000

## Identifying key amino acid types that distinguish paralogous proteins using Shapley value based feature subset selection

Anonymous Authors<sup>1</sup>

### Abstract

Paralogous proteins have a common ancestor but have diverged in functionality. Using known machine learning algorithms, we present a datadriven method to identify the key amino acid types that play a role in distinguishing a given pair of proteins that are paralogs. We use an existing Shapley value based feature subset selection algorithm, SVEA, to identify the key amino acid types adequate to distinguish pairs of paralogous proteins. We refer to these as the amino acid feature subset (AFS). For a paralog pair, say proteins P and Q, its AFS is partitioned based on protein-wise importance as AFS(P)and AFS(Q) using a linear classifier, SVM. To validate the significance of the AFS amino acids, we use multiple domain knowledge based methods : (a) multiple sequence alignment, and/or (b) 3D structure analysis, and/or (c) supporting evidence from biology literature. This method is computationally cheap, requires less data and can be used as an initial data-driven step for further hypothesis-driven experimental study of proteins. We demonstrate the results for 15 pairs of paralogous proteins. Code at https://anonymous. 4open.science/r/AFS AAC SVM-F3D9.

## 1. Introduction

Proteins form the fundamental machinery in living systems, having several vital functions such as DNA replication, catalysis, transport, environmental interaction, etc. Advancements in sequencing technologies have resulted in exponential growth of protein sequence databases (The UniProt Consortium, 2020). However, the number of experimentally verified annotations constitute a tiny fraction: only 0.57 of 250 million sequences in UniProtKB (The UniProt Consortium, 2020) have manually reviewed annotations. Experimental methods for determining biological process level functions (transcription, DNA repair, etc.) are high-throughput whereas methods for molecular function (catalysis, ligand specificity, etc.) are low-throughput and hence are not scalable. The relationship between sequence and function is subtle and has not been fully decoded yet.

Paralogs are proteins that have a common ancestor but have diverged functionally. The functional difference in two paralogous proteins is considered to arise due to evolutionary changes in the sequences (Yang et al., 2023). A typical experiment to investigate the role of an (or a group of) amino acid(s) in the function of a protein is to perform a site-directed mutagenesis experiment: replace one or more amino acids and test the effect of the sequence change (Kresge et al., 2006). In this work, we provide an algorithmic ML pipeline, consisting both feature engineering and feature subset selection, as a quick and resource-cheap test to assess the likely outcome from a site-directed mutagenesis experiment. We use a diverse dataset of 15 paralog pairs. Our datasets show a range of sequence and function diversity (details in Appendix B). Longest common subsequence score (lcss) is a metric to quantify sequence diversity and median within-class lcss is  $\leq 0.5$  in 12 of the 15 datasets, and the median interclass lcss for the corresponding classes is less than withinclass lcss. Functional diversity, as discerned from biology literature, also shows large diversity from subtle functional differences (e.g., trypsin/chymotrypsin) to drastic (e.g., lysozyme c/ $\alpha$ -lactalbumin). Function description is fine-grained (e.g., trypsin/chymotrypsin) as well as coarse grained (e.g, GPCRs).

Our findings are that small subsets of amino acids can discern differences between pairs of paralogs. The subset sizes are between 5 to 10, the median being 8. We provide validations from literature, MSA (a popular computational tool to assess evolutionary conservation) and logical consistencies; for many pairs such validations are more than one.

Towards this, we view a protein as the composite of its constituent standard 20 amino acids. We use amino acid composition (AAC) features, a Shapley value (Shapley, 1953) based feature subset selection algorithm (Shapley Value

<sup>&</sup>lt;sup>1</sup>Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

based Error Apportioning, SVEA) (Tripathi et al., 2020;
2021), and a linear support vector machine (SVM) classifier (Steinwart & Christmann, 2008) as tools to identify
key amino acid types that can distinguish a given a pair of
proteins that are paralogs. It yields quick results based on
which biologists can conduct detailed experiments which
are resource-intensive (time, cost, trained manpower, etc.).

062 063 The key results from our ML pipeline experiments are:

064

065

066

067

068

069

070

071

074

075

076

077

078

079

081 082

083

084

085

086

087

088

089

090

091

092

093

094

095

096

097

• Using known machine learning algorithms we demonstrate a data-driven method to identify key amino acids that distinguish two paralogous proteins.

- The SVEA algorithm identifies a subset of amino acid types (referred to as *AFS*) adequate for distinguishing two paralogous proteins. The size of *AFS* ranges from 5 to 10 amino acids out of 20. (Table 1)
- For a paralog pair, say protein families P and Q, the computed AFS is partitioned into AFS(P) and AFS(Q) using a linear SVM, to determine the family-wise importance of AFS. (Table 1)

• Domain knowledge based validation of AFS: The significance of the amino acids in AFS was validated for 14 datasets using various methods like (a) multiple sequence alignment (MSA) and/or (b) structural analysis and/or (c) supporting evidence from literature that report structural/functional role of these amino acids.

• Logical consistencies in the pair-wise AFS of three paralogous proteins (globins, Section 3.1.7, and GPCRs, Section 3.1.8). If families P vs Q and P vs R have  $AFS_1$  and  $AFS_2$ , then,

- we find common amino acids in  $AFS_1(P)$  and  $AFS_2(P)$ , except for one pair.
- amino acids in AFS<sub>1</sub> ∩ AFS<sub>2</sub> are either excluded from AFS<sub>3</sub>, which is from Q vs R, or have much lower Shapley value in AFS<sub>1</sub>, AFS<sub>2</sub>, or AFS<sub>3</sub>.

• Validation of *AFS* using test data (Section 3.2): The composition of amino acids is sufficient to classify several paralog pairs. A linear SVM classifies with high test scores (70-99%) using only the composition of *AFS* amino acids as features. (Appendix Table E5)

*AFS* are top ranked features with an alternate feature ranking measure, Marginal Contribution feature importance (MCI) (Catav et al., 2021). (Appendix Table E6)

Shapley values based feature attribution methods are popular for explaining machine learning models (Rozemberczki et al., 2022). One such method is SHAP (Lundberg & Lee, 2017), which assigns attribution scores to input features based on a model's output for a given instance input. Another method is SAGE (Covert et al., 2020), which assigns feature attribution scores based on a model's loss computed at the dataset level. Unlike these methods, where feature attributions are based on a trained model, the SVEA algorithm that we use for our task assigns scores to the features based on the distribution of the data points in the feature space and their ground truth labels. The SVEA algorithm uses a function v(S), which acts as a measure of inter-class linear separation between the data points in the space of the feature subset S. The scores assigned to the features are Shapley values computed using this function  $v(\cdot)$ . We also use an alternate feature ranking method, i.e. the Marginal Contribution Feature Importance (MCI) (Catav et al., 2021). MCI is an axiomatic approach that was proposed as an alternative to Shapley values to score and rank features. We find close agreement between the AFS computed using SVEA and the top-ranked amino acids using MCI.

Use of deep learning methods trained on large datasets is becoming commonplace in Biology; for example, prediction of molecular function via EC number or GO annotation (Bileschi et al., 2022; Sanderson et al., 2023), identifying input sequence regions relevant to model output (Zhou et al., 2016) and learning sequence-function mapping from deep mutational scanning experiment data (Song et al., 2021). The use of large datasets for training makes this approach highly resource-intensive. The approach we present herein needs much smaller datasets and, consequently, (i) is computationally cheap and (ii) has far wider applicability since labelled data validated by wet lab experiments is limited.

## 2. Methodology

We discuss the main components of our methodology.

### 2.1. AAC features

Consider a paralogous pair of proteins, families P and Q. We first curate a set of sequences, say  $D_P$  and  $D_Q$ , from a standard protein sequence database, SwissProt (The UniProt Consortium, 2020), with  $n_P$  and  $n_Q$  number of sequences each from families P and Q respectively. For a protein sequence  $\mathbf{p}^{(j)} = (p_1^{(j)}, p_2^{(j)}, \ldots, p_L^{(j)})$  of length L with  $p_k^{(j)} \in \{1, 2, \cdots, 20\}$  corresponding to the standard 20 amino acids, the AAC feature  $\mathbf{x}_j^{AAC} \in [0, 1]^{20}$  for  $\mathbf{p}^{(j)}$  is computed as follows,

$$x_{j,i}^{AAC} = \frac{1}{L} \sum_{k=1}^{L} \mathbf{1}_{\{p_k^{(j)} = i\}}, \, \forall i \in [20]$$

So  $x_{j,i}^{AAC}$  is the normalised count of the standard amino acid  $i, i \in \{1, 2, \dots, 20\}$ , in a protein  $\mathbf{p}^{(j)}$ .

#### 2.2. Feature subset selection using SVEA

Given a set, N, of features from the protein sequences of P and Q, we try to find the features  $S \subseteq N$  that contribute the most to the linear separation of P and Q sequences. With 110 *AAC* features, we have  $N = \{1, 2, ..., 20\}$  corresponding 111 to each of the standard 20 amino acid types.

112 We utilise the Shapley value based feature ranking and sub-113 set selection algorithm, SVEA (Tripathi et al., 2020; 2021), 114 to identify the most important feature subset  $S \subseteq N$ . Shap-115 lev value is a well known solution concept from cooperative 116 game theory (Shapley, 1953; Narahari, 2014) for distribut-117 ing the total worth of a coalition of players fairly among 118 each of them by quantifying each player's effective marginal 119 contribution. The SVEA algorithm considers the binary clas-120 sification task as a cooperative game among the features, 121 with a function v(S) as the worth of every feature subset 122 S. v(S) acts as a measure of linear separation between 123 the classes in the feature space of S. Accounting for class-124 imbalance, we define v(S) using a class-balanced hinge loss 125 function  $tr_{-}er(S)$ , which is defined as, 126

$$tr\_er(S) = \min_{w,\xi_j} \frac{1}{2n_P} \sum_{j=1}^{n_P} \xi_j + \frac{1}{2n_Q} \sum_{j=n_P+1}^{n_Q} \xi_j$$
  
s.t.  $y_j \left( \sum_{i \in S} w_i x_{j,i}^{AAC} + b \right) \ge 1 - \xi_j, \ \forall j \in [n_P + n_Q]$   
 $\xi_j \ge 0, \ \forall j \in [n_P + n_Q]$ 

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143 144

145

and  $v(S) = tr\_er(\emptyset) - tr\_er(S)$ . The minimizer in the above finds a linear hyperplane with the least class-balanced hinge loss in the feature space of S.  $\emptyset$  is the empty set and  $tr\_er(\emptyset) = 1$ , therefore,  $v(S) = 1 - tr\_er(S)$ .  $tr\_er(S) = 0$  implies v(S) = 1, i.e., the two classes are completely linearly separable in the feature space of S. The maximum value of  $tr\_er(S)$  possible is 1.

The Shapley value  $\phi(i)$  for a feature  $i \in N$  is computed as,

$$\phi(i) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v(S \cup \{i\}) - v(S))$$

146 Thus,  $\phi(i)$  is a weighted sum of the marginal contribution of 147 feature *i* to all the possible feature subsets that do not contain 148 *i*. Shapley values are unique solution concepts satisfying 149 the axioms - efficiency, symmetry and marginality (Young, 150 1985). The higher the  $\phi(i)$ , the higher the contribution 151 of feature i to the linear separation between the classes 152 and, consequentially, the higher the importance of feature i 153 distinguishing the classes. 154

Exact Shapley value computations are known to be exponential time. Hence, they are computed using a linear time
(in number of features) Monte Carlo approximation (Castro
et al., 2009) in the SVEA algorithm. As the number of
features is small (20), good approximations can be computed fast via larger sampling. More details of the SVEA
algorithm are given in Appendix Section C.

162 163 164 **Data-driven cutoff for selecting** AFS: The efficiency axiom of Shapley value implies,  $\sum_{i=1}^{20} \phi(i) = v(N)$ . If all features have equal contribution in achieving v(N), then  $\phi(i) = \frac{v(N)}{20}, \forall i \in N$ . Consequentially, if a feature *i* had lesser contribution than others then  $\phi(i) < \frac{v(N)}{20}$ . Therefore, we set  $\phi_{cutoff} = \frac{v(N)}{20}$  for selecting the key distinguishing amino acid feature subset,  $AFS = \{i : \phi(i) \ge \phi_{cutoff}\}$ . Each of the features in AFS uniquely corresponds to  $d \le 20$  amino acids from the standard 20.

### 2.3. Protein family-wise partition of AFS using SVM

We train a linear SVM, to classify P vs Q, using the composition of the amino acids in AFS as the features, i.e. using  $\mathbf{x}_{j}^{AFS} \in [0,1]^{d}$ , with  $x_{j,i'}^{AFS} = x_{j,i}^{AAC}$  and each  $i' \in \{1, 2, \dots, d\}$  uniquely maps to a  $i \in AFS$ . We use these linear SVM weights  $\mathbf{w} \in \mathbb{R}^{d}$  to divide the set AFS into disjoint sets AFS(P) and AFS(Q) based on the sign of the weights. Since  $x_{j,i'}^{AFS} \ge 0 \quad \forall i' \in [d]$ , the sign of the linear classifier weight  $w_{i'}$  indicates which class is relatively prominent in the amino acid corresponding to i'. So if the +1 class is P, then we divide AFS classwise as  $AFS(P) = \{i' \in [d] : w_{i'} > 0\}$  and similarly  $AFS(Q) = \{i' \in [d] : w_{i'} < 0\}$ . See Appendix Section D for details on SVM training.

A flowchart summarizing the steps for computing AFS(P)and AFS(Q) is shown in Figure 1.

### **2.4. Validation of** AFS

**Literature evidence**: For 14 different paralog protein pairs, we provide supporting evidence from protein biology literature for the significance of amino acids in AFS in the functional specificity of the protein pair.

**MSA analysis**: We also compute multiple sequence alignment (MSA) of randomly selected sequences from  $D_P$  and  $D_Q$  and analyze the conservation of AFS(P) and AFS(Q) amino acids within and across the respective families (Figure 2). MSA algorithms (Edgar & Batzoglou, 2006) aim to align multiple protein sequences by inserting gaps in the sequences while optimizing an objective. The objective is usually to minimize the number of gaps inserted while maximizing an overall score that promotes the alignment of similar (based on physicochemical properties) amino acids at a given position. The alignments are often used as a tool to determine homologous relationships between proteins and identify conserved or mutated regions in them.

**Structural analysis**: For paralog pairs that together function as heteromers (protein complexes made up of different types of proteins), we perform structural analysis to validate the role of AFS in the heteromeric structure formed by the paralog pair (Sections 3.1.7, 3.1.3 and 3.1.4).

**Using test data**: We test the classifier trained in Section 2.3 on a test data. (Details on test data in Appendix Section A.1).

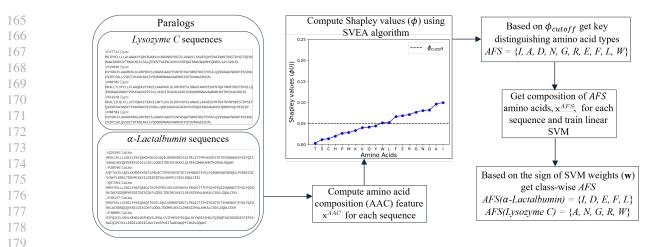


Figure 1: Flowchart summarizing the steps in our ML pipeline to compute the key amino acid types, AFS, that distinguish two paralogous proteins, using amino acid composition (AAC) features, Shapley value based SVEA algorithm for feature subset selection and class-wise feature subsets using linear SVM. Lysozyme C and  $\alpha$ -Lactalbumin are used here as representative examples of paralog pairs. AFS identified for other paralog pairs are given in Table 1.

In general, we find an imbalance in the number of sequences
for the two paralogous proteins. It is known that accuracy is
not a well-suited performance measure of the classifier in
class imbalance settings. Therefore, we use the arithmetic
mean of sensitivity and specificity (AM) to measure the
performance of the classifier (Brodersen et al., 2010).

<sup>192</sup> Using marginal contribution feature importance (MCI): <sup>193</sup> We check agreement of AES with another feature ranking

We check agreement of AFS with another feature ranking
method, MCI (Catav et al., 2021). See Appendix Section E.4
for details on MCI computation.

## 3. Results and Discussions

184 185

197

198

199

200

201

202

203

204

215

216

217

218

219

### **3.1.** Role of the amino acids identified in AFS

For 15 paralog pairs, we discuss the significance of the amino acids identified in the respective AFS (Table 1).

