGETQRAARETSAHRQQPHAMDEEKRPTASRTFSRPTGYGHIPAANTQRTRSEHPPRQPHVEVDPHP PTASTHRE 10 107 28|61|34|10.50 31081_5 DGASVTARTLHTPGATPPALTAAAPWDAPFTNKSHEHTRIHSQRLPTASTHREVRAQAREG
DLCGGRRG 10 109 28|43|37|4.05 1 10399_4 GRDCGCGGERRFRDVGGVLRWCDLCGGRRGGRGVDVGERAQRD EVAKPEEK 1064 30|73|48.0|15.880 19267_3
SRVSRENAIDGAKQKPENTKPEETRLEVAKPEEKKPEGAKPEETKLEVAKPEEKKMEDTKPEATKSEEPPKET PTQRSHTR 9 9 2 47|60|53.5|6.5 1 24666_3 PQLIPTANEKWQKHRGHHTTHTERAPTTPATRGNKISSHPALNGTRSPTQRSHTRCPA
YITPTPIE992 22|29|25.5|3.52 38649_2 TLMYLQPPCYITPTPIEPHLPDKITKSAK TGKSRPST995 37|73|55|13.260 37761_5 ARPRTGKSRPSTQKTYTVPEVTPNSFPQIRNKHTASHVQGRPLSHPPRGDKIYCCPALSGMRRPTPNGHTRYP PQQSRSPF994
30|40|37.0|3.94 1 9951_3 SKWDHKKHQFGNIRPPPQQSRSPFLTKAKACGGAADGGTH NDSVRAGD999 56|88|68|8.970 24806_5
NDSVRAGDRPRGGVGGTGRGDAASKHGRTGPRQPRHPLRGGIPDAAERTAAEEHRQGVGRDGVVPAGHWAEVDAQRSKPAAVEAAQRV SQVSRGWN994 39|58|48.0|7.60 27910_4
DTCEGWRCRSSQVSRGWNTAFIPQFIGDNRKGQHELRRTPMQLQDGIPCAASSPGEAE ARHAPYAA992 38|48|43.0|5.02 21239_4 EQQGRREVRQTHTEGTPHEERAAARHAPYAAHHARHTQPCRSNNTQME AQRNGRPL996 39|59|54.0|8.040 14571_4
HWRPAHTRTSSSLAPSRSTHAAQRNGRPLAPQPPPARGSRAGSDTLHRGTGQSMTHSTS QRAKDGQR 9 9 2 38|69|53.5|15.5 1 19417_5 PTTASRAGHHQKFQAYQRAKDGQRTSASSNDSRSPSATPLLQSPPAAEPRQSTGGPAAHAVGMPPSSAT
TPHSHGRA999 25|64|33|10.960 361_4 IPKKGQRLPDRLSSEDVLEMLASEPVAKSSTESVDSTPHSHGRAWGAKERAPRRPPREKPQTPA THTEETND992 39|81|60.0|21.00 45283_2
TTHTEETNDRREKHRHGPQPGRQLHTSPQTHTASSHPPGTKRKRESPPPPPVQSHGQRNTATAPQPTHKAQQAITQRATRR SPREMAKT993 43|64|56|8.65 1 33371_4
HRYSPREMAKTPRPPYQPHAPANPHPLMTRHKIPCHPALNGTRSPTQRGHTRCTTPAGKETRGA THRHRLQF99 1 54|54|54|0.00 23440_3 EDSALHTKAHTASSHTNARGEATHRHRLQFNPTDNGIRPQPCSQHKAQQAITQR RYFGVCDG999 48|71|69|8.140
29299_4 LREAERRAVERRYFGVCDGGSDAACKGGGRPKSDATKSTCWRKGKERSANACLDAPVGPATMNTADTQETG NTQMEEGE995 24|54| 28|10.840 19219_4 RRGQQSGMHRKPHTTHNTHSNPRQQAHTHTQPCRNTNTQMEEGEENRSGTQGST
NTKQQHRE 9 9 8 56|90|69.5|9.83 0 26623_6 EETHVAVEGAAARSKPVPFSKAASFQALNTKQQHREGRGQRQGGKDRGDKTEGKESMPTAKPPAENSGATISRPVRRAPAVVPSSQHPEK GPRPGRQL 9 9 2 27|51|39.0|12.0 1 41942_1
RRKGEKRQEAAQTHTEETNDRREKHRHGPRPGRQLHTSPHTRTIAVLSVRG GTASANPN999 21|62|21|12.640 18971_3 YQRDEKGKRAAPRRQSVRHLPADQASCAVGVAHRDAYVLVAPECGTASANPNTAVNTPLPQS
HTLIAPVK 9 9 3 40|51|43|4.64 2 9986_4 QPQNQKGKQNKITEAWDYQRTHTLIAPVKQNTHRGREREVAHSMQQKQNRN MHRTPHTT 9 9 2 25|52|38.5|13.5 0 13572_4 GQQPRMHRTPHTTHSTHSNPRQQAHTHTALPQQQQHRWKSEKQTGAAHKAPP
RMGTARAG 9 9 9 27|43|38|4.32 0 17242_3 FQTQASLFVPQDTQQKEHRNNRMGTARAGRDRDTRASPCIRSA SPPHSART996 46|58|53.5|3.730 8592_3 LHPPTQKGGVMGKGSPPHSARTKEYAAATPSTSHCKKEPCAHTHKRPQREGSGASKST
KPVAKTAA 9 9 9 16|42|35|7.6 0 16817_5 SAAKPAAKPAAKTAAKPVAKTAAKPAKKPTVKPAVKPAAKAA LYYTNPNR99 1 37|37|37|0.00 24058_2 SPHTTLNHHHLYYTNPNRTPPPRKIQKFHKIMYGGDA TRAPSRLR992 34|45|39.5|5.50 8166_4
YEVAAPENSIEEQKSNTMPTTMRDAATEASTTTTTRAPSRLREID PRPPHNAH992 43|59|51.0|8.04 21081_3 TASTQRKQRQQPHTMQEEKTSTTISPSSFPQPTGNGQKPRPPHNAHGASTHHTSHTRQW
IDASPPES996 21|28|25.5|2.540 41728_3 TALRQIDASPPESFTAAPCVVLPAQHSL LGSRGHQA992 60|62|61.0|1.00 38742_3 SERGSGGSRAQTRSQQHRGTADPKGQERGAAVPRGITGLGSRGHQAVSHGPPAAWRSHGGGN FFQHHDAA999 22|52|35|9.270
21133_3 PFLLDELKREYSHTDTRTVAAPYFFQHHDAAGCKTSRCRHKYEKENNMQRME DVGPRHVD 9 9 6 21|52|32.5|9.19 0 48407_5 RHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEQEDVGPRHVDPDHFRSTTQD RRCPNPSM998 34|92|60.5 |17.880
34748_5 MCSQRSYRTKSQQARRWEYDHCTYPSLASRQWRWNRNCMIARRCPNPSMSSPLPLPAPVSRSDAQQQDFMLHPSSCSRTVKSTHYYPSSTAS VCGAAGVC995 15|41|15|10.230 11137_5
PKGAPAQKKKHSEKVKESKQNYFLKAAGCGCVCGAAGVCQP RPSRTVHT 8 8 7 14|62|17|22.83 0 45130_3 RDGAPRIQREKEQTKRGEGSRPSRTVHTPRGSIHATQTTPAATHHAGGEDIHHHLPQLIPTA TKQRRRNQ885 65|79|78|6.71 1 41658_4
LHGQKGIKTATPVTKHSGTQKHPAIHNPRHSTTADTKQRRRNQVSDTIALSLPTQPHTVTVPSGGIPAKSHMTTGTAPV GRHGGPSP 8 8 5 31|52|40|7 .83 0 42035_4 VLKKWREQQQWQYIRHVSEEQSRAPTQRAGRETIRKGRHGGPSPTEKSSNTA
ATPRRNTG885 16|16|16|0.0 1 47590_2 ATPRRNTGGEDDKGKK LCRTGRQA888 24|44|34.5|6.30 18605_3 QRRVGACDCQHQLRCRGARLCRTGRQAERTNKTRGQQTLSFNVS TPTDEEKD887 40|87|56|15.650 16838_6
TDEGHTRPRTTPTDEEKDWRRISRGGQFLAVGRGSTANWSKKPLKQVDQNVGNGIVAFSGGLRAEDLDDGDFVTEVLRHPTIPTFVD HRQRRPSW884 22|51|26.0|11.52 1 2890_3 ARSNALPQSLPQRDKKESVDVCREHSDCHRQRRPSWHHPSRVSTQRATPET
AGRPRASG 8 8 5 40|85|61|15.78 0 41505_4 KRERNDAKKKIILLTDQVAGRPRASGQKKTSLRSGGRLATFSSFSSQCASPPTNHLFGPCSCGQTSRCWISRHSPRHRSEGKKRD52 1 2890_3
ARSNALPQSLPQRDKKESVDVCREHSDCHRQRRPSWHHPSRVSTQRATPET AGRPRASG 8 8 5 40|85|61|15.78 0 41505_4 KRERNDAKKKIILLTDQVAGRPRASGQKKTSLRSGGRLATFSSFSSQCASPPTNHLFGPCSCGQTSRCWISRHSPRHRSEGKKRD52
1 2890_3 ARSNALPQSLPQRDKKESVDVCREHSDCHRQRRPSWHHPSRVSTQRATPET AGRPRASG 8 8 5 40|85|61|15.78 0 41505_4
KRERNDAKKKIILLTDQVAGRPRASGQKKTSLRSGGRLATFSSFSSQCASPPTNHLFGPCSCGQTSRCWISRHSPRHRSEGKKRD
PLISTTPG886 36|68|45.5|11.880 15522_4 TPRDGTRHHHHTPLISTTPGKWPKAPRPPHNTHGASEPSVGNDAATKFTLPSSPHWNTEPNPTQPQEK GQRNSASS 8 8 7 31|64|57|12.0 1 22915_4
TSRVGCYRKCQWYQEAKHGQRNSASSNDNRGSSATPLLQPSLAAEHLRRSTEDSAAHAVGTPPP PRLTPCGH886 33|89|41.0|20.34 1 18484_4
KKERSTTSSSSTSFSQPTDMATSPQPAQGAHSHTTSHPPKLIPNHPRRGDMNLHWHPALNGMPRLTPCGHTRCTKPAEKESRAAHSLPP LTVRHGAH882 34|43|38.5|4.50 36268_4 LPPPRRNSVTPHSRSWPWHPTLTVRHGAHHPLPSERTQQGAQS
GVQSGSTT 8 8 5 21 |47|37|9.17 0 47611_4 VYGSTANGGNWYMPGKPITRVWGNSHNRSGHGVQSGSTTAIIEGKEV
SYQPRAGG888 38|93|61.5|15.220 42374_6 YNSYQPRAGGGFDGGQFRGGRGRGGGGYSRGGGLGQENGFNRPLQYQGESPSHAYHREEKPEEEIFKEHTPGINFDQYEAIKVHISPNDIPPM RHPRATRW887 32|63|47|8.320 24095_5
IAQEKGMSPTRHPRATRWQSCRGCPNATTTNSAATTEPCTLTHSVFSQSEQQGQRELRHTTHR ACSRRGAQ882 46|60|53.0|7.00 45428_4 TRHGSLSAPFVAAPPHACSRRGAQGHRHERVGGVRVEQLGGNTRHPSQEECSNKQLKREE EAARKRRG885
34|41|38|2.48 1 37298_3 LKDPVGNADTIADVDNILHRRGLEVEAARKRRGKPLQPDGL DPHRRRMP 8 8 3 25|39|35 |5.89 1 10656_3 MTLCTKGMELSPDPHRRRMPWKAEKECVPGVFHSSKEKM ERAAGPAA883 25|90|36|28.410 28073_6
RGEPRVCGGPQCAERSRQPPTLSSSHLPSPPNRPHTRAETERKRMQQEGAGRSTADTARERRQHEERAAGPAAQSTRHETASTQGWHTTP SSHPPCRR882 45|80|62.5|17.50 17118_4
HTSRDGVHARMAHNASSHPPCRRRRHPPPSPPAHLCSPREMAKTPRPPHQPHAPVDPHPPTMAETKFLPSRPQWDTEPNS QRSGKSGA 8 8 8 31|72|55.5|12.22 0 11300_4
VAARARQATQRSGKSGALRETPKFGGRREPLAPRTPVGRRDHNNASIGGASPQLSQQSTNNKPEVDDNRTVK THRHRLQF88 1 51|51|51|0.00 9718_3 PPHNKRTHRQQPHTNGRGEETHRHRLQFNPTADGIRPQPAAGTQHAHEASN GNTDDQAP 8 8 5
40|75|53|11.51 0 21775_6 SGALAPASSQTQNAGSHELLGTEMPVSGEHFPPNIDSPLMGQVDTADEESPRIGNTDDQAPHSVSPDVSESVGTN QQAQPHSS 8 8 2 43|53|48.0|5.0 2 41273_4
LIRGTEKKEWNRREGDGVQRIQRERADNMRRGQQAQPHSSQATRRHPRKEGTQ
RTGDRSHI883 23|93|73|29.440 44914_4 HLQKREGKNEGTIPPSRCGPTLSSAHRHRRTGDRSHIAEEAPQCAPHNRALKVGRDAARHRPAPFSHPQVLRAHKKQQTESECTQTQSSPSLA RARRPLPS 8 8 6 39|69|52.0|12.05 0 18720_3
PAATKYKSEGHASTQPQEEFDGWATAGAVGGVPNSQTERRARRPLPSECAQQTGAQSPSCHRRGEHKQS EKQRAAEA852 42|79|60.5|18.50 35863_5
MKVAEAEKQRAAEATKVAEAEKQRAAEATKVAEAEKQRAAEATKVAEAEKRKAAEAAKAMESQKQRFLERFAVLEEEKK LRSQGATF888 33|59|39.5|7.730 28960_5 KTVEVLRSQGATFGPVKAERKGKDAAAPARTEKKPKAAAAAADGAEEEDEAPREKKKPN
MIRENKKK 8 8 6 44| 63|54.5|5.74 1 37355_5 HVVDDALALLYPSVSLQAAETASFSGRREHIEHMIRENKKKKGEKGHETDGKKEETLRQLRED RQLHSIAR885 44|58|52|5.490 26639_3
PHSPPQPGSVARRQLHSIARPTRISEGYLCPPGRRSSTAAEDRCVPRTHTIRTPHTSP HTHTHTHT 8 7 2 39|40|39.5|0.5 1 26802_3 AGMRGTFEQPHRVAAQREHTHTHTHTRKRETEPERGPQEH TVRRSVQK888 48|70|62.0|7.680 16948_6
RPGTHSPTHTGNHTTSNDSSRTVRRSVQKDTARTHEKTAPSPPDRKHSPRDEGSGVEKAGEKKRNNKSHP
SGVRQRGC883 37|43|37|2.83 1 21592_3 PVAECWHAQPAHTSSAHSGVRQRGCRKQQRGTHPREAAEAAVD THTHRERE882 51|53|52.0|1.00 38920_4 VHSTTKSNKKTKQKAEAWDYQRTHTLIAPVKQNTHTHRERERGRPLNAIKTKQ TQQTHTPP882
65|78|71.5|6.50 26010_5 RSSTTHTEETNDRRGKHRQRPHPDRQLHTSPHTHKHTSSRCRQSPGERICSSTQQTHTPPAVTHNARGEATHSHHRLQ SSGERENP885 30|53|47|9.270 12928_6
STNADASRRFSSGERENPPKRRGGGGGCCSEEGFKGSTVRERCLQRSRCTWDS PHTLDKRT886 28|57|47.0|9.910 22102_4 LATVRTPVTPPSPHTLDKRTWAYRPWHAQTHRNKWEENSVQQRPHADTQNHTRHNSG GGAVGVRC885 31|38|33|2.42 1
5341_3 ERRCAVRDGRWSVARCGCGGERRGRTGGAVGVRCGGAE
QQQQQQQQ 8 2 1 61|61|61|0.0 0 41045_4 NIKQQQQQQQQQQPLQQKKKEKDYAAASVSPKGSGGAINTHTHTHTQSTRTHTHKKRERER TTTTTTTT84 1 38|38|38|0.00 7600_3 TTTTTTTTTKAPTTTTTTTTEAPTTTTTHAPSSIRKID HTPIHGSR882
21|40|30.5|9.5 1 15880_3 DTHRGHTTRGEGSSQGCTVRRTPRTTHTPIHGSRHTHTQP
DRVHTARS 8 8 8 29|78|43.5|18.85 0 15747_5 TAMGELTDDSPPPRQITSSQQRTPCEGCQWRSASPDRVHTARSSPMKRTPRGESPARPAWNQNRPTPRVKPTVSAPRT LYYTNPNR88 1 35|35|35|0.00 43633_3
PFHSHSLYYTNPNRTPTSRDNPQFWQNNVRVRCMN PNRTPPPR88 1 34|34|34|0.00 4199_3 HKHAIHNKPKHNHYSLLYYTNPNRTPPPRDNPQN WTRKDSKT 8 8 8 51|103|74.5|15.51 0 44656_8
LELLSGDMQHEAWEKTFSRHPVIHSIAQDMWTRKDSKTEEEVVTIERREREEDEKRMRRMSSSDWRAKFRDRERNPTAEEDEERPIYVIKPETFALLRLQPDV TRSGGSCR887 14|33|18|6.310 6269_1 RRRKPTACSEPLWSCHLRSTRSGGSCRCRRYHA
GPRGGGAE 8 8 8 28|35|30.0|2.28 0 12972_4 GRERRWGPRGGGAETPVRGLGHRLGSNGRRDAAEI
VQSHGQRD 8 8 3 56|59|59|1.41 1 44277_5 TVSQQNAGRRRQPRCHANNSVREPSQRCQSAGERIRPSTQQPHTPPPPPVQSHGQRDAA SMGRRSCD888 25|37|28.0|4.30 14252_5 WMRASSPADVLSMGRRSCDRPVFRPLSQSAPCRRQKE RVTVRCGP 8 8
8 16|26|19.0|2.93 0 20166_3 PPHTRRVTVRCGPPSRADERAEGSSN
KLRLQLRE885 55|70|63|5.9 1 21101_7 AALGEASGKLEAEELQRQLDALRRQKDKLRLQLREARRGEEKLDILRRHNEDLQSRLNDARRGQEKLDAV GGRGGDRG 8 3 3 32|55|38|9.74 0 18702_4
TKIQRTEQAGFRGGNSGGFRGGRGGDRGGFRGGRGGDRGGFHGGRGGDRGGFRGG LWATLPGD888 13|30|19.5|5.320 27923_4 LWATLPGDGGVTKGGIILSGHTDVVPVDGQ SPPSAGRQ776 35|69|53.5|10.95 1 28321_2
SARAHSSKRETTDQQIRRERQRIGVHSLLSPSPPSAGRQRHLLAPPHTKPAAGRDHAKNPQKHRKQTVL GKLRWRFQ 7 7 7 26|35|32|3.36 0 12648_3 VVFCPGTDGKLRWRFQGEKDWRKCARRPEEAPEVN PTQRSHTR772 30|50|40.0|10.03 14578_1