#### 3.1.1. Lysozyme C and $\alpha$ -Lactalbumin

205 **Literature evidence**: Amino acids D and E of 206  $AFS(\alpha$ -Lactalbumin) are found in the  $Ca^{2+}$  and  $Zn^{2+}$ 207 binding sites respectively of  $\alpha$ -lactalbumin (Permyakov & 208 Berliner, 2000; Permyakov, 2020). All  $\alpha$ -lactalbumins stud-209 ied so far are known to bind  $Ca^{2+}$  and  $Zn^{2+}$  whereas sev-210 eral (but not all) lysozymes do not bind  $Ca^{2+}$ .

212 **MSA analysis:** (Figure 2a)  $AFS(\alpha$ -Lactalbumin) and 213 AFS(Lysozyme C) amino acids (Table 1) are significantly 214 conserved in respective families.

3.1.2. TRYPSIN AND CHYMOTRYPSIN

**Literature evidence:** Y and W get the highest Shapley value  $\phi(\cdot)$  in AFS(Trypsin) and AFS(Chymotrypsin) re-

spectively (Table 1 and Figure E6b). In experiments to convert trypsin to chymotrypsin (Hedstrom et al., 1994; Hedstrom, 2002) it has been shown that Y to W conversion in loop-3 of trypsin leads to significant increase in chymotrypsin activity. We do not find S, H and D in AFS, which are important for the function of both families and are known as the catalytic triad (Dodson & Wlodawer, 1998).

#### 3.1.3. TUBULIN- $\alpha$ and TUBULIN- $\beta$

**MSA analysis**: (Appendix Figure E10) AFS(Tubulin- $\alpha$ ) and AFS(Tubulin- $\beta$ ) amino acids are significantly conserved in respective families.

**Structural analysis of** AFS: Tubulins typically exist as heterodimers, consisting of two subunits: tubulin- $\alpha$  and tubulin- $\beta$  (Mühlethaler et al., 2021). We looked at the contact residues of a tubulin- $\alpha$  chain and tubulin- $\beta$  chain in the 3D structure of tubulin- $\alpha/\beta$  heterodimer (PDB IDs: 3JAR, 5N5N). We see that the contact points of the tubulin- $\alpha$  chain in the heterodimer have more AFS(Tubulin- $\alpha$ ) amino acids than AFS(Tubulin- $\beta$ ). Similarly, AFS(Tubulin- $\beta$ ) amino acids are more than AFS(Tubulin- $\alpha$ ) at the contact point of the tubulin- $\beta$  chain in the heterodimer. Thus, the amino acids identified in AFS can be considered to be significant towards the quaternary structure of tubulin- $\alpha/\beta$  heterodimer. Appendix Section E.2 has more details.

### 3.1.4. HISTONE H2A AND HISTONE H2B

**MSA** analysis: (Appendix Figure E11), AFS(Histone H2A) and AFS(Histone H2B) amino acids are significantly conserved in respective families.

Structural analysis of AFS: Histones have a heterooc-

Paralog pair	Amino acid feature subset, AFS	Class-wise AFS parition				
Lysozyme C (74) and $\alpha$ -Lactalbumin (22)	$\{I,A,D,N,G,R,E,F,L,W\}$	$AFS(\alpha\text{-Lactalbumin}) = \{I, D, E, F, L\}$ $AFS(\text{Lysozyme C}) = \{A, N, G, R, W\}$				
Trypsin (66) and Chymotrypsin (17)	$\{Y, W, T, A, V, K, P\}$	$AFS(\text{Trypsin}) = \{Y, A\}$ $AFS(\text{Chymotrypsin}) = \{W, T, V, K, P\}$				
Tubulin- $\alpha$ (117) and Tubulin- $\beta$ (191)	$\{M,Q,K,N,F,I,H,A,C,Y\}$	$\begin{array}{l} AFS(\text{Tubulin-}\alpha) = \{K, I, H, C, Y\} \\ AFS(\text{Tubulin-}\beta) = \{M, Q, N, F, A\} \end{array}$				
Histone H2A (180) and Histone H2B (177)	$\{L,G,S,M,K,N,T,Y,F\}$	$AFS(\text{Histone H2A}) = \{L, G, N\}$ $AFS(\text{Histone H2B}) = \{S, M, K, T, Y, F\}$				
Interleukin-1 $\alpha$ (16) and Interleukin-1 $\beta$ (25)	$\{C,G,T,S,V,Q,A,N,P\}$	$AFS(\text{Interleukin-1 } \alpha) = \{T, S, A, N\}$ $AFS(\text{Interleukin-1 } \beta) = \{C, G, V, Q, P\}$				
Cytochrome P450 CYP3 (32) and CYP51 (32)	$\{H, F, G, K, A, P, N\}$	$AFS(CYP3) = \{F, K, P, N\}$ $AFS(CYP51) = \{H, G, A\}$				
Globins						
Myoglobin (107) and Hemoglobin- $\alpha$ (303)	$AFS_1 = \{E, S, Y, V, K, P, I, G, C, W\}$	$AFS_1(Myoglobin) = \{E, K, I, G, W\}$ $AFS_1(Hemoglobin-\alpha) = \{S, Y, V, P, C\}$				
Myoglobin (107) and Hemoglobin- $\beta$ (285)	$AFS_2 = \{K, V, C, E, W, N, F, M, Y, I\}$	$AFS_2(Myoglobin) = \{K, E, M, I\}$ $AFS_2(Hemoglobin-\beta) = \{V, C, W, N, F, Y\}$				
Hemoglobin- $\alpha$ (303) and Hemoglobin- $\beta$ (285)	$AFS_3 = \{W, P, N, S, G\}$	$AFS_{3}(\text{Hemoglobin}-\alpha) = \{P, S\}$ $AFS_{3}(\text{Hemoglobin}-\beta) = \{W, N, G\}$				
GPCRs						
Rhodopsin-like (181) and Glutamate-like (89)	$AFS_1 = \{D, Q, E, G, M, L\}$	$AFS_1(\text{Rhodopsin}) = \{M, L\}$ $AFS_1(\text{Glutamate}) = \{D, Q, E, G\}$				
Secretin-like (90) and Glutamate-like (89)	$AFS_2 = \{W, H, Y, V, D\}$	$AFS_2(\text{Secretin}) = \{W, H, Y\}$ $AFS_2(\text{Glutamate}) = \{V, D\}$				
Rhodopsin-like (181) and Secretin-like (90)	$AFS_3 = \{W, E, M, S, V, H, Q, A\}$	$AFS_3(\text{Rhodopsin}) = \{M, S, V, A\}$ $AFS_3(\text{Secretin}) = \{W, E, H, Q\}$				
Rhodopsin-like GPCRs						
Aminergic receptors (186) and Lipid receptors (113)	$AFS_1 = \{L, P, E, W, F, M, D\}$	$AFS_1(\text{Aminergic receptors}) = \{P, E, W, D\}$ $AFS_1(\text{Lipid receptors}) = \{L, F, M\}$				
Aminergic receptors (186) and Peptide receptors (367)	$AFS_2 = \{L, F, E, M, K, D, V, R\}$	$AFS_2(\text{Aminergic receptors}) = \{E, K, D, R\}$ $AFS_2(\text{Peptide receptors}) = \{L, F, M, V\}$				
Lipid receptors (113) and Peptide receptors (367)	$AFS_3 = \{P, R, G, I, W, S, V\}$	$AFS_3(\text{Lipid receptors}) = \{R, G, S\}$ $AFS_3(\text{Peptide receptors}) = \{P, I, W, V\}$				

257 tameric structure comprising of two H2A/H2B dimers and 258 one H3/H4 tetramer (Dutta et al., 2001). We looked at 259 the contact residues of an H2A chain and H2B chain in the heteroocatmer structure of histone (PDB IDs: 261 3KWQ, 1AOI). We find that the contact points of H2A chain in the heterooctamer have more AFS(Histone H2A) 263 amino acids than AFS (Histone H2B). This is interest-264 ing since AFS(Histone H2A) has only three amino acids, 265 while AFS(Histone H2B) has six amino acids. Simi-266 larly, the contact points of H2B chain in the heterooctamer have more AFS(Histone H2B) amino acids than AFS(Histone H2A). Thus, the amino acids identified in 269 AFS can be considered to be significant towards the quater-270 nary structure of the histone heterooctamer. See Appendix 271 Section E.3 for more details. 272

## 273

229 230

238

274

### 3.1.5. Interleukin-1 $\alpha$ and Interleukin-1 $\beta$

**Literature Evidence:** *C* has the highest Shapley value and is in AFS (Interleukin-1  $\beta$ ). Deleting *C* results in loss of activity in Interleukin-1  $\beta$  (Veerapandian et al., 1992). We do not find such studies for Interleukin-1  $\alpha$ .

**MSA** analysis: (Appendix Figure E12) AFS(Interleukin-1  $\alpha$ ) and AFS(Interleukin-1  $\beta$ ) amino acids show significant conservation in respective families.

### 3.1.6. CYTOCHROME P450 CYP3 AND CYP51

**Literature evidence:** H, F and G, in the respective order, have the highest Shapley value  $\phi(\cdot)$  for this paralogous pair (Table 1 and Figure E6f). H and G with the highest  $\phi(\cdot)$ in AFS(CYP51) have been reported (Nitahara et al., 2001; Lepesheva & Waterman, 2004; 2007; Strushkevich et al.,



Figure 2: Multiple sequence alignment of sequences from the respective families in (a), (b) and (c). Within each alignment, 15 sequences on the left are from one family, and those on the right are from the other family in each of (a), (b) and (c). The sequences are randomly selected from the train set of the families. For each aligned sequence in (a)  $AFS(\alpha$ -Lactalbumin) amino acids are in green and AFS(Lysozyme C) are in red, in (b) the amino acids in  $AFS_1(Myoglobin)$  are in green and  $AFS_1(Hemoglobin-\alpha)$  are in red, and in (c) the amino acids in  $AFS_2(Hemoglobin-\alpha)$ are in green and  $AFS_2(Hemoglobin-\beta)$  are in red. The intensity of the color is proportional to the Shapley value  $\phi(i)$  of the amino acid *i* (Figures 3 and E6).

2010) to be important in the enzymatic activity of CYP51. Mutation of these amino acids at specific positions has been shown to result in a decrease in the activity of the enzyme (Lepesheva & Waterman, 2007; 2004). Similarly, *F* with the highest  $\phi(\cdot)$  in *AFS*(CYP3) is also known to be important in the enzymatic activity of CYP3 (Qiu et al., 2008; Denisov et al., 2019; Zhang et al., 2024). A cluster of *F* residues in CYP3 is known to form a substrate-binding pocket with an

338 active site (Zhang et al., 2024).

340 3.1.7. GLOBINS

339

MSA analysis: (Figures 2b,2c and Appendix Figure E8)
For the three globin paralog pairs (Table 1), we observe in
the MSA, conservation of the class-wise partition of AFS
in the respective families.

346 Structural analysis of AFS: Myoglobin is a monomer, 347 while  $\alpha$  and  $\beta$  chains together constitute hemoglobin, a 348 tetramer of composition  $\alpha_2\beta_2$  (Dill et al., 2017). We su-349 perimposed the 3D structures of myoglobin, hemoglobin-350  $\alpha$  and hemoglobin- $\beta$  (PDB IDs: 3RGK, 1HHO) and 351 mapped the  $\alpha, \beta$  contact residues (based on (Shionyu et al., 352 2001)) of hemoglobin tetramer to that of myoglobin. We 353 find that the amino acids K, E, I, which are common 354 in  $AFS_1$  (Myoglobin) and  $AFS_2$  (Myoglobin), are less in 355 number at the contact residues of hemoglobin tetramer and 356 more in number at the corresponding locations in myoglobin, 357 which is a monomer (see Appendix Figure E7).

358 Literature evidence: W with a significantly high 359 Shapley value  $\phi(W)$  (Figure 3b), is present in 360  $AFS_3$ (Hemoglobin- $\beta$ ). It is highly conserved at po-361 sition 40 in the MSA (Figure 2c) in hemoglobin- $\beta$ 362 sequences as compared to hemoglobin- $\alpha$  sequences. This 363 W at position 40 has been determined to be present 364 in hemoglobin- $\beta$  at one of its contact positions to 365 hemoglobin- $\alpha$  in the tetrameric structure (Shionyu et al., 366 2001) and is, therefore, a structurally and functionally 367 significant residue. C, present in  $AFS_1$ (Hemoglobin- $\alpha$ ) 368 and  $AFS_2$ (Hemoglobin- $\beta$ ), has been shown to play an 369 important role in the tetrameric structure of hemoglobin 370 formed by  $\alpha$  and  $\beta$  hemoglobins (Kan et al., 2013). 371

372 **Logical consistencies in** AFS (refer to Table 1 (Globins) 373 for  $AFS_1, AFS_2, AFS_3$ ):

•  $AFS_1 \cap AFS_2 = \{E, Y, V, K, I, C, W\}$ . Except for W with the least Shapley value in  $AFS_1$  (Figure 3a), the remaining are excluded from  $AFS_3$ .

• Explanation: V, Y, C in  $AFS_1$  (Hemoglobin- $\alpha$ )  $\cap$  $\overline{AFS_2}$  (Hemoglobin- $\beta$ ) can be expected not to be *key* in  $AFS_3$  for distinguishing  $\alpha$  vs  $\beta$  hemoglobin.

 $\bullet AFS_2 \cap AFS_3 = \{W, N\}. N \text{ is excluded from } AFS_1,$ while W gets the least Shapley value in  $AFS_1$  (Figure 3a).

383 384

374

375

376

377

378

379

380

•  $AFS_3 \cap AFS_1 = \{W, P, S, G\}$ .  $\{P, S, G\}$  are excluded from  $AFS_2$ , while W gets the least Shapley value in  $AFS_1$ .

The Shapley value for W is very close to the cut-off in  $AFS_1$  (Figure 3a). If it is dropped from  $AFS_1$ , then the exclusion principle illustrated above would be more prominent as in GPCRs (Section 3.1.8).

### 3.1.8. G-PROTEIN COUPLED RECEPTORS (GPCRS)

**Literature evidence**: W (with highest Shapley value  $\phi(\cdot)$ ) and H common in  $AFS_2$ (Secretin) and  $AFS_3$ (Secretin) (Table 1 and Figure 3), are well conserved at multiple positions with structural importance and functional importance in secretin-like GPCR sequences (Cary et al., 2022; Harmar, 2001). Mutating certain conserved W leads to a loss in expression of this GPCR at the cell surface, where it functions (Cary et al., 2022). H present in the intracellular loop region is also known to be important in the activation of certain secretin-like GPCRs (Harmar, 2001).

M common in  $AFS_1$  (Rhodopsin) and  $AFS_3$  (Rhodopsin) has been found to be present at important binding pockets and a position important for activation of the GPCR (Okada et al., 2001; Sakmar et al., 2002). S from  $AFS_3$  (Rhodopsin) is found at multiple major phosphorylation sites (see Okada et al. 2001 for details) in Rhodopsin.

Mutating D at two positions has been shown to affect glutamate binding of glutamate receptor GPCRs (Jingami et al., 2003). D is common in  $AFS_1$ (Glutamate) and  $AFS_2$ (Glutamate) and has highest Shapley value in  $AFS_1$ .

E and D common in  $AFS_1$ (Aminergic) and  $AFS_2$ (Aminergic) are present at binding sites of important ligands (like histamine/serotonin) of aminergic receptors (Vass et al., 2019).

**Logical consistencies in** AFS **of GPCRs** (refer to Table 1 (GPCRs) for  $AFS_1$ ,  $AFS_2$ ,  $AFS_3$ ):

- $AFS_1 \cap AFS_2 = \{D\}$ , is excluded from  $AFS_3$ .
- $AFS_2 \cap AFS_3 = \{W, H, V\}$ , is excluded from  $AFS_1$ .
- $AFS_3 \cap AFS_1 = \{Q, E, M\}$ , is excluded from  $AFS_2$ .

**Logical consistencies in** AFS **of Rhodopsin-like GPCR subfamilies** (refer to Table 1 (Rhodopsin-like GPCRs) for  $AFS_1, AFS_2, AFS_3$ ):

- $AFS_1 \cap AFS_2 = \{L, E, F, M, D\}$ , is excluded from  $AFS_3$ .
- $AFS_2 \cap AFS_3 = \{R, V\}$ , is excluded from  $AFS_1$ .
- $AFS_3 \cap AFS_1 = \{P, W\}$  is excluded from  $AFS_2$ .