KTRSHHTTHTEQANPHPSITRHKISCHPAINGTRRPTQRSHTRCTTPAGK TRQPVTPH772 30|42|36.0|6.02 28862_4 NLKYTQKIKSIRDICYRGSNSVLGGRRCSTRQPVTPHSAANP YITPTPIE77 1 60|60|60|0.00 38933_4
NYNKFNSKSFLTNSLHNTQPFSKSNSTYKIHRIHFNLTHANPYITPTPIEPHLRVTTPKF RGCFVPQS 7 7 3 36|41|38 |2.05 2 27662_3 LAPKDSRGCFVPQSQWRWAWHSHSSVVGAPCSRTATPVGKT AEGTHQQH 7 7 1 68|68|68|0.0 0 15069_5
RKHPSCPLSSCRTPPRIPSAAERAARDQTLNRTAITSASNIGRHHTNTAPEHAEGTHQQHTASHHHRH RAAHQGSG777 35|56|46|6.170 44782_3 AEVQQTVVRAAHQGSGQNAARDEKKRQQRQQRQQQKQQQSDEMQSGTCDATALDTI ARRGHTPA 7 7 3
40|46|42|2.49 1 18322_4 TTARRGHTPAAHGQSPSSPPLPSLACSLQHKPTPRQRSENKERNTT ARHAPYAA 7 7 2 38|45|41.5|3.5 2 6913_3 PHEERAVARHAPYAAHHGSRHTHTQPCRNTNTQMEEGEANRSGTQ CSRVSRPP773 37|41|40|1.73 10916_4
AHTEQRTHTAKQKKREEKNDPKSRQQLLACRASCSRVSRPP RGHHTTHT 7 7 2 36|48|42.0|6.0 5 29112_3 TRRHPPPSPPAHSCCPREMAKPRGHHTTHTEQAPTTRGRSGSGPSSIH TPGVFETT777 29|29|29|0.00 43585_3
TKMPEERSTPGVFETTGLRLIDGVGSDAV PPPPPPTT774 34|36|34.5|0.83 1 25116_3 IQRRGWQLDRGRYHPSPSPPPPPPPTTTTTTTHSHD TAGWRRQG773 43|45|44|0.820 45876_4 ITSFAASGQRGKMERYAENAMPLHNTAGWRRQGSERFYPTLRRGQ
THRHRLQF77 1 50 |50|50|0.00 17171_4 ARRHPHNKRTHQQGEEVTHRHRLQFRPTANGIRPQPAANKHNAHGASNHP HRLQFNPT 7 7 1 49|49|49|0.0 0 9366_3 EKIRPSTQQTHTPPAATHNGRGKETHRHRLQFNPTANGMRPQPAANKHN AQPHSSHT772
24|36|30.0|6.05 8136_4 EGERRSAADTARERADKMRRGQQAQPHSSHTTRLGM RHPRNEGT77 1 90|90|90|0.00 19168_3
TTRRHPRNEGTQRQQPPTMQEEKTSTTISPSSFPRPPHNAHGASKPSSIHHAAPRFPAIPPSMGHGAQLNAVTRGASHPQENRLGVRSLS RPHFHPST 7 7 7 32|55|42|7.21 0 32217_3
MAGGVPSSATRRGSAAHRRPPDNVGRPHFHPSTPLPSRTSPLDARLPMEEDRSFC PSSCVQQA773 21|38|22|7.79 1 28484_1 LSAIKNKTRGKEMCAPPHRPSSRPSSCVQQARRTGPSS RNTGEKGE 7 7 6 28|41|34.0|4.02 1 1677_4
LTINMAREHAATPHRNTGEKGERGKGKKQHKQAEETNDGRE HESFLSLE 7 7 7 55|100|77|13.68 0 42297_8
EHGKGDPHKEATPTLTEIQPPQQLRQKQRLHESFLSLEKTGTEVERQLWSSPRRREKTEPRRPQAEGSPVFELPPSEVAAIESEDIVGRDGNDVPSHNLS FCPPGRQS774 24|49|42.5|9.340 25093_3
DRTVSGPHSPPQPGSAAPRQLHSIARPTRSKEGYFCPPGRQSSTAAEDH SEAGGVPP776 29|43|36.0|4.530 24179_3 AQRQTFRQSEAGGVPPQRRKPPQARRPPASPPVVCFPSARAAP GPRPGRQL 7 7 2 46|51|48.5|2.5 1 26768_3
GKKVRSSTTHTEETNDRREKHRHGPRPGRQLHTLKHTHSQQSPTRDEKEER RRWERRPD776 18|46|34.0|9.670 10334_4 SGWWEPDAERHSRRWERRPDAADGGAAAAVPLGKRPAARHAPELRA LPVPSTGE777 50|80|74|10.180 20957_5
AESRPTGKPSLPVPSTGESPPTTAGDTRNSTQTNEKTTTSGIGTDTEAPEPYSKNDVAEQHDEDRDPSDLIRSVNTGHPT SSDVIEPF 7 7 3 45|67|48|9.74 1 47877_6
VTPEAQHEAISSPQIQHSPAQPSESESGPVISKQSSSDVIEPFTSAGVGMAEDDSPQNGNTDDPAPQ THRDTPNR775 33|47|39|4.62 1 27883_4 PNIIPTHRDTPNRSIRHSTHLPTQSNSPRTTHGSSSHPSNKCSSDQC MHRTPHTT 7 7 2 51|51|51.0|0.0 0
30708_2 HMRRGQQPGMHRTPHTTHNTHSNPRQHTHTQPCRNTNNTHGRGRSKQERHT QRAKDGQR 7 7 2 56|64|60.0|4.0 2 30035_3 RRTNPGTPIVAEGLTCGFCDSSTVSPTSASRAGHHQKYQTYQRAKDGQRTSASSNDSRSPSAIP PFPPSSSG776
30|35|31.5|1.57 1 31214_4 WGLVGRSPFPPSSSGKAREGSRRIPTADARWLKEA AGHNQHGP776 22|43|34.5|6.80 44773_3 ATDKLQHSAVHHHRPSTSHHGHEGNAGHNQHGPHETQARQHQH CWVRQPPQ777 26 |46|34|6.160 19539_3
QTQASSRKCWSPCCMRDSDRRCWVRQPPQTVATPRSDVTAQHSLAA PNRTPPPR77 1 31|31|31|0.00 5181_3 TFNMFNHKYSNHKPPSLYYTNPNRTPPPRDN

APHRHHRS773 15|30|17|6.652 19695_1 THTDKQQQQHKGAGNNMAPHRHHRSHRHVS WRTARWWS 7 7 7 20|29|27|3.68 0 20834_3 ISFDVAWRTARWWSSHGSQWDTRPAQKSS NNHARWQH 7 7 3 24|38|28|5.89 1 28266_3
NTVRDTTGWQRYMRHGPDLNNHARWQHTQSKTKQKHGG

TRAPSRLR772 34|35|34.5|0.5 1 36599_3 TSTTATTTTAPSTTTTEMPNTATTRAPSRLRKIDG HTHTQRER772 63|66|64.5|1.50 11271_5 NSTPRHNSQTATHKQPKQKKTRMPPLPSPNKRPHTHTQRERGREISWVDGGKKEEEDVELTRNAPL PPHNKRTH 7 7 2
59|64|61.5|2.5 0 4237_4 SPRARGFGPPHNKRTHRQQSPTNEREREETHRHHRLQFSPTANGTRPQPRSQQTQRILSETSGS PAANTQRT772 55|59|57.0|2.0 1 14178_3 RGFGPPHNKRTHRQQPPAMEEEKRPNATIASSSIPRPTRHGHSPAANTQRTPSKLSPKE
LGSRGHQA772 31|55 |43.0|12.0 1 23528_4 RAQTQSQQYRSTADPSERERGAAMPQGITGLGSRGHQAALHGPPAAWRKHRGGHD VAPPSPHT773 34|59|55|10.96 1 31850_3
VRAEGSSTHTARHGPRTRVAPPSPHTLFKNSVQPRPHARNPNHARHNNDASRQPPTQQK APSRLREI772 25|36|30.5|5.50 4225_4 TTTTTTTTTGAPTTATTETPTTTMTRAPSRLREIDG LGEEEVPL772 45|52|48.5|3.50 4251_3
GKTTARSSTWEPRKEHQLALTLQGNKASVDVGGEPLGEEEVPLTGERPPEVL

WRQRHHHR 7 7 5 16|34|30|6.37 0 26765_3 GSWRQRHHHRAPAQYQQCAYCPSAKDGVHPFLFP DRRDYGDR 7 3 3 43|81|55|15.86 0 41451_5
VEPYCRNCGRNGHLSRDCRSGPRDNRRNQNDRREFVRRDYNDRRDYGDRRDYGDRRDYNDRRDYGDRREFVDRRDFNDRRD VTRWWKRP777 27|50|34|6.870 46260_3 SDRATTSSSVTRWWKRPNVLPHNRTVPPSSARSSPRGVTAESAHALVYSS
RNRRDREG 7 7 4 26|42|29.0|6.22 2 21573_4 ALSASRLSQRPTHTTEHRNRRDREGIDGVSGHQLSSSSLLVK THRHRLQL774 29|92|44.5|24.420 9732_6
CTSRKTCPNTMHNGKKPILKRRALSQRCQSAGERIWSSRQQTNTSPVITHRGRGGTTHRHRLQLNPTTNGIQPQPLTQRTTSKQSPKEPHAK TGAGENEA666 58|74|63.0|5.260 14508_6
NVSSLLSGELTGAGENEARRSGKGGRISASERITSGNLSRVSQSLNSGGVSRVGNRESVSQRLQEDGGASLISS LAKQQREE666 33|66|47.0|13.440 19284_5
RRAELAKQQREESRARKEELQRKQAEERRKKKEELQAETERLLAEARSAEEGEKKALAEKVRTGKE

YITPTPIE66 1 42|42|42|0.00 15927_3 MPHIIYLSNYHHKFKTNYNESIYITPTPIEPHLPENPAVSLK TLKKGGGG663 43|107|90|27.070 37195_3
TLKKGGGGGKGGTSGARQKQPKQPQKQKQHQYHHHQHQQHQRQQHQQQLRQQGMPMLMAQHVLLPSSSPTEGVFPFPARGSGSGVLPTPVPVFQQQQQPPPHLPFQP AEGTHQQH66 1 42|42|42|0.00 3352_1
NFSRISPAGTAPQHAEGTHQQHTASHHHRHHSCHLRAASSTN KRRHDPTV66 1 96|96|96|0.00 48227_4 TPSQKGRNKEKRRHDPTVPPRANNLLHPPPLPHRKTQSHRRVSTTMCTTPPCSERWARHSQTSPRSIFPPLKCCVPAQTGKHNQSIHKPNQHPHSL
GTHRHKPP 6 6 2 33|94 |63.5|30.5 0 5044_5 AGRDHAKNPQKHRKQTVLQRRSCDSNAKSRSAPTQQPPTHTKRGKGGGGKPAPKINRNGMKRKQCTLHYTSIKCHRSSKHPPPQTGTHRHKPPN RGGRGFGD622 50|56|53.0|3.00 6845_3
VENREENGYNGFGNRGGRGFGDRGGRGFGDRGGHGFGDRGGRGFGDRGGRGFGDRG KASSWMHS663 31|35|33|1.630 25340_1 DRWQKKASSWMHSGPSFRSDCWQTHASSRTCWSPC KECVPGVV663 25|34|29|3.682 4022_3
LHYSGHKGNETLSPDPHRRRMPWTAAKECVPGVV EARRHDGS666 15|19|15.0|1.890 29859_1 WGLREARRHDGSRVSPSGE

EKKLSLGE665 36|88|72|18.620 45297_5 EKMFQIMKPVTIQTLWKEKRAQAEEDDAATGKGAAEEEPQKSDRAKITKRDVTVPEKKLSLGEKLLLKAQERKKRERQERDGATNEEG WANDDRHC 6 6 6 25|72|45.0|15.85 0 19430_3
SHRRSQPRSLAASGMSPSRDGHKDVASNKTFRGTHWANDDRHCGMCGPSAAHLKSHFHASSSSSSGINTPVH RDGAQRIQ663 36|58|40|9.570 46128_3 SQLIRGTEKKEYNKRERDGAQRIQRESRQHGERAAGPAAHFTHHETAPTQGRHTTPAA
KWQHPSSR664 47|57|50.5|3.770 13489_4 SAQTNNTSTQGKSKRRDTAATTAHNHKWQHPSSRSPHTMRQLESVRLQDSHTALQTR LRRGQRSD664 30|43|34.5 |5.320 15763_4 ENAMPLHNTAGWRRQGSERFCPTLRRGQRSDDKACSLGTPNVV
HRLQFNPT 6 6 1 43|43|43|0.0 0 24387_3 NKRTHRQQSSANARGEATHRHRLQFNPTDSGIQPQPAANTQRT LTVRHGAH662 36|42|39.0|3.00 43888_3 TSRLTVRHGAHHPPNSKCNQQVAAQSPSCHCNGEQKQSSSRT RHPRNEGT 6 6 1 83|83|83|0.0 0
9415_4 RRGQQAQPHSSHTTKRHPRNEGTQRQQPPTMQEEKTSTTISPSSSLQPQGNGRNPRPPQNAHRACEPSSIHHAVKTSCHPAIN AGWRERSG663 24|64|43|16 .340 18729_4
GCAIAGWRERSGGKCVSAAVKGRAGCLAVEGISRPGQHRSCRAPSLSLTNPNPNPNPNPNPNPN HCRTGDRS662 27|54|40.5|13.5 1 46007_4 PTPKMSAHLHKREGTKRKGDMIPPSRRGSTLSSTHRHCRTGDRSHIAESAPQCA ERAQRDAG664
39|50|48.0|4.320 23321_5 CGGERRFRDVGGVLRWRDLRGGLRGGRGVDVGERAQRDAGELPVCGRSEP HHLPQLIP 6 6 2 46|52|49.0|3.0 2 36130_2 THHETASTQRRHTTPAATHHAGGQHIHHHLPQLIPAATTQRTRSRQTLIHPS IPAAPGKW662
21|31|26.0|5.03 17951_3 IPAAPGKWQKHRGHHNAHGASEPSSMTRHQN SATPASSR666 30 |68|55.0|13.520 23168_3 HSRTSASAPPSPTSPSCSPGQPMRPASAPCGSRHNQSVPSTHPSRGAATSATPASSRHHRHPAAPFHQ QAYGATRP666
49|80|63.5|12.150 5377_6 EDVLFYSVTTYQAYGATRPLESLRKRGYTEEEELVLGPIKRHARDARRQEQGEGGEEVEEEEEREESADALGGQLFIGDK SSTGRHAP665 30|69|45|12.850 45612_3
AQHDDHGGTPYRRDATPARSSTGRHAPCGSQGTLCMSMVPSFPHRGGRPASASAGTSSRREHGNSLSTS LPSPHGHG664 43|56|48.0|4.660 25856_3 SVSNDAATKFTLPSSPHWNTEPNPKQPQEKRRGLRTPSLLPSPHGHGRGQQPPHAT GHVLESNR 6 6 6
48|49|49.0|0.37 0 20955_4 GHVLESNRHVSRTERNERLRGTKGRERDGEATRNPFPFPSLSPPHPHPP PQPHPPHM666 54|80|61.5|10.230 2501_5
EGTTSRQGKEPPRPSPWPAPQPHPPHMRTVTTLPDRAREKAGPPHKKHTPCQKSPPTLFHRSGTNTPLVMCRGGPSATRH QRRPSWHR 6 6 2 28|65|46.5|18.5 0 47787_5
FGASPSSTAPAAAALDHASIAITLTEREESVDVRREHSNCHRQRRPSWHRPSCASTHRATPASGS YSAHTTSC665 33|69|45|13.330 33884_4 STGSAANIHHKNNSVAFGLPSYSAHTTSCSHRHHHCSSSTRIRDNKTQRMGGGVGGRGGAVAVGGGRSF
ERLRNVPG 6 4 3 42|47|47|2.36 0 46231_3 NEHGERLRNVPGERLRSVPGERLRNEHGKRPRNEHGERLRNVPGERL DRYTTHSH 6 6 5 21|36|33|5.27 0 26773_3 TPRQRRVEGKKEKHDRYTTHSHRTKTNGSSPPVATH HTHTHTHT65 1 47|47|47|0.00
16448_2 TTQFTHTHTHTHTHTHRLFTHQRRKPKKSPHAPPCLPKKPQKGKGEGDQ EKQTGAAH662 28|31|29.5|1.5 1 21008_3 PIHGSRHTHTHSPAATPTHRWKREKQTGAAH ELVAPAED632 45|52|48.5|3.50 45289_2
VKELVAPAENVKELVAPAEDVQELVAPAEDVQELVAPAEVVQELVAPAENVK SGDPLINR666 39|56|39.0|6.210 22042_5 DRLLGIHIRSGDPLINRAIGLHYRGGWKPWYNREDPPNSAISSTADEPGQEVGAAY YRRLSCHR666 30|41|33.5|3.530 12432_4
GGKEYRRLSCHRGRPVGAVFTAQKMPRCETEEAVGGFAESV CRGRSQCS 6 6 3 46|59|57|5.72 0 24713_5 SINSFREQPCPHPSVVCRGRSQCSPQRGRKPPGKSCSVDASETVLLHTLSPRHPQNYWK LQRRNHHK664 23|39|29.0|5.740 30330_3
VAAFPPPHVELQRRNHHKSRCEAPSKHPCGEASHSHRPT HREHTATA 6 6 5 40|59|43|8.03 0 1055_5 TITPAAVTPNNNNSRTQTHVTGEGRTLHREHTATATNSSRCHGWRFCHASAHTERRPNT LEGDALRR664 35|67|41.5|12.280 16715_4
SGRRVDAAGEGQRARRCLEGDALRRWRCDRCREHAELCAQPSAAAAEAVGEPFALGWGAFEGGVQRR TRTSIRPV663 47|56|50|3.740 44579_4 LRGSTHFFAVPDSLLEDNHSTPQTPREGSTRTSIRPVYMGRTPVPTRSSRKPQSGQ HTPIHGSR662
23|35|29.0|6.0 1 37810_2 QACTVRRTPRTTHTPIHGSRHTHTALPQHQHTDGR HTHSEARR662 37|46|41.5|4 .52 16345_3 HTHSEARRTPSAPLRVHSSGRPVTLAPSQWGNTNDRRRGPSMPLAE PNRTPPPR66 1 28|28|28|0.00 6875_3
NYKLHTYHLYYTNPNRTPPPRKSHGFGQ QQQHKSAR662 26|33|29.5|3.53 48033_3 RYMSSAPSRRHTDEQQQHKSARTNTAPHRHHRS SSHPPCRR662 52|61|56.5|4.50 2133_3
TPRDGIHATKAHNASSHPPCRRRHPPPSPPAHSRGHHTTHTEQANLIHPSRGTKLPAIPPS QGRRHERV 6 6 3 25|35|26|4.5 2 39424_3 QGRRHERVGGCAGGAVGRKHTAPSQEECSKKQLKE NLLAVAGD 6 6 6 43|69|53.5|7.99 0 32834_4
TGRLSARVVGKTNEEVFNRDNLLAVAGDHDSNRIPEAKRRRSLEEKERRETQAEKRNYKNDDNDDNDGA PRPPHNAH662 40|47|43.5|3.54 18806_3 PSSFPRPPHNAHGAGKHSSIHHAAQNFLPSRHQWDTEPNSAQSHAVH TTRALAYN663 45|107|86|25.750