The explanations for these consistencies are similar to that in globins (Section 3.1.7).

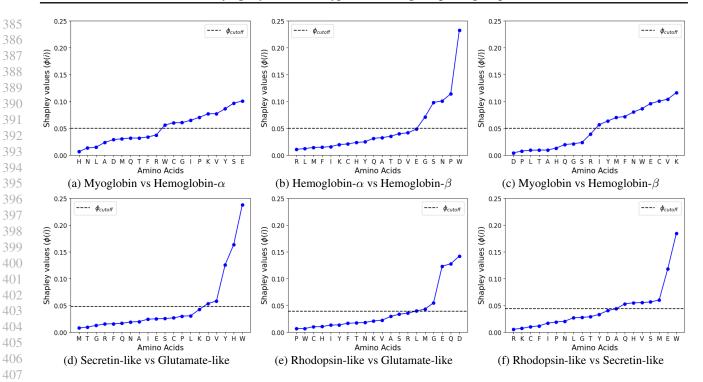


Figure 3: Shapley value ( $\phi(i)$ ) for AAC features computed using SVEA. See Appendix Figure E6 for remaining paralogs.

### **3.2. Validation of** *AFS* **using test data**

408 409 410

411

412

413

414

415

416

417

418

419

420

421

422

431

432

The classification scores on test data for the classifiers trained using AAC and AFS features, respectively, are reported in Appendix Table E5. Using AFS features, the test AM scores are at least 70%. For 13 of 15 paralog pairs, the scores are greater than 83%, and for 8 of 15 paralog pairs, it is greater than 90%. Details of the test data are provided in Appendix Section A.1.

# **3.3. Marginal contribution feature importance (MCI) of** *AFS*

423 For an AFS of size d, the top-d amino acids ranked by MCI 424 differ with AFS only in at the most two amino acids. For 8 425 of 15 datasets, AFS and top-d MCI sets are the same, while 426 only for two datasets do they differ in two amino acids. For 427 all 15 datasets, at least the top-3 MCI amino acids are in 428 AFS. For 11 of these datasets, at least the top-5 MCI amino 429 acids are in AFS. (Appendix Table E6) 430

## 4. Conclusion

433 We demonstrated an ML pipeline to identify the key amino 434 acid types, AFS, that distinguish a pair of paralogous pro-435 teins. The role of AFS in functionally distinguishing the 436 paralog pairs was validated using various sources of domain 437 knowledge. The robustness of this approach, as demon-438 strated by considering a diverse set of paralogous protein 439 pairs, illustrates its wider applicability. Identification of AFS can be used as an initial data-driven step before doing more detailed experimental investigations, like site-directed mutagenesis (Bachman, 2013) resolving sequence-function relationship. As the size of AFS is small (5-10 amino acids of 20), significantly less number of mutations can be tried.

As our pipeline works without using the sequence order information of the amino acids in the protein, it posits an interesting question to biologists : how amino acid composition by itself is able to distinguish paralogs given ample evidence that 3D structure and function are conserved despite sequence divergence (Lau et al., 2015)! Notably, amino acids in the AFS typically occur more than once in the sequence, but our method is silent on the specific positions where the amino acid has a functionally distinguishing role. This may be addressed by engineering features that incorporate sequence order information from the protein. However, these features can be very high-dimensional, for example,  $20^k$ -dimensional for k-mer features. The Monte Carlo based approximation algorithm for Shapley values would require exponentially more sampling (in number of features) for good approximations.

## 440 Impact Statement

This paper presents a computationally efficient data lean
ML pipeline. It can be used by biologists to decide whether
they should invest valuable resources (skilled manpower,
time, funds, etc.) for performing wet-lab experiments to
determine amino acid(s) that are critical for functional differentiation of paralogous proteins.

## 449 **References**

448

474

475

476

477

478 479

480

481

482

483

484

- 450 Bachman, J. Chapter ninteen - site-directed mutage-451 nesis. In Lorsch, J. (ed.), Laboratory Methods in 452 Enzymology: DNA, volume 529 of Methods in En-453 zymology, pp. 241–248. Academic Press, 2013. doi: 454 10.1016/B978-0-12-418687-3.00019-7. URL https: 455 //www.sciencedirect.com/science/ 456 article/pii/B9780124186873000197. 457
- 458 Begum, K., Mohl, J. E., Ayivor, F., Perez, E. E., and 459 Leung, M.-Y. GPCR-PEnDB: a database of pro-460 tein sequences and derived features to facilitate pre-461 diction and classification of G protein-coupled recep-462 tors. Database, 2020, 11 2020. ISSN 1758-0463. doi: 463 10.1093/database/baaa087. URL https://doi.org/ 464 10.1093/database/baaa087. 465
- Bileschi, M. L., Belanger, D., Bryant, D. H., Sanderson, T., Carter, B., Sculley, D., Bateman, A., DePristo, M. A., and Colwell, L. J. Using deep learning to annotate the protein universe. *Nature biotechnology*, 40(6):932–937, June 2022. ISSN 1087-0156. doi: 10.1038/s41587-021-01179-w. URL https://doi.org/10.1038/s41587-021-01179-w.
  - Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M. The balanced accuracy and its posterior distribution. In 2010 20th ICPR, pp. 3121–3124, 2010. doi: 10.1109/ ICPR.2010.764.
  - Cary, B. P., Zhang, X., Cao, J., Johnson, R. M., Piper, S. J., Gerrard, E. J., Wootten, D., and Sexton, P. M. New Insights into the Structure and Function of Class B1 GPCRs. *Endocrine Reviews*, 44(3):492–517, 12 2022. ISSN 0163-769X. doi: 10.1210/endrev/bnac033. URL https: //doi.org/10.1210/endrev/bnac033.
- 486 Castro, J., Gómez, D., and Tejada, J. Polynomial 487 calculation of the shapley value based on sampling. 488 Computers & Operations Research, 36(5):1726–1730, 489 2009. ISSN 0305-0548. doi: 10.1016/j.cor.2008.04.004. 490 URL https://www.sciencedirect.com/ 491 science/article/pii/S0305054808000804. 492 Selected papers presented at the Tenth International 493 Symposium on Locational Decisions (ISOLDE X). 494

- Catav, A., Fu, B., Zoabi, Y., Meilik, A. L. W., Shomron, N., Ernst, J., Sankararaman, S., and Gilad-Bachrach, R. Marginal contribution feature importance - an axiomatic approach for explaining data. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference* on Machine Learning, volume 139 of Proceedings of Machine Learning Research, pp. 1324–1335. PMLR, 18– 24 Jul 2021. URL https://proceedings.mlr. press/v139/catav21a.html.
- Covert, I., Lundberg, S. M., and Lee, S.-I. Understanding global feature contributions with additive importance measures. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), Advances in Neural Information Processing Systems, volume 33, pp. 17212–17223. Curran Associates, Inc., 2020. URL https://proceedings.neurips. cc/paper\_files/paper/2020/file/ c7bf0b7c1a86d5eb3be2c722cf2cf746-Paper. pdf.
- Denisov, I. G., Grinkova, Y. V., Nandigrami, P., Shekhar, M., Tajkhorshid, E., and Sligar, S. G. Allosteric interactions in human cytochrome p450 cyp3a4: The role of phenylalanine 213. *Biochemistry*, 58(10):1411–1422, 2019. doi: 10.1021/acs.biochem.8b01268. URL https://doi. org/10.1021/acs.biochem.8b01268. PMID: 30785734.
- Dill, K., Jernigan, R., and Bahar, I. Protein Actions: Principles and Modeling. CRC Press, 2017. ISBN 9781351815000. URL https://books.google. co.in/books?id=NHs2DwAAQBAJ.
- Dodson, G. and Wlodawer, A. Catalytic triads and their relatives. Trends in Biochemical Sciences, 23(9):347-352, 1998. ISSN 0968-0004. doi: https://doi.org/10.1016/S0968-0004(98)01254-7. URL https://www.sciencedirect.com/ science/article/pii/S0968000498012547.
- Dutta, S., Akey, I. V., Dingwall, C., Hartman, K. L., Laue, T., Nolte, R. T., Head, J. F., and Akey, C. W. The crystal structure of nucleoplasmin-core: Implications for histone binding and nucleosome assembly. *Molecular Cell*, 8(4):841–853, 2001. ISSN 1097-2765. doi: 10.1016/S1097-2765(01)00354-9. URL https://www.sciencedirect.com/ science/article/pii/S1097276501003549.
- Edgar, R. C. and Batzoglou, S. Multiple sequence alignment. *Current Opinion in Structural Biology*, 16(3):368–373, 2006. ISSN 0959-440X. doi: 10.1016/j.sbi.2006.04.004. URL https://www.sciencedirect.com/ science/article/pii/S0959440X06000704. Nucleic acids/Sequences and topology.

- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. CD-495 496 HIT: accelerated for clustering the next-generation se-497 quencing data. Bioinformatics, 28(23):3150-3152, 498 10 2012. ISSN 1367-4803. doi: 10.1093/ 499 bioinformatics/bts565. URL https://doi.org/10. 500 1093/bioinformatics/bts565.
- 501 502 Galozzi, P., Bindoli, S., Doria, A., and Sfriso, P. The revisited role of interleukin-1 alpha and beta in 503 autoimmune and inflammatory disorders and in comor-504 bidities. Autoimmunity Reviews, 20(4):102785, 2021. 505 ISSN 1568-9972. doi: 10.1016/j.autrev.2021.102785. 506 507 URL https://www.sciencedirect.com/ science/article/pii/S1568997221000483. 508
- 509 Hargrove, T. Y., Kim, K., de Nazaré Correia Soeiro, M., da 510 Silva, C. F., da Gama Jaen Batista, D., Batista, M. M., 511 Yazlovitskaya, E. M., Waterman, M. R., Sulikowski, 512 G. A., and Lepesheva, G. I. Cyp51 structures and 513 structure-based development of novel, pathogen-specific 514 inhibitory scaffolds. International Journal for Para-515 sitology: Drugs and Drug Resistance, 2:178–186, 2012. 516 ISSN 2211-3207. doi: 10.1016/j.ijpddr.2012.06.001. 517 URL https://www.sciencedirect.com/ 518 science/article/pii/S2211320712000206. 519 Including Articles from Keystone Symposium on "Drug 520 Discovery for Protozoan Parasites"; pp. 230-270. 521 522
- Harmar, A. Family-b g-protein-coupled receptors. *Genome biology*, 2(12):REVIEWS3013, 2001. ISSN 1474-7596.
  doi: 10.1186/gb-2001-2-12-reviews3013. URL https:
  //europepmc.org/articles/PMC138994.

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

- Hedstrom, L. Serine protease mechanism and specificity. *Chemical Reviews*, 102(12):4501–4524, 2002. doi: 10.1021/cr000033x. URL https://doi.org/10. 1021/cr000033x. PMID: 12475199.
- Hedstrom, L., Perona, J. J., and Rutter, W. J. Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant. *Biochemistry*, 33 29:8757–63, 1994.
- Jingami, H., Nakanishi, S., and Morikawa, K. Structure of the metabotropic glutamate receptor. *Current Opinion in Neurobiology*, 13(3):271–278, 2003. ISSN 0959-4388. doi: 10.1016/S0959-4388(03)00067-9. URL https://www.sciencedirect.com/ science/article/pii/S0959438803000679.
- Kan, H.-I., Chen, I.-Y., Zulfajri, M., and Wang, C. C. Subunit disassembly pathway of human hemoglobin revealing the site-specific role of its cysteine residues. *The Journal of Physical Chemistry B*, 117(34):9831–9839, 2013. doi: 10.1021/jp402292b. URL https://doi. org/10.1021/jp402292b. PMID: 23902424.

- Kresge, N., Simoni, R. D., and Hill, R. L. The development of site-directed mutagenesis by michael smith. *Journal* of Biological Chemistry, 281(39):e31–e33, 2006. ISSN 0021-9258. doi: 10.1016/S0021-9258(19)33938-9. URL https://www.sciencedirect.com/ science/article/pii/S0021925819339389.
- Lau, C. K., Turner, L., Jespersen, J. S., Lowe, E. D., Petersen, B., Wang, C. W., Petersen, J. E., Lusingu, J., Theander, T. G., Lavstsen, T., and Higgins, M. K. Structural conservation despite huge sequence diversity allows epcr binding by the pfemp1 family implicated in severe childhood malaria. *Cell Host & Microbe*, 17(1):118–129, 2015. ISSN 1931-3128. doi: 10.1016/j.chom.2014.11. 007. URL https://www.sciencedirect.com/science/article/pii/S1931312814004235.
- Lepesheva, G. I. and Waterman, M. R. Cyp51the omnipotent p450. Molecular and Cellular Endocrinology, 215(1):165-170, 2004. ISSN 0303-7207. doi: 10.1016/j.mce.2003.11.016. URL https://www.sciencedirect.com/ science/article/pii/S0303720703005148. Proceedings of the Serono Foundation for the Advancement of Medical Science Workshop on Molecular Steroidogenesis.
- Lepesheva, G. I. and Waterman, M. R. Sterol 14αdemethylase cytochrome p450 (cyp51), a p450 in all biological kingdoms. *Biochimica et Biophysica Acta* (*BBA*) - *General Subjects*, 1770(3):467–477, 2007. ISSN 0304-4165. doi: 10.1016/j.bbagen.2006.07.018. URL https://www.sciencedirect.com/ science/article/pii/S0304416506002145. P450.
- Lundberg, S. M. and Lee, S.-I. A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips. cc/paper\_files/paper/2017/file/ 8a20a8621978632d76c43dfd28b67767-Paper. pdf.
- McArthur, A. G., Hegelund, T., Cox, R. L., Stegeman, J. J., Liljenberg, M., Olsson, U., Sundberg, P., and Celander, M. C. Phylogenetic Analysis of the Cytochrome P450 3 (CYP3) Gene Family. *Journal of Molecular Evolution*, 57(2):200– 211, August 2003. doi: 10.1007/s00239-003-2466-x. URL https://link.springer.com/article/ 10.1007/s00239-003-2466-x.
- Menon, A. K., Narasimhan, H., Agarwal, S., and Chawla, S. On the statistical consistency of algorithms for bi-

nary classification under class imbalance. In *Proceed- ings of the 30th ICML - Volume 28*, ICML'13, pp.

- 552 III-603-III-611. JMLR.org, 2013.
- 553 Mühlethaler, T., Gioia, D., Prota, A. E., Sharpe, 554 M. E., Cavalli, A., and Steinmetz, M. O. Com-555 prehensive analysis of binding sites in tubulin. 556 Angewandte Chemie International Edition, 60(24): 557 13331-13342, 2021. doi: 10.1002/anie.202100273. URL 558 https://onlinelibrary.wiley.com/doi/ 559 abs/10.1002/anie.202100273. 560
- 561 Narahari, Y. Game Theory and Mechanism De-562 sign. WORLD SCIENTIFIC / INDIAN INST 563 OF SCIENCE, INDIA, 2014. doi: 10.1142/ 564 8902. URL https://www.worldscientific. 565 com/doi/abs/10.1142/8902. 566
- Nitahara, Y., Kishimoto, K., Yabusaki, Y., Gotoh, O.,
  Yoshida, Y., Horiuchi, T., and Aoyama, Y. The amino acid
  residues affecting the activity and azole susceptibility of
  rat cyp51 (sterol 14-demethylase p450). *The Journal of Biochemistry*, 129(5):761–768, 2001.
- 573 Okada, T., Ernst, O. P., Palczewski, K., and Hofmann, 574 K. P. Activation of rhodopsin: new insights from 575 structural and biochemical studies. Trends in Bio-576 chemical Sciences, 26(5):318-324, 2001. ISSN 577 0968-0004. doi: 10.1016/S0968-0004(01)01799-6. 578 URL https://www.sciencedirect.com/ 579 science/article/pii/S0968000401017996. 580
- 581 Permyakov, E. A. α-actalbumin, Amazing Calcium-Binding
   582 Protein. *Biomolecules*, 10(9):1210, Aug 2020. ISSN
   583 2218-273X. doi: 10.3390/biom10091210. URL http:
   584 //dx.doi.org/10.3390/biom10091210.

- Permyakov, E. A. and Berliner, L. J. α-Lactalbumin: structure and function. *FEBS Letters*, 473(3):269–274, 2000.
   ISSN 0014-5793. doi: 10.1016/S0014-5793(00)01546-5.
   URL https://www.sciencedirect.com/science/article/pii/S0014579300015465.
- 591 Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, 592 E. C., Couch, G. S., Croll, T. I., Morris, J. H., and 593 Ferrin, T. E. Ucsf chimerax: Structure visualiza-594 tion for researchers, educators, and developers. Pro-595 tein Science, 30(1):70-82, 2021. doi: 10.1002/pro. 596 3943. URL https://onlinelibrary.wiley. 597 com/doi/abs/10.1002/pro.3943. 598
- Qasba, P. K., Kumar, S., and Brew, D. K. Molecular divergence of lysozymes and α-lactalbumin. *Critical Reviews in Biochemistry and Molecular Biology*, 32(4):255–306, 1997. doi: 10.3109/10409239709082574. URL https://doi.org/10.3109/10409239709082574.

- Qiu, H., Taudien, S., Herlyn, H., Schmitz, J., Zhou, Y., Chen, G., Roberto, R., Rocchi, M., Platzer, M., and Wojnowski, L. Cyp3 phylogenomics: evidence for positive selection of cyp3a4 and cyp3a7. *Pharmacogenetics and Genomics*, 18(1):53—66, January 2008. ISSN 1744-6872. doi: 10.1097/fpc.0b013e3282f313f8. URL https:// doi.org/10.1097/FPC.0b013e3282f313f8.
- Rozemberczki, B., Watson, L., Bayer, P., Yang, H.-T., Kiss, O., Nilsson, S., and Sarkar, R. The shapley value in machine learning. In Raedt, L. D. (ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 5572–5579. International Joint Conferences on Artificial Intelligence Organization, 7 2022. doi: 10.24963/ijcai.2022/778. URL https://doi.org/10.24963/ijcai.2022/778. Survey Track.
- Sakmar, T. P., Menon, S. T., Marin, E. P., and Awad, E. S. Rhodopsin: Insights from recent structural studies. Annual Review of Biophysics, 31(Volume 31, 2002):443–484, 2002. ISSN 1936-1238. doi: 10.1146/annurev.biophys.31.082901.134348. URL https://www.annualreviews.org/ content/journals/10.1146/annurev. biophys.31.082901.134348.
- Sanderson, T., Bileschi, M. L., Belanger, D., and Colwell, L. J. Proteinfer, deep neural networks for protein functional inference. *eLife*, 12:e80942, feb 2023. ISSN 2050-084X. doi: 10.7554/eLife.80942. URL https://doi.org/10.7554/eLife.80942.
- Shapley, L. S. 17. A Value for n-Person Games, pp. 307–318. Princeton University Press, Princeton, 1953. ISBN 9781400881970. doi: doi:10.1515/ 9781400881970-018. URL https://doi.org/10. 1515/9781400881970-018.
- Shionyu, M., Takahashi, K., and Gō, M. Variable subunit contact and cooperativity of hemoglobins. J. Mol. Evol., 53(4-5):416–429, October 2001.
- Song, H., Bremer, B. J., Hinds, E. C., Raskutti, G., and Romero, P. A. Inferring protein sequencefunction relationships with large-scale positive-unlabeled learning. *Cell Systems*, 12(1):92–101.e8, 2021. ISSN 2405-4712. doi: 10.1016/j.cels.2020.10. 007. URL https://www.sciencedirect.com/ science/article/pii/S2405471220304142.
- Steinwart, I. and Christmann, A. Support Vector Machines. Springer Publishing Company, Incorporated, 1st edition, 2008. ISBN 0387772413.
- Strushkevich, N., Usanov, S. A., and Park, H.-W. Structural basis of human cyp51 inhibition by antifungal

- azoles. Journal of Molecular Biology, 397(4):1067-1078, 605 606 2010. ISSN 0022-2836. doi: 10.1016/j.jmb.2010.01. 607 075. URL https://www.sciencedirect.com/ 608 science/article/pii/S0022283610001324. 609 The UniProt Consortium. UniProt: the universal protein 610 knowledgebase in 2021. Nucleic Acids Research, 49(D1): 611 D480-D489, 11 2020. ISSN 0305-1048. doi: 10.1093/ 612 nar/gkaa1100. URL https://doi.org/10.1093/ 613 nar/gkaa1100. 614 615 Tripathi, S., Hemachandra, N., and Trivedi, P. Interpretable 616 feature subset selection: A Shapley value based approach. 617 In 2020 IEEE BigData, pp. 5463-5472, 2020. doi: 10. 618 1109/BigData50022.2020.9378102. 619 620 Tripathi, S., Hemachandra, N., and Trivedi, P. Inter-621 pretable feature subset selection: A shapley value based 622 approach, 2021. URL https://arxiv.org/abs/ 623 2001.03956. 624 625 Vass, M., Podlewska, S., de Esch, I. J. P., Bojarski, 626 A. J., Leurs, R., Kooistra, A. J., and de Graaf, C. 627 Aminergic gpcr-ligand interactions: A chemical and 628 structural map of receptor mutation data. Journal of 629 Medicinal Chemistry, 62(8):3784-3839, 2019. doi: 630 10.1021/acs.jmedchem.8b00836. URL https://doi. 631 org/10.1021/acs.jmedchem.8b00836. PMID: 632 30351004. 633 634 Veerapandian, B., Gilliland, G. L., Raag, R., Svens-
- son, A. L., Masui, Y., Hirai, Y., and Poulos, 635 636 T. L. Functional implications of interleukin-1 $\beta$ based on the three-dimensional structure. Proteins: 637 638 Structure, Function, and Bioinformatics, 12(1): 10-23, 1992. doi: 10.1002/prot.340120103. URL 639 640 https://onlinelibrary.wiley.com/doi/ 641 abs/10.1002/prot.340120103. 642
- Yang, Y., Xu, T., Conant, G., Kishino, H., Thorne, J. L.,
  and Ji, X. Interlocus gene conversion, natural selection,
  and paralog homogenization. *Molecular Biology and Evolution*, 40(9):msad198, 09 2023. ISSN 1537-1719.
  doi: 10.1093/molbev/msad198. URL https://doi.
  org/10.1093/molbev/msad198.
- 649
  650
  651
  651
  651
  652
  653
  653
  654
  655
  7276. doi: 10.1007/BF01769885. URL https://doi.org/10.1007/BF01769885.
- Zhang, Y., Wang, Z., Wang, Y., Jin, W., Zhang, Z., Jin, L.,
  Qian, J., and Zheng, L. Cyp3a4 and cyp3a5: the crucial roles in clinical drug metabolism and the significant implications of genetic polymorphisms. *PeerJ*, 12:e18636, 2024.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. Learning Deep Features for Discriminative Localization . In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2921–2929, Los Alamitos, CA, USA, June 2016. IEEE Computer Society. doi: 10.1109/CVPR.2016.319. URL https://doi.ieeecomputersociety.org/ 10.1109/CVPR.2016.319.

## A. Data collection and code

We discuss the details of the data collection procedure for the datasets used in our computational experiments.

## A.1. Datasets of 15 paralog pairs

We apply our method for identifying amino acid types that distinguish paralogous proteins using the datasets described in Table A2. Only the train set is used for computing AFS, while the test set is used for computing classification scores for the linear SVM trained using the train set.

Table A2: The number of sequences in the train and test sets of the protein families considered in computational experiments.

Family	Train (Swiss-Prot)	Test (TrEMBL)
Lysozyme-like		
$\alpha$ -Lactalbumin	22	53
Lysozyme C	74	14
Trypsin-like		
Trypsin	66	3813
Chymotrypsin	17	281
Tubulin		
$\alpha$	117	190
$\beta$	191	347
Histone		
H2A	180	16599
H2B	177	7599
Interleukin-1		
α	16	12
$\beta$	25	194
Cytochrome P450		
CYP3	32	818
CYP51	32	601
Globins		
Myoglobin	107	479
Hemoglobin- $\alpha$	303	525
Hemoglobin- $\beta$	285	261
	(GPCR-P	EnDB)
	Train (80%)	Test (20%)
GPCR families		
Rhodopsin-like	181	45
Lipid receptors	113	28
Peptide receptors	367	92
Aminergic receptors	186	47
Glutamate-like	89	23
Secretin-like	90	23

All datasets are taken from publicly available databases (UniProt (The UniProt Consortium, 2020) and GPCR-PEnDB (Begum et al., 2020)). Well-known pairs of paralogous proteins were curated from millions of sequences from UniProt considering the number of sequences and manually reviewed labels available for them.

For all datasets except GPCR, we use manually curated Swiss-Prot sequences for training and electronically annotated TrEMBL sequences for testing. These proteins have very specific functions. In contrast, GPCRs are a large and diverse group of transmembrane proteins that mediate cellular responses to extracellular signals. We chose to use an already curated dataset in this case. For each of the GPCR families considered (Table A2), the sequences are randomly split as 80%-train/20%-test. The use of GPCR-PEnDB data is to illustrate the effectiveness of our method with random slicing, which is inevitable when additional curated data are not available. If one or many UniProt entries in a dataset had identical sequences, then only one of them was retained, and the remaining were deleted.

Identifying key amino acid types that distinguish paralogous proteins

715	The following queries were used for collecting data from UniProt (The UniProt Consortium, 2020),
716	• lysozyme C: (protein_name:"lysozyme C") AND (fragment:false) NOT (existence:4) NOT
717	(existence:5) AND (length:[* TO 200]) AND (ec:3.2.1.17) AND (xref:cazy-GH22) AND
718	(reviewed:true)
719	• $\alpha$ -lactalbumin: (protein_name:"alpha lactalbumin") AND (fragment:false) NOT
720	(existence:4) NOT (existence:5) AND (length:[* TO 200]) AND (reviewed:true)
721	-
722	• myoglobin: (protein_name:"myoglobin") AND (xref:interpro-IPR002335) AND
723	(fragment:false) NOT (existence:5) NOT (existence:4)
724	• hemoglobin-α: (protein_name:"hemoglobin alpha") AND (xref:interpro-IPR002338) AND
725	(fragment:false) NOT (existence:5) NOT (existence:4)
726	• hemoglobin- $eta$ : (protein_name:"hemoglobin beta") AND (xref:interpro-IPR002337) AND
727	(fragment:false) NOT (existence:5) NOT (existence:4)
728	• trypsin: (protein_name:trypsin) AND (fragment:false) AND (ec:3.4.21.4) NOT
729	(existence:5)
730	• chymotrypsin: (protein_name:chymotrypsin) AND (fragment:false) AND (ec:3.4.21.1) NOT
731	(existence:5)
732	• <b>tubulin-</b> $lpha$ : (protein_name:"tubulin alpha") AND (family:"tubulin family") AND
733	(length:[300 TO 600]) AND (fragment:false) NOT (annotation_score:1) NOT
734	(annotation_score:2)
735	• tubulin- $\beta$ : (protein_name:"tubulin beta") AND (family:"tubulin family") AND
736	(length:[300 TO 600]) AND (fragment:false) NOT (annotation_score:1) NOT
737	(annotation_score:2)
738 739	• interleukin-1 $\alpha$ (protein_name:"interleukin-1 alpha") AND (family:il-1) AND
740	(fragment:false) NOT (existence:4) NOT (existence:5) AND (length:[200 TO 400]) NOT
740	(annotation_score:1)
742	• interleukin-1 $\beta$ : (protein_name:"interleukin-1 beta") AND (family:il-1) AND
743	(fragment:false) NOT (existence:4) NOT (existence:5) AND (length:[200 TO 400]) NOT
744	(annotation_score:1)
745	
746	• Histone H2A: (protein_name: "histone h2a") AND (family:histone) AND (fragment:false) NOT (existence:4) NOT (existence:5) AND (length:[* TO 200])
747	• Histone H2B: (protein_name:"histone h2b") AND (family:histone) AND (fragment:false)
748	NOT (existence:4) NOT (existence:5) AND (length:[* TO 200])
749	• Cytochrome P450 CYP3: (family:"Cytochrome P450") AND ((gene:cyp3) OR
750	(gene:cyp3A*)) AND (fragment:false) NOT (existence:4) NOT (existence:5) NOT
751	(annotation_score:1)
752 753	• Cytochrome P450 CYP51: (family:"Cytochrome P450") AND ((gene:cyp51) OR
	(gene:cyp51A*) OR (gene:cyp51B*) OR (gene:cyp51C*)) AND (fragment:false) NOT
754 755	(existence:4) NOT (existence:5) NOT (annotation_score:1)
756	The GPCR sequences were collected from the GPCR-PEn database (URL: https://gpcr.utep.edu/) (Begum et al.,
757	2020). Sequence redundancy of the rhodopsin-like family was reduced using CD-hit (Fu et al., 2012) with 30% sequence
758	similarity cutoff.
759	Similarly Outon.
760	
761	A.2. Code
762	

The code to reproduce the computational experiments is available at https://anonymous.4open.science/r/ AFS\_AAC\_SVM-F3D9. Protein sequences used in the computational experiments along with their UniProt IDs, are provided in the datasets folder as .csv files for each family.

- 766
- 767
- 768 769

## 770 B. Sequence and function diversity of protein classes within a dataset

Paralogous proteins have a common ancestor but have diverged in functionality. Protein functions are an aggregate of descriptors describing protein's activity and influence at various levels. They can be at the molecular level, like binding with specific molecules and catalysing reactions, to the biological process level, like energy metabolism. In B.1, we discuss the diversity of the functions of the proteins considered in our datasets.

As paralogs have a common ancestor, high sequence similarity would suggest high evolutionary conservation in the proteins. In B.2, we discuss the extent of sequence diversity in protein classes considered in our datasets.

We see that the dataset of proteins considered in our computational experiments are diverse in their function and sequences.

## 780781 **B.1. Function diversity**

773

774

775 776

777

We have considered paralogous proteins with varying functional differences. We find very subtle differences in the functions of trypsin and chymotrypsin. On the other hand, the function difference is drastic in the case of alpha-lactalbumin and lysozyme c.

Trypsin and chymotrypsin are a family of enzymes that break peptide bonds in proteins. The difference in the function of these proteins is fine-grained; trypsins cleave only the peptide bond following a basic amino acid (K and R), while chymotrypsins cleave the peptide bond following a hydrophobic amino acid (F, W, and Y) (Dodson & Wlodawer, 1998).

GPCRs constitute a large and diverse class of cell surface receptor proteins. They trigger intra-cellular pathways in response
 to external signals. These signals are in the form of small molecules, called ligands. Depending upon the nature of
 ligands and other 3D structural similarities, GPCRs are grouped into distinct classes. We consider three such classes viz.,
 rhodopsin-like, secretin-like, and glutamate-like. Further, we consider pairwise three subfamilies of rhodopsin-like GPCRs
 viz., aminergic receptors, lipid receptors, and peptide receptors.

Lysozyme C and  $\alpha$ -lactalbumin are sequence and structure homologs with mutually exclusive functions and high fold conservation. Based on phylogenetic analysis, they are considered to have diverged from a common ancestor millions of years ago (Qasba et al., 1997).

Globins are a superfamily of functionally divergent homologous protein families with a high level of fold conservation. We consider three well-known globin families viz., myoglobin, hemoglobin- $\alpha$  and hemoglobin- $\beta$ . Myoglobin is a monomer that binds and releases oxygen as per physiological requirements. On the other hand,  $\alpha$  and  $\beta$  chains together constitute hemoglobin, a tetramer of composition  $\alpha_2\beta_2$  (Dill et al., 2017), that transports oxygen in red blood cells.

Tubulin- $\alpha$  and tubulin- $\beta$  are similar to the hemoglobin- $\alpha$  and hemoglobin- $\beta$  pair in that they both share sequence and 3D structural similarities but have subtle functional differences. One copy each of tubulin- $\alpha$  and tubulin- $\beta$  form a functional dimer. Notably, neither two copies of tubulin- $\alpha$  nor two copies of tubulin- $\beta$  can form a functional dimer. Tubulin- $\beta$  has a catalytic activity (GTP hydrolysis) that is absent in tubulin- $\alpha$ . This is one of the several subtle functional differences between tubulin- $\alpha$  and tubulin- $\beta$ .

Interleukin-1 alpha and interleukin-1 beta are both proteins involved in the immune system. They differ from each other in their occurrence within the body (on cell surface or in blood circulation), activation mechanisms, and associated signalling pathways (Galozzi et al., 2021).