6421_2 GYDGTDRSLPRMLLPRTDSHDPDTNDADPRSTSSAPSVSTPALSFTTTADALASTPLPNTNPCAAYTWLPDTRSTAPPDATTRALAYNAYSPPNTAASSATPPPTSS THTQRERE662 26|27|26.5|0.54 47049_2 SWRGSKHEHPPAHTHTQRERERERIQKQ
AGAVSGVI666 36|55|47.5|7 .070 1333_4 EYGMDEELGQAGAVSGVISSDHTRRSSLSRWRSNEGPERKMPEESQLQKAGAASG GTRHRRWA664 40|63|55.0|8.790 31578_3
AGVSGEASQTVAAHRYLREACGTVLGEATGGGAGGGTRHRRWAQTPQEKPGWRAAPDEGCGDA ALENTAKE666 46|66|52.0|8.170 14782_4 TRIDGIQRVDALENTAKEGRDDTPAKSKKGDLASLPKKPRVGAVGERKEKGKSGRGGEWYNTPCSE HGTPSTPV
6 3 2 33|48|40.5|7.5 0 26549_6 STPVDSSAHGTPSTPVDSSAHSTPSTPADSSANGTVLILPDGAALSTF THTHRERE66 1 48|48|48|0.00 16423_5 SIQSKSINKQNHTKDFITLHPPTTPRSSPSLLHRHTHTHRERERERIPL APHHSTQR 6 6 5 36|61|47|8.01 0
25179_2 TGRRVHRQHQLQFRTKFHDHKKHHNPQHRPTPLAPHHSTQRAHTSSTRPGTIIATTPVTCV00 16423_5 SIQSKSINKQNHTKDFITLHPPTTPRSSPSLLHRHTHTHRERERERIPL APHHSTQR 6 6 5 36|61|47|8.01 0 25179_2
TGRRVHRQHQLQFRTKFHDHKKHHNPQHRPTPLAPHHSTQRAHTSSTRPGTIIATTPVTCV00 16423_5 SIQSKSINKQNHTKDFITLHPPTTPRSSPSLLHRHTHTHRERERERIPL APHHSTQR 6 6 5 36|61|47|8.01 0 25179_2
TGRRVHRQHQLQFRTKFHDHKKHHNPQHRPTPLAPHHSTQRAHTSSTRPGTIIATTPVTCV

SLHPPTQK664 34|48|42.5|5.580 4331_4 PHSLRHLHSLHPPTQKGGAMGKRSPPHSACIKEYAAATPYTSHCNKEP DSHRTGTH 6 6 6 27|51|39.0|9.09 0 15028_4 THGGNRDVASAGHKLNAKGVPHASIQWRRDSHRTGTHAAASPSQHKTRRPN RTPVAPPS663
47|70|56|9.460 36937_3 PLATVRTPVAPPSPHKHAKVGGEFAAATATRRNPNHTHRNSGASRQSPTQQNNLPGSWPDREIIYGNQTS AAGPAAQF662 32|61|46.5|14.54 18406_4
GVSQLIGETERKKNGTRGRGTEHSGYSERERRQHEERAAGPAAQFTHHETASTQRRHTTPA EALGTLSR664 23|46|41.5 |9.03 1 42247_4 DGRRRVYESADKGESWTEALGTLSRVWGNNQKRHEKDVGSGFSTAT
RGDERCGR 6 6 4 16|39|19.0|9.42 0 8955_3 MESNRGDERCGRHTHRGHTTRGENSSQACTVRRTPRTTH KAVKETQA664 46|87|78.5|15.880 4085_6
KEIGAMPPKGGGTNGKHQHQQQQQQQRQGGKKGSKKSDDDEFDALLAKAVKETQAALPKAENGHHQKKQNNGAGKQQQKKTSATEEL PALNGMRR 6 6 4 30|41|37.5|4.39 1 41886_4 APPATRRGGSSAIHDATPKCHCRPALNGMRRPTPRGHTECN
GIRPQPRS 6 6 2 30|79|54.5|24.5 0 37468_4 HTLPEINHRGQGQEATHRHRLQLNPTDNGIRPQPRSQHTTHTKQAITQRATRGSDSSSIHDAVPHSMRSHGVHHFPNRT KPAAHQRV 6 6 6 31|43|39.5|5.24 0 36345_3
EAFHRNAESRRKPAAHQRVHPSSASRPPALPLHCAAFCKRTQA HSTSKQQC 6 6 3 40|47|45|2.94 0 39568_3 TAQESLRTQPRRHSTSKQQCQEAISAAAGSACFHKHANRPAGNLAPS SIGTGRGT663 35|55|49|8.380 1024_3
GAGRFHWHMLNKTSRASVPDPESIGTGRGTPPKDVDTAAPQKHPSPEEKEVASGA HRYSPREM 6 6 2 58|68|63.0| 5.0 1 45727_4 HRYSPREMAQKPRPPQQPHAPANPRPSMTRHKIPAIPPSMEHGAQLNAVTRGAPHPQENGVGLRSLSS
AGWQRYMR662 22|45|33.5|11.50 26823_3 VWVSVEEAEKRSTGRHINTARGTAGWQRYMRYDPDINNHARWRHT PPHACSRR 6 6 1 59|59|59|0.0 0 22701_2 QKQSKTRGKKCVLRPTGPRRTPTRHGSLSAPFVAAPPHACSRRGAQGRRHERVGGVRVE
VFGAPSST644 32|49|41.0|6.180 30000_3 PSSTAAKPPAESPFKNVFGAPSSTAAKPPAESPFKNVFGAPSSTDAKPP PHRSTRVG665 31|54|42|8.90 19242_5 NFPHRSTRVGAPRADCSSSSSSCKERKRQEWQDSRASESQEASPMLETSRPLLL HRPRCGQY 5
5 5 30| 46|31|5.98 0 28990_4 AAHRPRCGQYTCRRENMATQREEYRASVTGSHYTKTAAFLERTPAV TRSPTQRS55 1 68|68|68|0.00 31703_4 RHHLPQLIPAAPGKWQNPAATTQRTRSKHPPHEAEAEADPHPSMTRHKISSHPALNGTRSPTQRSHTR
TRQPVTPH55 1 53|53|53|0.00 37790_4 GSNSVQGGRKCPTRQPVTPHRAANPTVLTEAEEEVAAFRSRTHCRHRCHSPAT PGRQLHTS 5 5 1 69|69|69|0.0 0 13498_4
RAGTLEKRVRRGKTTSSTTHTEETNDRREKHRHGPRPGRQLHTSPRTHAPSQCCQSAGEKIRPSPQHTT

IRYFPTST555 36|83|72|16.040 4555_7 TMSSTLCCTSCTPDDTPLQAIRYFPTSTSRMDADSADDIKGRGSRKNKHDEERKRQAQGDSHVKLKYLESFDAYEDDVEEDGS YITPTPIE55 1 28|28|28|0.00 39925_2 LTHNQTTSYITPTPIEPHLPENLAFWPK
PSPTHAHS554 53|61|55.5|3.420 17904_4 AAKKDRKHGRDNQQAGERTTKTLPLAGATAPSPTHAHSDNAARPRTGKSRPSTQKTYTVPE QRSRLAGD 5 5 3 39|50|46|4.55 0 32878_2
QAAVGNIAREHSPQRSRLAGDCSTSVGRNGTGAGARRAQPALAHRQRCRT RFGRGRTT 5 5 5 40 |51|46|3.72 0 3356_4 PAESARFGRGRTTACFSEGCSRRESSEASMWRKLPSTLKEKSSRTSKFTAL AEGTHQQH 5 5 1 41|41|41|0.0 0 15398_3
STTAGTAPQHAEGTHQQHTASHRHHRHGCTQSREGKREEDE KRRHDPTV55 1 86|86|86|0.00 25435_4 RKEKRRHDPTVTPRANALLCPPPLPRRRPQSNRRVSTTMCTTPSCSERWARRGQASPRSILPPLKYCVPAQSSKQNQRAHKPNQHP RQSHRVRP 5 5 4
27|39|29.5|4.6 0 4162_3 RQSHRVRPPRSKEGYFCPPGRQSSTAAEDCWFRAPTQHK RSRHPAFQ 5 5 2 25|56|40.5|15.5 0 8050_5 VDRRAGRPAGSHRSNHEFGEVPTEYIPDAAHDCVRSRHPAFQTARPSLHRCSPDQH PAQSSKHN554 28|36|31.0|3.20
34013_3 SPRSIFPPLKCCMPAQSSKHNQSARKPNQHPHSLKH IYEVERNI555 51|68|60|6.020 42910_4 ERNEEVNRRIYEVERNIAEQRRLSHKNQAEFNKALAEQKRREAIRDKEEDTRKGLEEIRYHLEGDFLN QSQRRGSG 5 5 4 28|47|35.5|7.02 1 28239_3
QSQRRGSGAAQPSFVDTCDRWRCRRSQVSRGWNTTFIPQFIGDNRKG RSHGKVWR553 16|44|31|11.44 1 25059_3 VAEVHREHDGAGKWAHAEGGGLHAVGASSDRRSHGKVWRTTRAT KFYTVRNW 5 5 5 24|31|26|2.65 0 45287_3
SLRKFYTVRNWFTSQGQGAGNLSSGADYTGP PAHPHAAQ552 37|73|55.0|18.00 40683_3 TRIPAPTRRLRHPPAHPHAAQGSQCSQRQLSAVAGAPIAAIHSFLAWHRDRCDSSLLPSPSPSCCSMPSTSSQ AARHRPAP553 28|53|30|11.340 33222_4