Cytochrome P450 (abbreviated as CYP) is a family of proteins whose function is clearance of 'foreign' molecules (drugs; also called as xenobiotics) as well as in certain biosynthesis pathways e.g., of steroid hormones. CYP3 and CYP51 are two of the several classes of CYPs; CYP3 metabolizes lipophilic molecules (McArthur et al., 2003) whereas CYP51 is involved in steroid biosynthesis (Hargrove et al., 2012).

<sup>816</sup> Hemoglobin- $\alpha$ /hemoglobin- $\beta$ , histone H2A / histone H2B and tubulin- $\alpha$ /tubulin- $\beta$  are paralog pairs that together function as heteromers (protein complexes made up of different protein subunits).

# 819820**B.2. Sequence Diversity**

The dataset of the 15 paralog pairs in our experiments comprises 21 protein families (Table A2). For these families, we compute the within-class sequence similarities (for sequences within a protein family). We also compute the inter-class sequence similarities (between sequences from two different protein families) for each paralog pair. These are shown in Appendix Figure B4. We use a longest subsequence based similarity score, *lcss*, that is defined in B.2.1. In B.2.2, we see that *lcss* significantly varies across the 21 protein families we are considering as compared to its variation between the two protein sequences of any paralog pair.

### 829 B.2.1. LONGEST COMMON SUBSEQUENCE BASED SIMILARITY SCORE (*lcss*)

We compute the longest common subsequence (lcs) based similarity score (lcss) between a pair of protein sequences. We define lcss between two sequences as the length of their longest common subsequence, lcs, divided by the length of the longest sequence from the two. For a pair of protein sequences,  $\mathbf{p}^{(i)} = (p_1^{(i)}, p_2^{(i)}, \dots, p_{L_1}^{(i)})$  of length  $L_1$  and  $\mathbf{p}^{(j)} = (p_1^{(j)}, p_2^{(j)}, \dots, p_{L_2}^{(j)})$  of length  $L_2$ , their lcss is,

$$lcs(\mathbf{p}^{(i)}, \mathbf{p}^{(j)}) = \max_{\mathbf{q}} k$$
  
s.t.  $\mathbf{q} = (q_1, q_2, \dots, q_k)$   
 $(q_1 = p_{x_1}^{(i)} = p_{y_1}^{(j)}, q_2 = p_{x_2}^{(i)} = p_{y_2}^{(j)}, \dots, q_k = p_{x_k}^{(i)} = p_{y_k}^{(j)})$   
 $x_1 < x_2 < \dots < x_k$   
 $y_1 < y_2 < \dots < y_k$ 

lcs based similarity score, lcss, is defined as,

828

850

851 852

853 854

855

856

857

865 866

867

868

869

877 878

879

$$lcss(\mathbf{p}^{(i)}, \mathbf{p}^{(j)}) = \frac{lcs(\mathbf{p}^{(i)}, \mathbf{p}^{(j)})}{\max(L_1, L_2)} \in [0, 1]$$

 $lcss(\mathbf{p}^{(i)}, \mathbf{p}^{(j)}) = 1$  if and only if  $\mathbf{p}^{(i)} = \mathbf{p}^{(j)}$ , i.e., sequences are identical. Whereas  $lcss(\mathbf{p}^{(i)}, \mathbf{p}^{(j)}) = 0$  if and only if  $p_x^{(i)} \neq p_y^{(j)}, \forall x, y$ , i.e., there are no amino acids common to both the sequences.

B.2.2. WITHIN-CLASS AND INTER-CLASS *lcss* FOR THE 15 PARALOG PAIRS

Within-class lcss:  $lcss(\mathbf{p}^{(i)}, \mathbf{p}^{(j)})$  are computed with  $\mathbf{p}^{(i)}, \mathbf{p}^{(j)}$  from the same protein family. These are shown in *blue* and *magenta* in Figure B4 (with box-plots) for each of 21 protein families in the 15 paralog pairs.

• 12 of 21 protein families have median within-class *lcss* greater than 0.5. This implies less sequence diversity in this set of families from the remaining families. These are,

	Fami	ly	$\alpha$ -la	ctalbumin	lysoz	yme C	myoglob	oin	hemoglob	in- $\alpha$	hemoglo	bin- $\beta$	tubulin- $\alpha$	
	Median	lcss		0.6	0	.59	0.81		0.63		0.6	7	0.83	
Fa	mily	tubuli	$n-\beta$	interleuki	n-1 $\alpha$	interlet	ukin-1 $\beta$	his	stone H2A	histo	one H2B	cytoch	nrome P450 C	CYP3
Medi	an <i>lcss</i>	0.8	2	0.72		0	.66		0.65		0.68		0.7	

Table B3: The median within-class *lcss* between sequences from the respective families. See boxplot in Figure B4.

- Median  $lcss \ge 0.6$  for 11 of these 12 families and  $\ge 0.8$  for 3 families (high level of sequence conservation).

- For 7 out of the 15 paralog pairs, the median within-class lcss > 0.5 for both families of a paralogous pair.

• For the remaining 9 protein families, the median within-class *lcss* is less than 0.5. This implies high sequence diversity in this set of families from the remaining families. These are,

F	amily	trypsin	chymotrypsin	rhodopsin-like rec	eptor	glutamate-l	ike receptor	secretin-like r	eceptor
Med	lian <i>lcss</i>	0.47	0.45	0.34		0.	35	0.36	
	Fa	mily	aminergic recepto	or lipid receptor	pept	ide receptor	cytochrome	P450 CYP51	
	Medi	an <i>lcss</i>	0.39	0.37		0.37	0	0.47	

Table B4: The median within-class *lcss* between sequences from the respective families. See boxplot in Figure B4.

- For 7 out of the 15 paralog pairs, the median within-class lcss < 0.5 for both families of a paralogous pair.

880 881	• For the paralog pair Cytochrome P450 CYP3 vs CYP51, the median sequence similarity for CYP3 is greater than 0.5, while for CYP51, it is less than 0.5.
882 883 884	<b>Inter-class</b> <i>lcss</i> : <i>lcss</i> ( $\mathbf{p}^{(i)}, \mathbf{p}^{(j)}$ ) are computed with $\mathbf{p}^{(i)}, \mathbf{p}^{(j)}$ respectively from two protein families that are paralog pairs. These are shown in <i>cyan</i> in Figure B4 (with box-plots) for each of the 15 paralog pairs.
885 886	• The median inter-class <i>lcss</i> is less than 0.5 for all paralog pairs. This implies sequences of the proteins across the classes are not very similar.
887 888 889	<b>Distinguishing paralog pairs based on within-class and inter-class</b> <i>lcss</i> : If we analyse the box plots in Figure B4 - two paralog pair proteins can be considered to be distinguishable based on sequence similarity if the upper-whisker of inter-class <i>lcss</i> is lower than the lower-whiskers of the respective within-class <i>lcss</i> scores.
890 891 892	• Apart from paralog pairs, tubulin- $\alpha$ vs tubulin- $\beta$ (Figure B4c) and interleukin-1 $\alpha$ vs interleukin-1 $\beta$ (Figure B4d), no other paralog pair is distinguishable based on sequence similarity.
893 894	• For Trypsin vs Chymotrypsin and the 6 GPCR pairs (Figures B4b and B4j to B4o), the median inter-class <i>lcss</i> scores are close to the within-class <i>lcss</i> scores making them indistinguishable based on sequence similarity.
895 896	
897 898	
899 900	
901	
902 903	
904 905	
906	
907 908	
909	
910 911	
912	
913 914	
915 916	
917	
918 919	
920	
921 922	
923 924	
925	
926 927	
928	
929 930	
931 932	
932 933	
934	

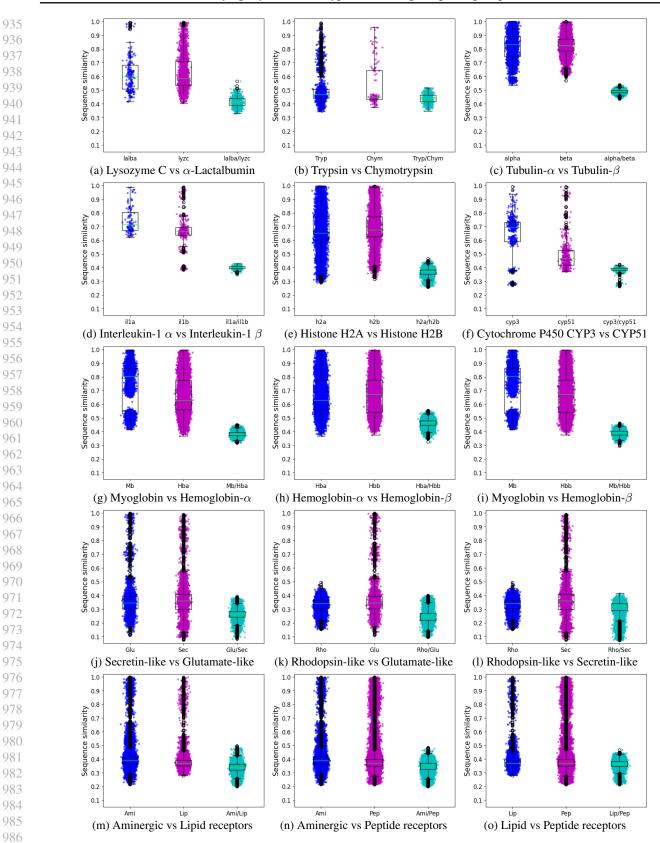


Figure B4: *lcss* sequence similarity scores for the 15 paralog pair datasets. In the boxplots, the lower and upper whiskers
are at 1.5 IQR (inter-quantile range) values away from the first and third quartiles respectively.

## **C. The SVEA algorithm for AFS**

<sup>992</sup> Algorithm 1  $\phi_i$  Monte-carlo approximation algorithm as suggested in (Tripathi et al., 2020; 2021)

**Input:** Feature set  $N = \{1, 2, \dots, 20\}$ , Number of sample permutations samPerm, Datasets  $(D_P, D_Q)$ , Set of coalitions  $Sam_co_set = [()]$ Initialise:  $v(()) = 0, \, \hat{\phi}_i := 0 \, \forall i \in N$ Append N to  $Sam\_co\_set$ . for  $s = 1, 2, \ldots, samPerm$  do Take  $\pi \in PermSet(N)$  with probability  $\frac{1}{20!}$ . for  $i = 1, 2, \ldots, 20$  do Compute  $Pred^{i}(\pi) = \{\pi(1), \pi(2), \dots, \pi(k-1) | i = \pi(k)\}$ if  $Pred^{i}(\pi)$  not in  $Sam_{co\_set}$  then Compute  $v(Pred^{i}(\pi)) = 1 - tr_er(Pred^{i}(\pi)).$ Append  $Pred^{i}(\pi)$  to  $Sam_{co\_set}$ . end if if  $Pred^{i}(\pi) \cup i$  not in  $Sam_{-}co_{-}set$  then 

Compute  $v(Pred^{i}(\pi) \cup \{i\}) = 1 - tr\_er(Pred^{i}(\pi) \cup \{i\})$ . Append  $Pred^{i}(\pi) \cup \{i\}$  to  $Sam\_co\_set$ . end if  $\hat{\phi}_{i} = \hat{\phi}_{i} + v(Pred^{i}(\pi) \cup \{i\}) - v(Pred^{i}(\pi)))$ 

 1010
 end for

 1011
 end for

 1012
 end for

 $\begin{array}{ll}1012\\1013\end{array}\quad \hat{\phi}_i = \frac{\hat{\phi}_i}{samPerm}, \forall i \in N\end{array}$ 

## **D. SVM training for** AFS partition

We provide details for the linear SVM classifier discussed in Section 2.3. We use 5-fold cross-validation to tune the SVM regularisation hyperparameter *C* from {0.1, 1, 10, 100, 1000} that gives the best average classification score for the 5 folds. *C* is inversely proportional to the strength of regularisation. In general, we find that there is an imbalance in the number of sequences that we find for the two paralogous proteins, i.e. say  $n_P >> n_Q$ . It is known that accuracy is not a well-suited performance measure of the classifier in class imbalance settings. Therefore, we use the arithmetic mean of sensitivity and specificity (AM) to measure the performance of the classifier (Brodersen et al., 2010). Further, we use a class-balanced version of hinge loss for training the SVM as suggested in (Menon et al., 2013) for statistical consistency with the AM score. Appendix Table E5 reports the train and test scores of the trained linear SVM with AAC and *AFS* features, respectively, on the protein family datasets (See Appendix Table A2) considered in our computational experiments.

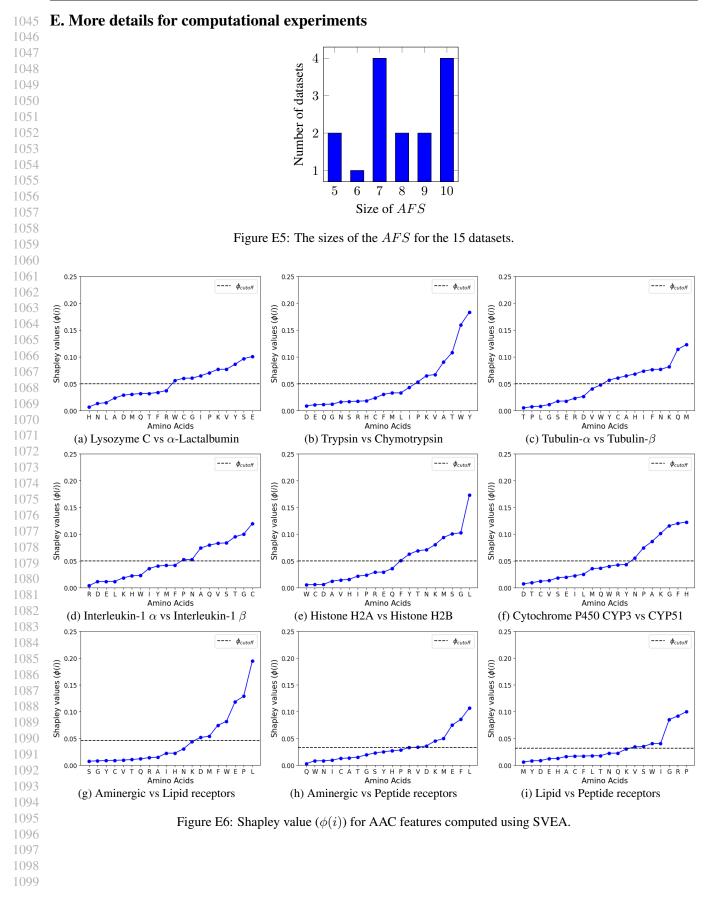


Table E5: Classification scores for different pairs of paralogous proteins using the train/test datasets described in Table A2, using AAC and AFS features. The AFS amino acids computed for each pair are given in Table 1. The train score is the mean ( $\pm 1$  standard deviation) 5-fold cross-validation score. AM is the arithmetic mean of specificity and sensitivity. Acc is the accuracy.

	(a) Ly	ysozyme C vs a		n				
		AAC	AFS			Dataset	-	•
	Train A		0.993 (±0.0	(13)		,	Tra	i
	Test A		0.898	2	Sec	retin-like 🗕	Tes	_
	Train A Test A		$0.99(\pm 0.0)$ 0.881	(2)	vs C	Glutamate- 📙	Tra	
	Iest A	0.850	0.881		like	:	11 a	
	(b)	) Trypsin vs ch					Tes	5
_		AAC		FS		,	Tra	i
	rain AM	0.992 (±0.01		$\pm 0.031$ )		odopsin-		
	lest AM	0.873		335	like		Tes	
	rain Acc Test Acc	$0.988 (\pm 0.02)$ 0.844		$\pm 0.047)$ 756			Tra	İ
	1			750	like		Tes	1
	(c	) Tubulin- $\alpha$ vs					Tra	
_		AAC		FS				•
	rain AM	0.996 (±0.00		$\pm 0.006)$		odopsin-	Tes	5
	Test AM	0.992		992	like	retin-like	Tra	i
	rain Acc	0.997 (±0.00	· · · · ·	$\pm 0.008)$	Sec	reun-like		
	Test Acc	0.991	0.9	994			Tes	
	(d) H	listone H2A vs	Histone H2H	3		'	Tra	i
		AAC	Α	FS	Am	inergic vs	T	_
Т	rain AM	0.983 (±0.0	16) 0.983	$(\pm 0.01)$	Lip	id	Tes	
	Test AM	0.91		934	rece	eptors	Tra	I
	rain Acc	0.983 (±0.0		$(\pm 0.01)$			Tes	-
	Test Acc	0.889	0.9	922			Tra	
		(e) Glob	ins					•
D	ataset		AAC	AFS		inergic vs,	Tes	51
		Train AM	0.998	0.994		eptors	Tra	i
			$(\pm 0.003)$	$(\pm 0.009)$			-	
	globin vs	Test AM	0.968	0.97			Tes	
Hem	oglobin- $\alpha$	Train Acc	0.998	0.995			Tra	1
		Test Acc	$(\pm 0.005)$ 0.969	$(\pm 0.006)$ 0.971		id vs —	Tes	.,
		Train AM	1.0	1.0	-	tide	Tra	
			$(\pm 0.0)$	$(\pm 0.0)$	rece	eptors		•
Mvo	globin vs	Test AM	0.957	0.936			Tes	
	oglobin- $\beta$	Train Acc	1.0	1.0	L	/ \ <b>T</b> ·	1	-
	0		$(\pm 0.0)$	$(\pm 0.0)$		(g) Inte	erle	u 
		Test Acc	0.949	0.919		True tre A D.C.		7
		Train AM	0.983	0.976		Train AM Test AM		(
Ham	oglobin- $\alpha$		$(\pm 0.008)$	$(\pm 0.007)$		Train Acc	- (	5
vs	ogi00iii-α	Test AM	0.961	0.935		Test Acc	+ (	,
	oglobin- $\beta$	Train Acc	0.983	0.976				-
	- <b>3</b> -0-0-0-1		$(\pm 0.008)$	$(\pm 0.006)$	(h)	Cytochrome	P45	(
		Test Acc	0.966	0.947	Γ			
						Train AM	0.	9
						Tost AM	1	

	(f) G	PCRs	
Dataset		AAC	AFS
	Train AM	0.933	0.95
Secretin-like		$(\pm 0.042)$	$(\pm 0.032)$
vs Glutamate-	Test AM	0.888	0.845
	Train Acc	0.933	0.95
пке		$(\pm 0.042)$	$(\pm 0.032)$
	Test Acc	0.889	0.844
	Train AM	0.884	0.85
Rhodopsin-		$(\pm 0.042)$	$(\pm 0.045)$
like vs	Test AM	0.967	0.934
Glutamate-	Train Acc	0.867	0.837
like		$(\pm 0.038)$	(±0.032)
	Test Acc	0.956	0.926
	Train AM	0.917	0.878
DI 1 '		$(\pm 0.051)$	$(\pm 0.065)$
Rhodopsin- like vs	Test AM	0.934	0.846
	Train Acc	0.908	0.863
Secretin-like		$(\pm 0.06)$	$(\pm 0.073)$
	Test Acc	0.941	0.853
	Train AM	0.949	0.943
A minanaia wa		$(\pm 0.014)$	$(\pm 0.005)$
Aminergic vs	Test AM	0.922	0.843
Lipid receptors	Train Acc	0.943	0.94
receptors		$(\pm 0.017)$	$(\pm 0.008)$
	Test Acc	0.92	0.84
	Train AM	0.835	0.818
Aminergic vs		$(\pm 0.06)$	$(\pm 0.053)$
Peptide	Test AM	0.844	0.79
receptors	Train Acc	0.83	0.819
receptors		$(\pm 0.06)$	$(\pm 0.051)$
	Test Acc	0.827	0.784
	Train AM	0.829	0.76
Lipid vs		$(\pm 0.022)$	$(\pm 0.035)$
Peptide	Test AM	0.845	0.709
receptors	Train Acc	0.838	0.75
receptors		$(\pm 0.018)$	$(\pm 0.032)$
	Test Acc	0.858	0.725

(g) Interleukin-1	$\alpha$ vs Interleukin-1	β
-------------------	---------------------------	---

AAC	AFS
0.98 (±0.04)	0.98 (±0.04)
0.979	0.985
$0.975(\pm 0.05)$	0.975 (±0.05)
0.961	0.971
	$\begin{array}{c} 0.98 (\pm 0.04) \\ 0.979 \\ \hline 0.975 (\pm 0.05) \end{array}$

50 CYP3 vs Cytochrome P450 CYP51

	AAC	AFS
Train AM	$0.967 (\pm 0.041)$	0.933 (±0.062)
Test AM	0.902	0.92
Train Acc	0.969 (±0.038)	0.936 (±0.062)
Test Acc	0.894	0.908

1155	E.1. Globin Family
1156 1157	The 3D structures of hemoglobin- $\alpha/\beta$ (PDB ID:1HHO) were aligned with myoglobin (PDB ID:3RGK) using the on-
1158	line pairwise structure alignment tool available at https://www.rcsb.org/alignment, with the default param-
1159	eter settings (algorithm: jFATCAT(rigid) — RMSD Cutoff: 3 — AFP Distance Cutoff: 1600 —
1160	Fragment Length: 8).
1161	>3RGK.A (Myoqlobin)
1162	
1163	GLSDGEWQLVLNVWGKVEADIPGHGQEVL <mark>IR</mark> LF <b>K</b> GHP <b>E</b> TL <b>EK</b> FDRFKHLKSEDE MKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKH <b>KIPVK</b> YLEFIS <b>E</b> AIIQVLQS <b>K</b> HPGD
1164	FGADAQGAMNKALELFRKDMASN <mark>Y</mark> K
1165	
1166	>1HHO.A (Hemoglobin- $\alpha$ )
1167	VLSPADKTNVKAAWGKVGAHAGEYGAEAL <mark>ER</mark> MF <mark>LSF</mark> PTTK <mark>TY</mark> FPHFDL
1168	SHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKL <mark>RVDPV</mark> NFKLLS <mark>H</mark> CLL <mark>V</mark> TL <mark>AAH</mark> L <mark>P</mark> AE
1169	<mark>FTP</mark> AV <mark>HA</mark> SL <mark>D</mark> KFLASVSTVLTSK <mark>Y</mark> R
1170 1171	
1171	
1172	>3RGK.A (Myoglobin)
1174	GLSDGEWQLVLNVWGKVEA
1175	DIPGHGQEVLI <mark>R</mark> LF <mark>KGHPE</mark> TL <mark>E</mark> KFDRFKHLKSEDEM <mark>K</mark> ASEDLKKHGATVLTALGGILKKKGHHEAEIK
1176	PLAQSHATKH <mark>K</mark> I <mark>P</mark> VK <mark>Y</mark> LEFIS <mark>E</mark> AIIQVL <mark>QS</mark> KH <mark>PG</mark> DFGAD <mark>AQG</mark> AM <mark>N</mark> KALELFRKDMASNYK
1177	>1HHO.B (Hemoglobin- $\beta$ )
1178	HLTPEEKSAVTALWGKV-
1179	NVDEVGGEALG <mark>R</mark> LL <mark>VVYPW</mark> T <mark>QR</mark> FFESFGDLSTPDAV <mark>M</mark> GNPKVKAHGKKVLGAFSDGLAHLDNLKGTFA
1180	TLSELHCDKL <mark>H</mark> V <mark>D</mark> PE <mark>N</mark> FRLLG <mark>N</mark> VLVCVL <mark>AH</mark> HF <mark>GK</mark> EFTPP <mark>VQA</mark> AY <mark>Q</mark> KVVAGVANALAHKYH
1181	$\Gamma'_{1} = \Gamma'_{2} T_{1} + \frac{1}{2} \frac{1}{1} \frac{1}$
1182	Figure E7: The highlighted AMINO ACIDS in myoglobin chain correspond to (after structure alignment) the positions which are hemologlobin- $\alpha/\beta$ tetramer contact points (as identified in Table 3 and Table 4 of (Shionyu et al., 2001)). We
1183	find that the amino acids $K, E, I$ , which are common in $AFS_1$ (Myoglobin) and $AFS_2$ (Myoglobin), are less in number at
1184	the contact residues of hemoglobin tetramer and more in number at the corresponding locations in myoglobin, which is a
1185 1186	monomer.
1180	
1187	
1189	
1190	
1191	
1192	
1193	
1194	
1195	
1196	
1197	
1198	
1199	
1200	
1201 1202	
1202	
1203	
1201	
1206	
1207	
1208	
1209	

1224 1225		AT I GG SLAGNE WYN FTOR I SE GAL SREFA JOG N WYN KHAGK WL LAFGOLAI RALD KLOF ALG KULUW LARHEG. ET FUOA JOG WU AGAALARTH HELE ALG WULW LARHEG. FUT BEAT STATAL ALG WU AND	RL FKGHPETLEKEDKKKHLKTEADMKASEDLKKHGNTKLTALGATLKKGHHDAELKPLAQSHATKHKIPIKVLETISATIHNLHSRHPA. EFGADAQGAMNKALELFRDIAKYKELGFH RLEKTHPETLEKEDKKKHLKTEDEMKASADLKKHGWLTALGSILKKKGQHEAELKPLAQSHATKHKISIKFLEFISEATIHNLQSKHSA. DFGADAQAMMGALELFRDIAK 40 40 10 10 10 10 10 10 10 10 10 10 10 10 10
1226			1 I I I I I I I I I I I I I I I I I I I
1227		DKLF DKLF DKLF DKLF DKLF DKLF DKLF DKLF	TKHN TKHN
1228		Е Ц Н С Е Ц В С Е	Q S H A
1229		AXLS AQLS AQLS AQLS AQLS AULS AULS AULS AULS AULS CLS AULS CLS AULS CLS CCLS CCLS CCLS CCLS CCLS CCLS CC	CPLA
1230		DTFA GTFA GTFA GTFA GTFA GTFA GTFA ATTY NTTA NTTY NTTY NTTY ATTA ATTA ATT	AELN
1231		DNIN DNL K DNL K DNL K DNL K DNL K DNL K DNL K DNL K DNL K DNL K C C C C C C C C C C C C C C C C C C C	GHHD
1232 1233		I I N I I A H L I A H L I A H L I A H L I A H L I A H L I A H L I A H L I A H L I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A H I A	LKKK 80
1233		<pre>6 GD A 5 D G G 5 S D G 6 GD A 6 GD A 7 S E G 6 GD A 7 S E G 7 S</pre>	GAII
1234		LTAR LGAR LGAR LTSE LTSE LTSE LTSE LTSE LTSE LTSE LTSE	LTAI
1236		6 K K V 6 K K V 7 K V 6 C K V 7 V 7 V 7 V 7 V 7 V 7 V 7 V 7	G V V
1237		VKAH VKAH VKAH VKAH VKAH VKAH VKAH VKAH	LKKH
1238		N P K N P K N P K N P K N P K N P K N P K N P K N N P K N N P K N N P M N N P R N N N N P K N N N N N N N N N N N N N N	S E D I
1239		A 1 C C A A W C C A A W C C A A W C C A A W M C A A W M C A A W M C A A W M C A A W M C A A W M C A A W M C A A W M C A A M C A A M C A A M C A A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M C A A M M C A A M M C A A M M C A A M M C A A M	DMK A EMK A
1240		GMF WMF WT QRLF SE FG ML SS PT A I GGN PK V KAHG KK VL TA FG DA IR NLDN I RLL VY PWT QR FF FF FG DL SS PD ANMGN PK V KAHG KK VL DS F SNG MKH LDD LI RLL VY PWT QR FF FS FG DL SS PD ANMGN PK V KAHG KK VL DS F SNG MKH LDD LI RLL I VY PWT QR FF SS FG LL SS PA MNG NF V KAHG KK VL TS FG DA VK ND NL RLL I VY PWT QR FF SS FG LL SS PA MNG NF V KAHG KK VL TS FG DA VK ND NL RLL I VY PWT QR FF SS FG LL SS PA MNG NF V KAHG KK VL TS GL AL DN LI RLL VY PWT QR FF SS FG LL SS A MNG NF V KAHG KK VL TS GL AL DN LI RLL VY PWT QR FF SS FG LL SNA MAHG I KV LHG LD SS GL GA LD NL RLL VY PWT QR FF SS FG LL SNA MAHG I KV LHG LD SS GL GA LD NL RLL VY PWT QR FF SS FG LL SNA MAHG I KV LG SS GL GA LD NL RLL VY PWT QR FF SS FG LL SNA MAHG I KV LS GS GL GA LD NL RLL VY PWT QR FF SS FG LL SNA MAHG I KV LG SS GL GA LD NL RLL VY PWT QR FF SS FG LL SNA MAHG I KV LG SS GL GA LD NL RLL VY PWT QR FF SS FG LL SNA MANN PK V KAHG KK VL SS FG LA LD NL RLL VY PWT QR FF SS FG LL SNA MANN SH KK KL GS SG LA LD NL RLL VY PWT QR FF SS FG LL SNA A A MS I KK GK KL GS SG LA LD NL RLL VY PWT QR FF SS FG LL SNA A A MS I KK GK VL SS FG LA LD NL RLL VY PWT QR FF SS FG LL SNA A A MS I KK GK VL SS FG LA LD NL RL L VY PWT QR FF SS FG LL SNA A A MS I KK GK VL SS FG LA LD NL RL FAAPF FL LF K FK KL KL FD FK KK GK VL TA LG GL L KK GG HH RL FG PF FL L KF LK KL SE DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL FG HPF FL L KF LK KL SE DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF LK KT FS ED MR A SE DL KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF LK KT SE DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF LK KT SE DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF LK KT SE DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF LK KT FS DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF LK KT SE DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF LK KT SE DD MR SS ED L KK HG YV L TA LG GL L KK GG HH RL F K HPF FT L KF KT KT SE DD MR SS ED L KK HG YV L A LG GL L KK GG HH RL K K HPF FT L KF KT KT	KTED
1241 1242		7 6 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	THX:
1242		F F F F F F F F F F F F F F F F F F F	FDKI
1243		TQR TQR TQR TQR TQR TQR TQR TQR TQR TQR	TLEK
1245		Mad Yang       Way and Yang <th>CHPE 40</th>	CHPE 40
1246		A RILV A	ALFK ALFK
1247			
1248		WATIGGESLA WEVGGEALG WEVGGEALG WEVGGEALG VADCGAEALA VADCGAEALA VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEVGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALG VEGGGEALI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGGQEVLI LAGGG	нсор
1249			DLAG
1250		QLITCLWCKVD. AATTALWCKVN. AATTALWCKVN. AATTSLWCKVN. QLITGLWCKVN. CLITGLWCKVN. AATTSINDCKVN. SAUTSLWCKVN. SAUTSLWCKVN. SAUTSLWCKVN. AATTSLWCKVN. AATTSLWCKVN. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATTSLWCKVL. AATT	VETI 20
1251		QL ITCLWGKVD AAVTGEWGKVD AAVTGEWGKVD QL TGLWGKVD QL TGLWGKVD AAVTGEWGKVD CLUGKVD CLUGKVD CLUGKVD CLUGKVD QL TSLWGKVD QL TSLWGKVD QL TSLWGKVD QL TSLWGKVD QL ULWVGKVD QL ULWVWGKVD QL ULWVWGKVD QU ULTVWGKVD QU ULTVWGKVD	VWG K
1252 1253		VTA VTA VTA VTS VTS VTS VTS VTS VTS VTS VTS VTS VTS	VLNI
1255 1254		EKQL EKAAA EKAAA EKAAA EKAA EKTI ERAA ERAA ERAA ERAA ERAA ERAA ERAA ERA	EWQL
1254		<ul> <li>WHTAEEKQLITCLUGKVD WATIGGESLA</li> <li>WHLTPEEKAATTSLWGKVN VDEVGGEALG</li> <li>MHLTAEEKAATTSLWGKVN VDEVGGEALG</li> <li>MHTAEEKQLITCLWGKVN VDEVGGEALG</li> <li>WHTAEEKQLITCLWGKVN VDEVGGEALG</li> <li>WHTAEEKANTGLWGKVN VDEVGGEALG</li> <li>WHTAEEKANTGLWGKVN VEEVGGEALG</li> <li>WWSDSERTINGFSULD EKVGGEALG</li> <li>WWNTPEEKANTALWGKVN VEEVGGEALG</li> <li>WHTGDEKAANTGLWGKVN VEEVGGEALG</li> <li>WHTGDEKAANTGLWGKVN VEEVGGEALG</li> <li>WHTGDEKAANTALWGKVN VEEVGGEALG</li> <li>WGLSDGWQHULTIWGKVESDLGHGQOELL</li> <li>MGLSDGEWULULWWGKVEADLAGHGQEUL</li> <li>MGLSDGWQUULUN WAVVENDVAHGQULLI</li> <li>MGLSDGWQUULUWWGKVENDLAGHGQEULI</li> <li>MGLSDGWQUULUWWGKVENDLAGGGEALI</li> <li>MGLSDGWQUULUWWGKVENDLAGGGEULI</li> <li>MGLSDGWQUULUWWGKVENDLAGGGEULI</li> <li>MGLSDGGWULUNWWGKVENDLAGGGEULI</li> <li>MGLSDGWQUULUWWGKVENDLAGGGEULI</li> <li>MGLSDGWQUULUWWGKVENDLAGGGEULI</li> <li>MGLSDGWQUULUWWGKVENDLAGGGEULI</li> <li>MGLSDGGWULUUWWGKVENDLAGGGEULI</li> <li>MGLSDGGWULUUWWGKVENDLAGGGEULI</li> <li>MGLSDGGWULUUWWGKVENDLAGGGEULI</li> <li>MGLSDGGWULUUWWGKVENDLAGGEULUI</li> <li>MGLSDGGWULUUWWGKVENDLAGGEULUI</li> <li>MGLSDGGWULUUWWGKVENDLAGGEULUI</li> </ul>	MGLSDGEWQLVLNVWGKVEADLAGHGQDIL MGLSDGEWQLVLKVWGKVETDITGHGQDVL 20
1255		нну нну нну нну нну нну нну нну	MGI MGI
1250		(HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb) (HHb)(	3 Mb 4 Mb
1258		P67824 Hbb P02030 Hbb P02030 Hbb P02410 Hbb P024140 Hbb P02140 Hbb P02140 Hbb P02140 Hbb P02140 Hbb P02140 Hbb P02140 Hbb P02140 Hbb P02138 Hbb P02137 Hbb P02137 Hbb P022138 Hbb P022138 Hbb P022138 Hbb P022138 Hbb P022138 Hbb P022138 Hbb P022138 Hbb P022138 Hbb P022137 Hbb P022138 Hbb P022137 Hbb P02213	P02173 Mb P02194 Mb
1259			
1260	Figure E8: Multiple se	quence alignment of hemoglobin- $\beta$ and myoglobin sequence	es. 15 sequ
1261	<b>c</b> 1 11' 0	d on the right are from myoglobin. The sequences are randomly	1 10