TGDRSQIAESAPQCAPRHRAVNDGRDAARHRPAPFSHHSSIACPHKVANRIRV LDLARDYK555 37|47|42|4.260 15031_5 NINSTRHALDLARDYKMRCFIPSTIAAFGDKCGKVNTKDDTILNPST VVERQREH 5 4 3 41|61 |50|8.18 0 30333_3
LEAALLEDEAPHAVVERQREHRLGLDHDEAQPLEAARKRRGKPLQPDGLFAFDDVEPDDRE HLPHHHPF 5 5 4 31|52|42.5|9.57 0 47502_3 AFPSRTRNQHHHLPHHHPFLVPSLHSQQKSSVPLAHVGYWRFSPQLSSPVLL AERRERLR 5 5 3 32|59|51|11.32 0
22339_4 MLQRHGSQKAKLEAERRERLRVDRYGTMRDASQDTTERNSSTGRPSTPQAPLNDRWPRP GSHSRCRL 5 5 5 36|60|51|8.13 0 37244_5 GNGAGGRRWCIGKKDPSLEGTPRAGRADGSHSRCRLLHGSDDGGVRGYGVLPRELEHGGA LRVSLSPH555
38|41 |40|1.170 1978_3 LRVSLSPHPIPSHPHPLPREKNKTQKNPDEKKKGGRGRGNK DEIPTSTW 5 5 5 50|75|61|8.21 0 10698_6 SMPVTPPLSEDSGSRNSRADEIPTSTWRARRWSDFPYTRQSEKNPQKVNNQPHIVPQQRNGTMGAVNCEEFDAPN
GEGAERAC 5 5 5 14|19|18|1.94 0 38528_2 GGHCPCGCGEGAERACGGC HRLQFNPT55 1 41|41|41|0.00 9621_3 GKATHRHRLQFNPTDNGIRPQLTANKHNTHGASNHPKSHTQ IPSAAESA553 42|62| 43|9.20 2405_4
HSHQQRQPRQTAEAPTDDGHPQRHAHRPSPDARIPQRQTHSSCLLSFCRPPPRIPSAAESAV AAEAAKAV543 34|62|41|11.9 1 40592_4 MKVAEAEKRKAAEAAKAVETEKQRAAEATKIAEAEKQKAAEAAKAVETEKQKAAEATKVAEA HRAQVHRH 5 5 3
25|38|25|6.13 0 20196_4 VQGVSERVHRAQVHRHSDRPAEHHRVEARRRPSTANRH IKTKQKQL555 34|39|37|1.620 42130_3 PLNAIKTKQKQLRCSAHAARQDGSGACRAKESTNGPAWV RRRDPQPH 5 5 5 42|105|53|26.33 0 33041_4
DCEAAHTHTASSHPPWTRRRDPQPHPPPPHSHSQRIWPRPRSQHNAPTPTPPATRRKQTLSHPRRGDMNLHCHPALNRVPRLTPCGHTRCTKPTEKESRAAHSLP TPRDRWNQ553 25|39|31|5.73 1 29663_3
PLHHAQQQTAAPQQQQKRTPRDRWNQCDCRARTPPRKPQ KRIKWKDN553 57|58|57|0.470 3671_5 KRIKWKDNDGVSRISVGEREYWSRLTGGGGSGIETEDGTLVFPVEGTKKGEAPNDKKT TVPHDVTG555 38|58|43|7.180 9809_5
TPRWTAAHFATSHPLLNDITVPHDVTGENQTRNEGRAKPWRHQRGGGPKSSTHGHNGP PCSRTATP55 1 40|40|40|0.00 39410_3 CFVPQSQWRWAWHSHGGAVGAPCSRTATPVGETVKGRHPR PSFSHRGA554 39|79|63.0|14.710 30083_6
PSFSHRGATPPTPSPASSPPTQRRDNSRRCNTRNDQAAAESRCSRHDQTSGMQEIKEKGKQLPPQASQHNVEAPQPAHK DDQHLPRR 5 5 2 47| 79|63.0|16.0 0 16360_6
AARSSQLALNDDQHLPRRRHPANLLGNGDTTPICRHGRNTCRTERVSSATPAALCPPRCERMNVPLPPVSTAGAPPTSA GLPPRPSR 5 5 2 47|47|47.0|0.0 1 37848_3 AHWFGYRKSKQEPLPRAAPQNAVLPSRGLPPRPSRWRGEHPDNTRGP KRWWIFGK555
36|97|63|21.050 44885_7 SLGDMKASRERLVGRWRSEKSESHEYANNNPSSVGAFATGWSASSRQHENAGDAAAKSAPAPLAESREKRWWIFGKNGSGDKNAGFSSTAGTPSKVN EGEANRSG552 36|38|37.0|1.0 1 18737_4
TQRAKYMGRGQQLGMHRTPHNNTQMEEGEANRSGTKGS RAPEPQVK 5 5 2 60|63|61.5|1.5 0 2354_5 FLYNRPLNSTERTAIKDRKPVPKRAPEPQVKITPQPVAPAVPAGPAGPAVPAGPAGPAVPAGP MHRTPHTT 5 5 2 36|44|40.0|4.0 0 3938_3
RRGQQPRMHRTPHTTHDTHSNPRQHAHTHSPTATPTHRWKREKQ QRAKDGQR552 52|60|56.0|4.02 12819_4 GFCDSSTVSPTSASRAGHYRKCQAYQRAKDGQRTSASSNDSRSPSATPLLQPPPAAEPRQ GPSSRACR552 38|40|39.0|1.0 1 7915_3
GPSSRACRWCASGAVGREHTAPLTGRVQQQTIKEGRVEKK QRRQANPR 5 5 4 44|57|54.0|5.12 1 2842_5 GAAQRRQANPREWKAPNARNEADSHPRPLPRTHPLPQPSRTHQGKDTERKSPPHNAR HTHTHTHT54 1 38|38|38|0.00 36326_3
GGGGGFAGMRGTFEQPHRVAAQREHTHTHTHTQAGNRA LRTHRVNG554 33|47|36.0|5.340 43402_4 VRCISQGSCGALRTHRVNGWMPRSWGDQLSSPPSVARGNRESTASLP SEPVFEVG553 56|81|76|10.80 14378_5
QQQPQRRQQQTYQQGHDLTSSTLQPLSRGHIPSEPSSEGLNSVSSSHSRFSRSSEPVFEVGRTHMRLPERRRNTHSNVGHD IGNTSGTM554 30|40|38.0|3.84 1 14219_4 GHAQHQVRIGNTSGTMPSTHREPTANATRGKWGHNQRPSS ARSCLPRS553 37|41
|40|1.70 13290_3 VAARRSCSSSRSDVTASRLCSTSVFACSTARSCLPRSSLVA KREELQRH555 59|85|62|9.70 19014_5 NQTKALLLDRKMSEVEHNQEQKREELQRHAQERNEAMLAVAERRRNLSQERIERQQQREQQRRENLRRHEEQKKLEKDLKEQRAE
PSSRTHRS552 21|50|35.5|14.50 46799_4 IASEASIPRKRHRPPSSRTHRSMDGEPFIFQTPATLKLPDEMEEEDFSYH RGCRCAAT554 30|46|40.0|6.120 27536_1 MARCWCCGGRRRLCPGRGCRCAATASAAGSCRCRRIRAVRSSCGAR IHYRMCVG554
25|40|25.0|6.50 26644_1 IAPKAKKGKKKRKHKKRKELRKAIHYRMCVGVGEGASVPH PNRTPPPR55 1 28|28|28|0.00 27136_3 HKSHYNITNLKPLYYTNPNRTPPPRDNP QWRSGRRD 5 5 4 33|50|43.0|6.42 0 2155_4
PQQSRSPFLTNGKGLQWRSGRRDTQSQSNGLSREARNGDAPLLWEAQQKT

RRSAARYE555 30|41|37|4.030 4412_3 SSSGRRSAARYEREGVGAGGAAAGWGRWRGSGGVVGGSRAS AERVAAER553 45|50|49|2.160 15418_5 LRAVDGAGRRADCCAERVAAERRCAVSDGRWSAARCGGGGEQRGRAGGAV VRVGRCTH554
16|25|19.5|3.39 1 18813_1 DPRVVAAHGARARVRVGRCTHHWGG ARRTGPSS553 23|27|25|1.63 1 3141_3 VQQARRTGPSSWACRWCAGGAVGRETH DSSGCRHS552 46|85|65.5|19.50 26888_4
RTLLPRPRTHTRTCDSSGCRHSEHTAHPPAQPPHPHHQIHLQHSQQPSHTHGRAAAAAQERKDQHGAAPSPPPSSLPSSPVVESL LRSGRLWG555 23|41|25|6.570 35251_3 QLGFEYYYLRSGRLWGPPATRPRRCPSPSSTPSPSPEVKTH SSRNRRQP553
24|35|31|4.55 1 12699_3 SSTKYYDYRGANHDDHPRTVTSSRNRRQPQQLCVG

PKGRIKPA554 38|87|61.0|20.70 19915_4 RQEGHTKTRHAPKGRIKPANSAHRSNTAGSSRTHAPSSSFLISGSLPPHPTPAPHNTRGNQKTAEAATTAKPSPCCADATARKGAPA LGSRGHQA552 36|52|44.0|8.02 40016_3
ADPSERERGAAMPRGITGLGSRGHQAALHGPPAAWRSHRGGHDCGPPACGQL SSEGRCSR 5 5 5 29|39|38|3.66 0 6201_3 RSSAATSSEGRCSRWSVCSRTPRWTSVPCTTWCSSAAPP

INSRGLQS554 47|61|52.5|5.310 33298_5 TLKSPAAERQVSAVATPINSRGLQSRPGSSNKQQQQRPWGNEKGAKREAEAEEKSAFVSQH SKQSPKEP55 1 74|74|74|0.00 34134_3