Figure E8: Multiple sequence alignment of hemoglobin- $\beta$  and myoglobin sequences. 15 sequences on the left are from hemoglobin- $\beta$  and on the right are from myoglobin. The sequences are randomly selected from the train set of the protein families. AFS(Myoglobin) amino acids are in green and AFS(Hemoglobin- $\beta$ ) in red. The intensity of the color is proportional to the Shapley value  $\phi(i)$  of the amino acid *i* (See Figure 3c)

## 1265 E.2. Tubulin

The inter-chain contact residues from the tubulin- $\alpha/\beta$  heterodimer were identified using ChimeraX 1.4 (Pettersen et al., 2021). The *Contacts* tool available in *Tools*  $\rightarrow$  *Structure Analysis* was used with settings as shown in Figure E9. For PDB ID:3JAR we count the residues of chain-A (tubulin- $\alpha$ ) and chain-B (tubulin- $\beta$ ) which are in contact with the residues of other tubulin chains. Similarly, for PDB ID:5N5N we count the residues of chain-G (tubulin- $\alpha$ ) and chain-B (tubulin- $\beta$ ) which are in contact with the residues of other tubulin chains. The code for counting the *AFS* residues at the identified contact points of the respective chains is available at https://anonymous.4open.science/r/AFS\_AAC\_SVM-F3D9.

Contacts
Interaction parameters
Find pairs of atoms with:
OVDW overlap ≥ -0.40 🔹 Å
⊖center-center distance ≤ 4.00 ♣ Å
Limit by selection $\Box$ with at least one end selected $ullet$
Ignore interactions between atoms 4 🔹 or fewer bonds apart
Ignore interactions between residues < 5 🔹 apart in sequence
Include intermodel 🗹 Include intraresidue 🗌 Include intramodel 🗹 Ignore hidden models 🗌 Include intramolecule 🗌
Treatment of results
Select atoms
Reveal atoms of interacting residues 🗌
Assign atomic attribute named Overlap
Display as pseudobonds
Color
Radius 0.075
Dashes 6
Distance label 🗌
Group name contacts
Urite information to:
Frequency of checking
• when OK/Apply clicked
Check O continuously (until dialog closed)
OK Reset Close Apply Help

Figure E9: ChimeraX 1.4 settings for identifying inter-chain contact points from the tubulin- $\alpha/\beta$  heterodimer and from the histone heterooctamer

### 1301 E.3. Histone

The inter-chain contact residues of histone H2A and H2B were identified from its heterooctameric structure comprising of two H2A/H2B dimers and one H3/H4 tetramer, using ChimeraX 1.4. The *Contacts* tool available in *Tools*  $\rightarrow$  *Structure Analysis* was used with settings as shown in Figure E9. For PDB ID: 1AOI and 3KWQ, we count the residues of an H2A and an H2B chain, which are in contact with other histone chains in the heterooctameric structure. The code for counting the *AFS* residues at the identified contact points of the respective chains is available at https://anonymous.4open. science/r/AFS\_AAC\_SVM-F3D9.