YGHSPAANTQRTRSKQSPKEPHTQKRFLTHPQRSAPHHQEETLVLRAPSLLPSLHAHGKKQPPHAPQKNHPNPF SGVRQRGC 5 5 2 42|45|43.5|1.5 0 39819_4 FSHLFTSDHSRRAQIHAVECWHAQPTHTSSAQSGVRQRGCGKQQR ARHAPYAA552
35|42|38.5|3.53 15593_3 TPHEERAAARHAPYAAQHTLHSTATGTHRHTALPQRQHTDGR VLWCGAGG554 58|69| 64.5|4.060 2149_6 TGRVLWCGAGGRCEGEAEVLGSRGELPPAPEELATLPQDTQGLQGGVSGVEDKLLPASSRPLEEEDDDS
TTTTRAPS552 27|38|32.5|5.50 30282_3 TTTTTTTAPEAPSITTTETPNTTTTRAPSSIRRIDGSL APSRLREI 5 5 2 27|30|28.5|1.5 1 24746_3 KPPNTTTTTTTQAPSTTTTHAPSRLREIDG SNPLCRWC552 22|36|29.0|7.00 37345_4
LLGRPAHALQITSNPLCRWCRPLTLKKQDRPDLLRN THTHRERE55 1 47|47|47|0.00 37968_1 NSTPRHNSQTATHKQPKQKKHACPLFHPPTNVHTHTHREREGERYHG KASVYIDG552 27|33| 30.0|3.00 414_4 MLQGKKASVYIDGTSLGEEDVPLTGEAPLGLVH
VDQNTTGD555 58|70|63|4.020 1718_5 TNEAKHMNGGGVKTPVDQNTTGDGAVANSKKGASGQQKKQHPPKRGAEKQRTEDDVRSPVGKDVNTQVTT MANIAEKN 5 5 5 35|56|39|8.68 0 35819_5
ATLAGVNGVLNHKPKTAENNIAMANIAEKNRSESRMKTRDYTGHKQPARFRTGWET DAAPHSMR 5 5 4 32|63|48.0|11.76 0 44459_4 TRRPLPHHQPPAEADPQPSTTRRHEFTLPSRPQQDAAPHSMRSHAMHQTRGKGESGCALPPSS RWEHRCRP 5 5 3 39|45
|41|2.49 1 38652_4 VRSSRVALQRQGCVTPRWEHRCRPASHAQEALLLPLKKEKNQPQQ EPQVKIAP553 37|52|47|6.24 1 12908_5 PLNSTEMGAIKDRKPVPKRAPEPQVKIAPKPAAPAVPAVPAGNEGMEREKGD PRRPAVRV555 22|31|22|4.410 42437_3
VRQEVSLHGAERRARGVQPRRPAVRVKGQTE RPPPTRTA 5 5 2 52|80|66.0|14.0 0 47945_4 ASTRRPGSPSNKTPPRSPSSTKSRSAQAPQRSPTRSQPHAQTTHACTAGSRCQQAPPRSRRWQQHRPPPTRTAQAPSAPT PTRTAPAP 5 5 4 32|65|44.5|12.97 0
30184_3 TRSQPHARTPHACTAGSRCQQTPPRSRHWQQHRPPPTRTAPAPSAPTARPSTALRATTAPRRQQT VQSHGQRD 5 5 2 30|48| 39.0|9.0 1 9881_3 VSNHPPWKRRRDTLSPPEVQSHGQRDTTTAPLSTHNAHRASTHSTSHT NGAPIEDG 5 5 5
49|64|63|5.75 0 22626_5 PPLEELRAANGAPIEDGFDAYDRREDDRAARRERVRVVRGELNHPRGKPRQNTILKLDDSDEEK97 0 30184_3 TRSQPHARTPHACTAGSRCQQTPPRSRHWQQHRPPPTRTAPAPSAPTARPSTALRATTAPRRQQT VQSHGQRD 5 5 2
30|48|39.0|9.0 1 9881_3 VSNHPPWKRRRDTLSPPEVQSHGQRDTTTAPLSTHNAHRASTHSTSHT NGAPIEDG 5 5 5 49|64|63|5.75 0 22626_5 PPLEELRAANGAPIEDGFDAYDRREDDRAARRERVRVVRGELNHPRGKPRQNTILKLDDSDEEK97 0 30184_3
TRSQPHARTPHACTAGSRCQQTPPRSRHWQQHRPPPTRTAPAPSAPTARPSTALRATTAPRRQQT VQSHGQRD 5 5 2 30|48|39.0|9.0 1 9881_3 VSNHPPWKRRRDTLSPPEVQSHGQRDTTTAPLSTHNAHRASTHSTSHT NGAPIEDG 5 5 5 49|64|63|5.75 0
22626_5 PPLEELRAANGAPIEDGFDAYDRREDDRAARRERVRVVRGELNHPRGKPRQNTILKLDDSDEEK

K-mer Frequency Unique Sequences Sequences in TheLlargest Cluster Low Frequency | Bigger Frequency | Medium Size | Standard Deviation Cluster ID Representative Sequence ID Sequence PAAGGFGS 9979 5103 4924 12|86|42.0|13.92 0 2443_3
SAAHTSTPAVGGFGSATTTSAPAAGGFGSAAHTSTPAAGGFGSATTTSTPAAGGFGSAAHTSTPAVGGFGSATTTSTPAVGGFGSA QAAAGDKP 4468 1932 1719 12|80|34|13.67 0 11060_8
AAGDKPPLFGQAAAGDKPSLFGQAAAGDKPSLFGQAAAGDKPSLFGQAAAGDKPSPFGQAAAGDKPSLFGQAAAGDKPSP PDHFRSTT 1692 1342 1245 12|87|40|15.48 0 21057_6
VDPDHFRSTTQDAYRPVDPSAYKRALPQEEQEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHVDPDHFRSTTQ ELLGTEMP 1210 1210 1171 20|118|67|17.11 0 8533_7
ISKQSSSDVIEPFTSAGVGMAEEESPGSGALAPASSQTQNAGSHELLGTEMPVSGEHFPPNIDSPLMGQVDTADEESPRIGNTDDQAPHSVSPDVSESVGTNSDPDSFSSTNVSGGVD PAVGGFGS 1116 0130|6.
AMGEAGGR 397 397 388 13|63|37.0|6.54 0 7179_3 VATRSSMNASSSGSKKGRQDSVSDLASQQAAVAAAAKTAMGEAGGRSWSNVVKSPHSPRDATV FGQAAAGD 374 217 211 16|63|31|9.15 0 13365_5
EKTKTEQKTAPFVQAAADDKPSPFGQAAAGDKPSPFEQAAAGDKPSPFGQAAAGDKPSPFGQA AALEESMN 237 235 227 35|94|66|11.59 0 17125_6
RAFLDQKPEGVPLRELPLDDDSDFVAMEQERRQLLEKDPRRNAKEIAALEESMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDDSDFVA EERCTPGV 225 225 225 25|60| 48|7.47 0 4481_4
VRAMEGATDMPVACTPRVTEGLRLVDGRFSTKMPEERCTPGVFETTGLRLIDDVGSDAVL FGSATTTS 206 185 110 12|72|32.5|11.32 0 7203_3 GSAAHTSTPAAGGFGSATTTSTPAAGGFGSAAHTSTPAVGGFGSAAHTSTPAVGGFGSATTTSAPAAGGFGS
DVGPRHVD 204 197 166 15|73|34.0|8.98 0 2975_6 LPQEEQEDVGPRHVDPDHFRSTTQDAYRPVDPSAYKRALPQEEQEDVGPRHVDPDHFRSTTQDAYRPVDPSAY
RRSAQGGW 186 186 186 12|47|14.0|3.72 0 3726_1 AARSPRTPSHHRRSAQGGWPRWSRTAPQAAQSSPCASPSGRPSASCS PAVGGFGS 120 95 66 12|46|21.0|6.13 3 2114_4 AAGGFGSATTTSAPAVGGFGSAAHTSTPAAGGGNLGNSASAATSGT

```
IDSPLMGQ 99 99 72 15|106|41.5|22.43 0 7431_8 ATSSPQIQHSPAQPSESESGPVISKQSSSDVIEPFTSAGVGMAEEESPGSGALAPASSQTQNAGSHELLGTEMPVSGEHFPPNIDSPLMGQVDTADEESPRIGNTD SAAHTSTP 85 73 58 12|54| 17.0|8.86 0
13000_3 REQGRFGEGPFGGSTFAGGGFGFGSATTTSTPAAGGFGSAAHTSTPAVGGFGSA GVAVCGAA 57 57 56 16|68|16.0|6.89 0 12723_3 RRQGVAVCGAAGVCPPWPQESKTRGKEMCAPPTGPRRTPTRHGSLSAPFVAAPPHACSRRGAQGRRHE
SWSNVVKS 56 56 56 24|49|37.0|5.84 0 1455_3 ASQQAAVAAAAKTAMGEAGGRSWSNVVKSPHSPRDATVVRAVEPVNFEA RSTTQDAY 55 55 49 12|48|19|8.93 0 13385_5 PDHFRSTTQDAYRPVDPSAYKRALPQEEEEDVGPRHGDPDHFRSTTQD
VFGAPSST 48 28 28 20|75|41.5|15.1 0 17815_5 KPPAESPFKSVFGAPSSTAAKPPAESPFKSVFGAPSSTDAKPPAESPFKSVFGAPSSTDAKPPAESPFKSVFGAP KAAAAPAK 45 14 14 14|55|31.5|11.32 0 5130_3
APAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAAAAPAKAA SGDHFPPN 39 39 16 57|97|67.0|13.38 2 1778_6
VISKQTSSDVIEPFTSADMGKVEEEAPDSGTLAPASTPTQSAGGRELLGTEMPASGDHFPPNMASPLMGQVETVDEDSPRNGNTDDRAPHSISSDVL TSTPAVGG 32 19 10 12|54|25.5|12.89 0 17630_3
PAAGGFGSATTTSAPAAGGGFGFGSATTTSTPAAGGFGSAAHTSTPAVGGFGSA GRETEHSG 29 29 29 14|26|17|3.63 0 16531_1 VGQKKKNGTGGRETEHSGYSERAQTT PNIDSPLM 29 29 18 32|70|48.0|10.66 1 12780_6
EMPVSGDHFPPNIDSPLMGQVDTADEESPRIGNTDDQAPHSVSPDVSESVGTNSDPDSFSSTNVSGGADA SWCLDAEL 29 29 29 31|45|35|3.37 0 8144_4 YNACSDVTVTSLPGSLNGGDSWCLDAELVEKKDDNSKHKSVKGVC GGFGSAAH 27 18 9
12|48|35|13.63 0 5732_3 AHTSTPAVGGFGSAAHTSTPAVGGFGSAAHTSTPAVGGFGSAAHTSTP SSDVIEPF 26 26 17 36|81|68|12.02 1 1309_6
HSPAQTSESESGPVISKQSSSDVIEPFTSADVGMAKEESPGSGALAPASSQTQNAGSHELLGTEMPVSGDHFPPNIDSPLM
MRRGQQAQ 25 25 21 20|97|54|18.77 0 20859_4 WSQLISDKEGKQWNERERDGTQRIQRERADNMRRGQQAQPHSSHTTRRHPRNEGTQRQQPPTMQEEKTSTTISPSSFPQPQGNGQNTAATTQRTWSK FETTGLRL 25 25 22 30|49|48.0|5.33 0 10184_3
VACTPRVTEGLRLVDGRFSTKMPEERCAPGVFETTGLRLIDDVGSDAVL RHVDPDHF 25 25 14 28|33|31.0|1.25 3 18560_4 YKRALPQEEEEDVGPRHVDPDHFHSTTQDAYRP TNKTRGQQ 24 24 24 32|34| 34.0|0.4 0 20125_2
WQLRCRGAYLCRTGRQAERTNKTRGQQTLSFNVS SLPRRHPS 20 18 16 14|28|18.5|3.77 1 14861_3 SLPRRHPSPSATQHSLPRRHPSPSAAQH FGSATTTS 16 14 9 18|38|26|6.14 3 19197_3 FAGGGFGFGSATTTSTPAAGGFGSATTTSAPAVGGFGS
MNARAQEL 16 9 8 65|88|83.0|7.24 0 5256_6 SMNARAQELAREKKLADRAFLDQKPEGVPLRELPLDDDSDFVAMEQERRQLLEKDPRRNAREIAALEESMNARAQELAREKKLADRAF PDHFRSTT 15 15 9 12|47|16|13.28 0 2198_4
GGNIYSKMGPSAQNYDTQEEEDVGPRHVDPDHFRSTTQDAYRPVDPS PFGQAAAG 14 11 7 15|55|26|12.14 0 4428_6 AAGDKPSPFGQAAAGDKPPPFGQAAAGDKPSPFGQGTVFDASRSTVFANAPGVAQ ELLGTEMP 11 11 10 30|115|58.5|26.55 0 11486_7
REPSRPANVPVVMPEAQQEATSSPRSQLSPAQKSESKSDPVISKQTSSDVIVPSTSADVGKVEEEAPDSGTLAPASTPTQSAGGRELLGTEMPASGDHFPPNMASPLMGQVETVD GNTDDQAP 11 11 9 31|58|38|11.1 1 3101_5
IDSPLMGQVDTADEESPRIGNTDDQAPHSVSPDVSESVGTNSDPDSFSSTNVSGGVDA AAKAPAPK 10 6 6 18|43|31.5|9.44 0 15314_3 NKPASKPAAKPAAKPAAKAPAPKAEKKGAAKAPAPKAAAPAPK TTQDAYRP 10 10 10 12|17|14.5|1.5 0 12041_3
STTQDAYRPVDPSAYKR5 0 12041_3 STTQDAYRPVDPSAYKR5 0 12041_3 STTQDAYRPVDPSAYKR
AAGGFGSA 9 7 6 12|40|25.5|9.25 0 18188_3 TTSAPAAGGFGSATTTSAPAVGGFGSAAHTSTPAAGGGNL KVAEAEKQ 8 4 4 36|65|51.0|11.39 0 9639_7 KVAEAEKRKAAEAAKVAEAEKQRAAEATKVAEAEKQKAAEAMKVAEAEKRKAAEAAKAVETEKQR
PCCDKRAG 8 8 8 32|51|40.0|8.25 0 534_4 PQTVPQPAPETKAPPQSPCCDKRAGNAGGLAPHFRFGRPDRKKDTAEETRT TRRVTVRC 7 7 7 16|24| 19|2.62 0 10548_3 PPHTRRVTVRCGPPSCADERAEGS ALAPASSQ 7 7 6 19|100|23.5|28.75 0 17788_7
VIEPSTSAGVGMAEEESPGSGALAPASSQTQNAGSHELLGTEMPASGDHFPPNMASPLMGQVETVDEDSPRNGNTDDRAPHSISSDVLESVHDEPSNAKT DSSAHSTP 7 2 1 73|73|73|0.0 0 7080_7
HSTPSTPVDSSAHSTPSTPADSSAHSTPSTPVDSSAHSTPSTPVDSSAHGTPSTPVDSSAHSTPSTPADSSAN SSDVIEPF 6 6 3 57|80|58|10.61 0 4439_7
EAQHEATSSPQIQHSPAQTSESESGPVISKQSSSDVIEPFTSAGVGMAEDDSPQNGNTDDQAPQGTSTDVLESVHDEPSN PFGQAAAG 6 4 4 19|33|22.5|5.54 0 15447_5 PPPFGQAAAGEKPPFGQAAAGDKPPPFGQAAAG
THLPPQWW 6 6 6 27|66|35.0|13.28 0 1515_3 RKKKKADRTSAVCTAALTRPLPSCLVAAANLPSTAPPRRTHLPPQWWRPGTAGTPTQDRSCWTPPR YKRALPQE 6 6 3 30|39|34|3.68 0 20273_3 YKRALPQEEEEDVGRATLIPTTSARRLRTRTGPLIPRRT
GGSCRCRR 6 6 6 14|21|18.0|2.67 0 19665_1 WSCHLRSTRRGGSCRCRRCHA VDPDHFRS 6 6 3 32|35| 33|1.25 0 12326_4 ALGQLYEEERERGRSRDVGPRHVDPDHFRSTTQDAY HSSHTTRR 5 5 2 55|67|61.0|6.0 2 1653_4
VDNTKRGQQAQQHSSHTTRRHPPPPSHPAHFHNPREVAKTPRPPHNTHGASEPSVSNGAATKFTLPS
Fgsattts 5 4 3 13 | 44 | 16 | 13.96 0 18860_1 htstpavgfggsatststPavgglalprtlllprtl Regpaqer 5 5 14 | 15 | 0.4 0 12635_1 AREGPAQEREGDRCA SAYKRALP 5 5 3 30 | 34 | 30 | 1.89 1 8456_4
VDPPPSDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD.
YITPTPIE 5 5 3 27|29|28|0.82 1 18591_2 HAYITPTPIEPHLPVTTPNFIEIMYGGDA AEDNRLDI 5 5 5 38|41|38|1.17 0 9638_4 NELAEDNRLDILPGGSPNSLREKTRWNVNTELHPADRAEIG ALAPPSLG 5 5 5 57|73|69|5.73 0 19848_4
ALAPPSLGARWLAALTRHSAARSTKPAVCIIVRKAERIMFVKATTPAPASSTAAMQKYWTWQTPRGGCTAAIR RRQGVAVC 4 4 2 16|66| 41.0|25.0 0 15493_2
RRQGVAVCGAAGVCQPRPQNTTRGKEMCAPPHRPSSHAHTSRLPQRAVRRGPSSCVQQARRTGPSS
MRRGQQAQ 4 4 3 80|88|88|3.77 0 6038_4 VRRKGKEWNERKRDGAQRIQRESADNMRRGQQAQPHSPHTTRRHPRNEGTQRQQPPTMQEEKTSTTISPSSFPRPPHNAHGAGGPSSI PPCRRRRH 4 4 1 96|96|96|0.0 0 9668_1
SRTVHTPRDGIHARKAHNASSHPPCRRRRHPPPSPPAHSRSPREMAKPRGHHTTHTEQANPHPSITRHKITCHPAINGTRSPTQRSHTRCNTPAGK ANGIRPQP 4 4 2 53|78|65.5|12.5 1 5484_3
THTEETNDRREKHRHGPQPGRQLHTSPQTHTQPAVTHQGRRGRENHRHRLQFNPMANGIRPQPRSPHTKHSKQSPKEP GSAAHTST 4 4 3 16|27| 21|4.5 1 17075_4 GSAAHTSTPAAGGGNLGNSASAATSGT YGPLRPTG 4 4 4 48|63|48.0|6.5 0 1322_5
SGSVTSTEPTDGPMEPDYGPLRPTGMWNVEEVVDVKNSTVDFRRIDDVESEVIEALSQPDDAV PHRHHRSH 4 4 4 13|15|13.0|0.87 0 18522_1 NSTAPHRHHRSHHRP
```