1320				N E		
1321	HILDOUTETTO PERFORMANCE HILDOUTETTO PERFORMANCE HILDOU	-2	AEE. EVA. PADDA PEEGI EAEDI GEYQO	ZEADI EVA. EN. EDEQ.		
	1001741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 1000741 100070	10444	E E E E E E E E E E E E E E E E E E E	VAEEL FEEGF FEEGF FEEGF FEEGF FAEEL	GME. GDEY. GDEY. BF. BY. CAY.	EY EM 50EEN 50EEN 57GDI
1322	CRATIK 17P7YG CRATIK 17P7YG CRATIK 17P5YG CRATIK 17P5YG CRATIG 12P7YG CRATIG 12P7YG CR	4 <b>4</b> 4 7	GEFE EFEE EFEE EFDPE GDYE GDYE GDYE AFED MEY	GEYE GEFE GEFE GEYE	GEEE GEEE GEDE NEGE DVEA AEEG	GEDE DEGE EAEE GADE GEEG GEEG GEEG
1323	DI CENTELLI PTIVOE NEL OL CLATTAL TPIVOE NEL OL CLATTAL SE YOOL NEL OL CLATTAL TPIVOE NE		DVA ARENGE REREGA DVA SADDGE Y ELAN AREN DVA ELGCE FERKET VA DVA ARENGE FERKET VA DVA ARENGE FEREVA DVA ARENGE PENDELE VARDON DVA ARENGE FEREVARD DVA ARENGE FEREVARD DVA ARENGE FEREVARD DVA ARENGE FEREVARD DVA ARENGE FEREVARD	ADEF AEEE AEEE AEEE CVEEE	raege saege svedae svedae sveae	STTEL SFEER SAEGL VGDD VGDD VGDD SVEGE SVEGE SVEGE SVEGE
1324	ALYD ALYD ALYD ALYD ALYD ALYD ALYD ALYD	, TUUA		SEX YOOY ON TADE REFEDE BOG ADT SEY QUYDAT AE REBELEYALE YA. (SEY QUYDAT AE REBELEYALE A. (SEY QU JGKR.). AFY QOYDAT YE REFEDE REGEN. AFY QOYDAT YE REFEDE SEVED FO	VG1ET VGAD VGVD VGVD VGAD VGAD	VGID VGAD VGAD VGAE VGAE VGVD VGTE VGTE
1325	ALL OF A CONTRACT OF A CONTRAC	DOV 3	EY00 EY00 EY00 EY00 EY00 EY00 EY00	EYQQ EYQQ GYQQ EYQQ EYQQ	DYEE DYEE DYEE DYEE DYEE DYEE	DYEE DYEE DYEE DYEE DYEE DYEE
1326	WEDTWEEPMAIL SURGUED IN THE AND INTERVIDED IN THE AND INTERVIDED IN THE AND INTERVIDED INTERVIDADI INTERVIDADI INTERVIDADI INTERVADADI INTERVIDADI INTERVI	an re	ME FLAAS SMOULY SE YOY HE FLAES SMOULY SE YOY	VIOX VIOX VIOX VIOX	AALEK AALEK AALEK AALEK AALEK AALEK AALEK	PRIVORO ELECTIAL REDULAL EXOREVOID TELOEDEET PRIVORENEEROE STATEDIALI EXOREVOIDS FERENEET TRIVOLOGIE LECTIALI EXOREVALTI VONELVOLATIONO LECHTIA PRIVOLOGIE LECTIALI EXOREVALTI VONELVOLATIONDALE LEL TRIVOLOGIE LECTIALI EXALENALLI KONTEVOLATIONO REGEDOO EN TRIVOLOGIE LEVI STAREDIALI KONTEVOLATI SOUCHE EN OORIGIE TRIVOLOGIE STAREDIALI KONTEVOLATI SOUCHE EN OORIGIE TRIVOLOGIE STAREDIALIKOUTI SOUCHE SOU
1327	VENA VENA VENA VENA VENA VENA VENA VENA	E C MIL	ESNM ESNM ESNM ESNM ESNM ESNM ESNM ESNM	E S NM E S NM E S NM E S NM E S NM	REDL. REDL. REDL. REDL. REDL. REDL.	REDL REDL REDL REDL REDL REDL REDL REDL
	DEPARTMENT OF THE	200 200	FTEA FTEA FTEA FTEA FTEA FTEA FTEA FTEA	IEFTEAESNM IEFTEAESNM IEFTEAESNM IEFTEAESNM IEFTEAESNM IEFTEAESNM	EFSEA EFSEA EFSEA EFSEA EFSEA EFSEA EFSEA EFSEA	FAEA FSEA FSEA FSEA FSEA FSEA FSEA FSEA
1328	NATL NATL NATL NATL NATL NATL NATL NATL	an au			MEEGI MEEGI MEEGI MEEGI MEEGI MEEGI	MEEGI MEEGI MEEGI MEEGI MEEGI MEEGI MEEGI
1329		LGEO	HWY 16E 600 BE HWY 76E 640 EE HWY 76E 640 EE	HWYTGEGMDE HWYTGEGMDE HWYTGEGMDE HWYTGEGMDE HWYTQEGMDE HWYTGEGMDE	VGEG VGEG VGEG VGEG VGEG VGEG	VGEG VGEG VGEG VGEG VGEG VGEG VGEG
1330	VSDT VSDT VSDT VSDT VSDT VSDT VSDT VSDT	AUG 1	CWH11	ГЦНИУ ГЦНИУ ГЦНИУ ГЦНИУ ГЦНИУ ГЦНИУ	CWHY Y	ZWHY 3 ZWHY 3 ZW
1331	660165604TLI I SUVEEPPORUM IS EVERY SUVEPYMET SUPERVALLENDE OF DOTATE ALT DE CENTLE LTP YOOL NO 660165604TLI SULVEEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE ALT DE CENTLE LTP YOOL NO 660165604TLI SULVEEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE OF CENTLE LTP YOOL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOUL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOUL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOUL NO 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE CONTONE AT DE CENTLE LTP YOUL NE 660165604TLI SULVEPPORUM IS VERY SPINSDIVVEPYMET SUPERVALLENDE VERY AND CONTOLE SULVEPACUAL DE PRY YOUR YOUR OBDE CONTONE AT DE CENTLE LTP YOUR YOUR OBDE CONTONE AT DE CENTLE LTP YOUR YOUR OBDE CONTONE AT DE CENTLE LTP YOUR YOUR YOUR AND CONTONE AT DE CONTONE AT DE CONTONE AT DE YOUR YOUR YOUR AND CONTONE AT DE PRY YOUR YOUR YOUR YOUR YOUR YOUR YOUR YO	V A G G	GLI PRAVING PLANARANA ANA PLANARANARANA PLANARANA PLANARANARANA PLANARANA PL	A 1 FOXS 10 FOXS 10 FOX 20 FOX 20 FOX 20 FOX 1 FOX 10 FOX	IR AVEL IL SY LA AVE YER JOHK DA LY MARAA VEHWYGHG HEGE SA RALEA LAID YA LYG FT AGGUE HEGE L DAROULINS STATA SA LONDALANA AN UNWYGHGALEGGE SA RALEA LAID YA URANASA KOLED BOOLY DAROULINS TTA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEA LAID YA URANASA KOLED BOOLY DAROULINS TTA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEA LAID YA UNYER YA DAROUNS YA DAROUN DAROULINS TTA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEAL ALAID YA DAROUNS YA DAROUNS DAROULINS TTA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEDA ALAID YA DAROUNS YA DAROUNS DAROULINS TTA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEDA ALAID YA DAROUNS YA DAROUNS DAROULINS TTA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEDA ALAID YA DAROUNS YA DARONA DAROULINS TTA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEDA ALAID YA DAROUNS YA DARONA YA TA DAROUNS SA LYA ALAMAZ NIKE DA LYAKAA VEHWYGHGOLEGGE SA RALEDA ALAID YA DAROUNS YA DARONA YA TA LAID YA DAROUNS SA LA ALAMAZ NIKE DA LYAKAA VEHWYGHGO HEGGE SA RALEDA ALAID YA DAROUNS YA DARONA YA TA LAID YA	A WOLLS NT ALA, E. WARLIBHERDLIYAKEA, IY YENYOGO E GE GE ALA KEDLALLEKIVEN DISTELODEEY AUVOLLSSTT ALA WARLIBHERDLIYAVKAA IY VEWYVGOGI EGE SA KEDLALLEKIVEKUDIS FERERGERT AUVOLLSSTT ALA WARLIBHEDLIKYDLIAVKAA IY VEWYVGOGI EGE SA KEDLALLEKIVEKUDIS FERERGUD GELS ALLOLISTT ALAYS'S RUDIKTDLIKYDLIAVKAA IY VEWYVGOGI EGE SA RAEDLALLEKIVEKUDIS FERERGUD GELS ALLOLISTT ALAYS'S RUDIKTDLIKYDLIAVKAA IY VEWYVGOGI EGE SA RAEDLALLEKIVEKUTIS EGE GEGOETS AUVOLLSSTT ALAYS'S RUDIKTDLIYAKAA IY VEWYVGOGI EGE SA RAEDLALLEKIVEKUTIS EDOOLDIST ALLOLISSTT ALAYS'S RUDIKTDLIYAKAA IY VEWYVGOGI EGE SA RAEDLALLEKIVEKUSTIS EDOOLEST AVOLLSSTT ALAYS'S RUDIKTDLIAVKAAA IY VEWYVGOGI EGE SA RAEDLALLEKIVEKUSTIS EDOOLEST AVOLLSSTT ALAYS'S RUDIKTDLIYAKAA IY VEWYVGOI EGE SA RAEDLALLEKIVEKUSTIS EDOOLEST AVOLLSSTT ALAYS'S RUDIKTDLIAVKAAA IY VEWYVGOI EGE SA RAEDLALLEKIVEKTIS EDOOLEST AVOLLSSTA ALAYS'S RUDIKTDLIAVKAAA IY VEWAY AN IXVISTANT AVOLUSTAT AN AVOLUSTAT AN AVOLUSTAT AVOLUSTAT AVOLUSTAT AN AVOLUSTAT
1332	1         5         V           1         5         V         V           1         1         5         V         V           1         1         5         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V         V	TAME	TAMP F SAMP F TAMP F SAMF F SAMF F TAMP F TAMP F QAMF	FTAMF FTAMF FTAMF FTAMF FTAMF FTAMF	AWTG3	AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI AMIDI
1333	646115 01 53 20 00 00 00 00 00 00 00 00 00 00 00 00	U a v	VAEQ VGEQ VGEQ VGEQ VGEQ VGEQ		[SNSTALAEVER[DHKPD] [SNSTALEVER[SRIDHKPD] SNTTALAEAVARLDHKPD] SNTTALAEAVARLDHKPD] SNTTALAEAVARLDHKPD] SNTTALAEAVEAUS SRIDHKPD] [SNNTAVAEVE SRIDHKPD] SNTTALAEAVEAU SRIDHKPD]	RAYCH SYTTA ALARWALDHYDD AVYDRIASTTA ALARWALDHYDD AVYDRIASTTA ALWALDHYDD AVYDRIASTTA ALWALDHYD ALALDHISSTTA ALARVASTA BADHYDD ALALDHISSTTA ALARVASTA BADHYDD AVYDRIASTTA ALARVASTA BADHYDD AVYDRIASTTA ALARVASTA BADHYDD AVYDRIASTA ALARVASTA BADHYDD AVYDRIASTA ALARVASTA BADHYDD AVDRIASTA ALARVASTA ALARVAD
1334	FI (466-163-014) XI (1110-0163-0140) XI (1110-0163-0140) XI (1110-0140) XI (11	d A SIV	MV1F 045110 UBHKWUR 1851 VOSTA UBH FRAVE ASTF 0551A UBH FRAVE ASTF 10551A UBH FRAVE ASTF 105513 UBH FRAVE ASTF 05513 UBH FRAVE	A STFIGNSTSIGEMFRANSE (Satfignmaaigelficvse antfignstaigemfranse astfignstsigemfranse Astfignstaigelfranse Satfignstsvgelforvant	EVFSR EVFSR EVFSR EAWAR EAWAR EAWAR EAWAR EAWAR	AWAR AWAR AWAR E V F SR E V F SR E AWAR E V F SR E V F SR
1335	SKVR SKIR SKIR SKIR SKIR SKIR SKIR SKIR SKI	1		TS 10 AA 10 TA 10 TA 10 TA 10 TA 10 TA 10 TA 10 TA 10	TAIAI TSVAI TAIAI TAIAI TAIAI TAIAI TAVAI TAIAI	TAIAI TAIAI TAIAI TAIAI TAIAI TAIAI TAIAI TAIAI
	6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 6711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 7711/15 771	-	KANTYUGNSTAT KISSTVGNSTAT SLASTFIGNSTAT SLASTFIGNSTST KASTFIGNSTST KASTFIGNSTST KASTFIGNSTST KASATIGNSTST KASATIGNSTST	I GNS I GNN V GNS V GNS I GNS	II SNS II SNS II SNT II SNT II SNT II SNT II SNN II SNN	L SNT L SNT L SNT L SNT L SNT L SNT L SNS L SNS L SNS
1336	react search to the search of	AVT.	MAVT ISST MAAT MAST MSST MSST MSAAI MSAAI MSAT	MASTI MSATI MAVTI MASTI MAVTI MAVTI MSATI	RAVCA RAVCA RAVCA RAVCA RAVCA RAVCA RAVCA	RAVCA RAVCA RAVCA RALCA RALCA RAVCA RAVCA RAVCA RAVCA
1337	166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166671 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 166771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 167771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 177771 1777771 1777777	6	N N N S N S	N N N N		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
1338	GEO TEST GGGT GS GGT LL SK VELEY DALIN SYNE SYN SYN WYN YN SWL YC UNYA, DF CHYDAR AL DD CAHLLT FFYGAL HU GO CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TL WYU UY NAD FLYDDAR AL DD CAHLLT FFS GOL HU GO CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR AL DD CAHLLT FFF GGD HU GO CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR AL DD CAHLLY FFF GGD HU GO CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR AL DD CAHLLY FFF GGD HU GO CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR AL DD CAHLLY FFF GGD HU GO CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR AL DD CAHLLY FFF GGD HU GG CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR AL DD CAHLLY FFF GGD HU GG CL LIST GGGT GS GGT LL SK LELE PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR AL DD CAHLLY FFF GGD HU GG CL RIST GGGT GS GGT LL SK LEFF PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR YW D C FHLLK FFF GOD HU GG CL RIST GGGT GS GGT LL SK LIEF PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR YW D C FHLLK FFF GOD HU GG CL RIST GGGT GS GGT LL SK REFF PDRAM I S WYFSYG YWYN TH WYU UY NAD FLYDDAR YW D C FHLLK FFF GOD HU GG CL RIST GGGT GS GGT LL SK REFF PDRAM I S WYFYN HW WYL WYW WYN WYN WYN YW D C FHLLK FFF GOD HU GG CL RIST GGGT GS GGT LL SK REFF PDRAM I S WYFYN WYWW WYW WYN WYN WYN WYN WYN WYN WYN				PGGDI PGGDI PGGDI PGGDI PGGDI PGGDI	PGGDI PGGDI PGGDI PGGDI PGGDI PGGDI PGGDI
1339	PCL00011153 PCL00011153 PCL00011153 PCL00011153 PCL00011153 PCL0001201153 PCL0001201153 PCL0001201153 PCL0001201153 PCL001201153 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011753 PCL0012011	1040	PKGL PRGL PRGL PRGL PRGL PRGL PRGL PTGL PTGL PTGL	PRGL PRGL PRGL PRGL PRGL	VVT q VVT q VVT q VVZ q VVZ d VVZ d	NT OP F T V P G D L AK NY OP F T V P G D L AK NY A B S V P G D L AK NY A P S V P G D L AK NY A P S V P G D L AK NY A P S V P G D L AK NY A P S T V P G D L AK NY OP F T V P G G D L AK NY OP P T V P G G D L AK NY OP P T V P G D L AK
1340	Herein II Herein II	LCD 1		VCD11 VCD11 VCD11 VCD11 VCD11 VCD11 VCD11 VCD11	INYQI INYQI INYQI IOYUI IOYUI IOYUI INYQI INYQI	NYQ NYQ NYQ NYQ NYQ NYQ
1341	REEAL REEAL REEAL REEAL REEAL REEAL REEAL REEAL REEAL REEAL REEAL RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU RELAU	1 V C C	I K S S I Q T A V K T A V K T A V K S S V Q T A	VKSS VKTA VKSS VKSS VLTA VLTA	FKCG FKCG FKVG FKVG FKVG FKCG FKVG	FKVG FKVG FKCG FKCG FKCG FKCG
1342	e1. Discribbwark and clicition of this (ded fostigation) is (Nue FPD and Information Structure and Structure an	AN A LAN		MI PNN MI PNN MI PNN MI PNN MI PNN	WCPTG WCPTG WCPTG WCPTG WCPTG WCPTG	QF UDWC PTGFKVGT QF UDWC PTGFKVGT QF UDWC PTGFKVGT QF UDWC PTGFKCGT QF UDWC PTGFKCGT QF UDWC PTGFKCGT QF UDWC PTGFKCGT QF UDWC PTGFKCGT QF UDWC PTGFKCGT
1343	LIDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYLL LUDSYL		YFVE YFVE SYFVE SYFVE SYFVE SYFVE SYFVE SYFVE	SYFVE SYFVE SYFVE SYFVE SYFVE AYFVE	IQF VD IQF AD IQF VD IQF VD IQF VD IQF VD IQF VD	105 VD 105 VD 105 VD 105 VD 105 VD 105 VD 105 VD
1344	EGAE EGAE EGAE EGAE EGAE EGAE EGAE EGAE	-	NKN S/ NKN S/ SKN S/ NKN S/ NKN S/ NKN S/ NKN S/ NKN S/ NKN S/ NKN S/ NKN S/ S/ NKN S/ S/ NKN S/ S/ NKN S/ S/ S/ NKN S/ S/ S/ NKN S/ NKN S/ NN S/	NKNS DKNS NKNS NKNS NKNS	TKRT TKRT TKRT TKRT TKRT TKRT TKRT TKRT	TKRT TKRT TKRT TKRT TKRT TKRT TKRT TKRT
		I NVO	A SUBPART LIA REAL TENDER OF DONES AND WAY AND WAY AND	I NVQ I NVQ I NVQ I NVQ I NVQ I NVQ I NVQ	APY 1S LEAVIED US AL IT SAFES MARCEMENTANCE. INFORM POWARY INFORMED OF ACCINYOPT TWAR ACCINSS IS ALLEAVIED ALLEAVIER OF THE A	ACCLI Y&OVVEXNYALIATIRIRAS I GYUBYCPTORIVO ACCLI Y&OVVEXNYALIATIRI URVINETI OFWARTOFIC ACCLI XBOVVEXNALIATI URVINETI OFWARTOFICKU ACCLI XBOVVEXNALIATIRI URVINETI OFWARTOFICKU ACCLINADIVEXNALIATIRI URVINETI OFWARTOFICKU ACCLINADIVEXNYANYI IKIRATI OFWARTOFICKU ACCLINADIVEXNYANYI IKIRATI OFWARTOFICKU
1345	DI I. FOL DIA VIA P 60, H PANA VIG F ADAWAGE TA ANALON MACHINE TA TA ANALON MACHINE AND MALE AND MA	vnew	AL SO OFFICIAL TARANT CONSTRUCTION OF TARANTAL AND A CONSTRUCT	CAADPRARTYT ASAMERCENSITEEDED UNIV AACOPRAGYLI YAALI RORMMELDE OUNIV CAADPRAGYLI YAALI RORMSTRETDE OUNIV CAADPRAGRI TAANI RORMSTRE DE OONIV CAADPRAGRI TAAVEROKYSIKEEDE OONISI TACYROWORT I TAAVEROKI SKEETE DOOMISI	VNAA VNAA VNAA VNVA VNVA AHAA VNAA VNAA	. RGDVVFKDVKAI AT VRGDVVFKDVVAA AT VRCDVVFKDVTAAVAA VRCDVVFKDVTAAVAA VRCDVVFKDVVAAVAT VRCDVVFKDVVAAVAT VRCDVVFKDVVAAVAT
1346	QTGA QTGA QSGA QSGA QSGA QSGA QSGA QSGA QSGA QS	1 T T	STKE VANKE SNKE SNKE SNKE SNKE STKE STKE ISTKE	ASTKEVDE AFMEEVDE ASTKEVDE ASTKEADE VSMKEVEE VSMKEVEE I SMKEVDE	VVPKD VVPKD VVPKD VVPKD VVPKD VVPKD	VVPKD VVPKD VVPKD VVPKD VVPKD VVPKD VVPKD
1347	47 V FG 47 V F	100	FRGR FRGR FRGR FRGR FRGR FRGR	FRGR FRGR FRGR FRGK	YRGD YRGD YRGD YRGD YRGD YRGD	YRGD YRGD YRGD YRGD YRGD YRGD
1348	FRPD	TACAT	TASAL TVAAV TVAAV TVAAV TVAAV TASAA TASAA TASAA	TASAA IVAA TASAI TASAI TASAA TVAAV TCSAI	ACCLA ACCLA ACCMI ACCLI ACCLI ACCLI ACCLI ACCLI	ACCMI ACCMI ACCLA ACCLA ACCLA ACCLA ACCLA
1349	SNK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV SNK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDS VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPGT IDD VIA AGP GGL RP JNV (SOK VP RAUL JDL FPG	IVED	AND PERFORMENT OF ALL PEOR AND PEOPLE ACOPRELIGY LITYAN FEOR AND PEOPLE ACOPRELIGY LITYAN FEOR AND PEOPLE ACOPRELIGY LITYAN FEOR AND PEOPLE ALL PEOPLE AND PEOPLE CALOPPERGY LITYA AND PEOPLE ACOPRELIGY LITYAN FEOR AND STREPT ACOPRELIGY LITYAN FEOR AND STREPT ACOPRELIGY LITYAN FEOR AND STREPT ACOPRELIGY LITYAN FEOR AND STREPT ACOPRELIGY LITYAN FEOR AND STREPT	CAADPRHGRYLTASAMFRGAMS AACDPRHGCYLTASALFRGAMP CAADPRHGRYLTASALFRGRMS CAADPRHGRYLTASALFRGRMS CAADPRHGRYLTASAMFRGRMS AASDPRHGRYLTVAAVFRGRVS TACNFONGRFLTCSALFRGKIS	PAT 13 A LEAVING QL SVALE IT SAVE TA SAVE DREADS WALCOPENDEX VALET PAT 13 A LEAVING QL SVALE IT SAVE TA SAVE DREADS WALCOPENDEX VALET PAT 13 A LEAVING QL SVALE IT SAVE TA SAVE DREADS WALCOPENDEX VALET PAT 13 A LEAVING QL SVALE IT SAVE TA SAVE DREADS WALCOPENDEX VALET PAT 13 A LEAVING QL SVALE IT SAVE TA SAVE DREADS WALCOPENDEX VALET VALET SAVE VALET SAVE TA SAVE TA SAVE TA SAVE DREADS WALCOPENDEX VALET VALET SAVE VALET SAVE TA SAVE TA SAVE VALET SAVE SAVE VALET SAVE VALET SAVE VALET SAVE VALET SAVE SAVE SAVE SAVE VALET SAVE VALET SAVE SAVE SAVE SAVE SAVE SAVE SAVE SAVE	LGKYM IGKYM IGKYM IGKYM IGKYM IGKYM
1350	DSYRACFAC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSOFIC DSYRACSO	a a a a	ADPR SDFR CDPR CDPR ADPR ADPR CDPR CDPR	ADPRI CDPRI ADPRI ADPRI SDPRI CNFO	CDPRI CDPRI CDPRI CDPRI CDPRI CDPRI CDPRI	CDPRI CDPRI CDPRI CDPRI CDPRI CDPRI CDPRI
1351	61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/00 61/000	VUID		AMMCA AMMAA AMMCA AMMCA AMMCA AMMCA	SMMAK SMMAK SQUVK SQNVK SQNVK SMMAK VQMVK	QMVK QMVK QMVK MMAK MMAK QMVK SMMAK SMMAK
1352	TGGRYPEAUTUDLEPUT SXSXYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUDLEPUT SCGNYPEAUTUD	-	QQNFDAKNMA QQMBDFKNMA QQMEDAKNMA QQMEDAKNMA QQMTAKNMA QQMTDSKNMA QQMFDSKNMA QQMFDSKNMA QQMFDSKNMA QQMFDSKNMA	WDAKI FDAKI WDAKI WDAKI FDAKI	L SVAELTNSAFEPASM L SVAELTNSAFEPSSM L SVAELTNACFEPANQI L SVAELTNACFEPANQI L SVAELTSSCFEPNSQ M SVQEMTMSCFEPNNQI L SVAELTNAVFEPSSM L SVAELTNACFEPANQI	LTVAETTSSCFEPANO LISVAETTSSCFEPANO LISVAETTSSCFEPANO LUVAETTSSCFEPANO LUVAETTNYEFPASIA LUVAETTNYEFPASIA LIVAETTNSCFEPANO LISVAETTNSAFEPASIA LISVAETTNSAFEPASIA
1352	RAILD RAILD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAUND RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD RAULD	TOOT	TOQN TOQN	100M 100M 100M 100M 100M 100M	ITNSA ITNSA ITNAC ITSSC ITSSC ITNAV ITNAC	ITNAC ITNAC ITNAC ITNAC ITNAC ITNAC ITNAC ITNAC
	TGGETVPPALI SNRAVPALI TGGOXVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOCKVPALI SOC	TUPE	RALTVFEL FRAVTVFEL FRAUTVFEL FRALTVFEL FRALTVFEL FRALTVFEL FRALTVFEL FRALTVFEL FRALTVFEL	(RALTVPELT (RALTVAELT (RALTVPELT (RALTVPELT (RAUTVPELT (RAVTVPELT (RAMSVPELT)	SVAE SVAE SVAE SVAE SVAE SVQE SVPE SVAE	TVAE SVAE TVAE SVAE SVAE TVAE SVAE SVAE
1354	Transaction of the second seco	VeAu	A DI SANCA DI TAVA TA DI QUA DI SANCA DI ANALANA MANANA DI ANALANA DI ANALANA DI ANALANA DI ANALANA MANANA DI ANALANA DI ANALANA DI ANALANA DI ANALANA MANANA DI ANALANA DI ANALANA DI ANALANA DI ANALANA MANANA DI ANALANA DI ANALANA DI ANALANA DI ANALANA DI ANALANA DI ANALANA MANANA DI ANALANA	APLISEGSQQYAALTVPELT APLISEGSQQYAALTVAELT APLISEGSQQYAALTVPELT APLISEGSQQYAALTVPELT APLISEGSQQYAAUTVPELT APLIAEGSGAQSFRAMSVPELT APLISEGAQSFRAMSVPELT	PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS) PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS) PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS SIBIAL COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS COPRERS). PUT SALEAVILE (L. SVIET ITS AN ET AS COPRERS).	PVI SVEKAME GLTV AFT TAK CEP AND IVVAKAVNEGLTV AFT TAK CEP AND VI SALKAVNEGLTVALTI AKC EPAND VI SALKAVNEGLTVALTI AKC EPAND VI SALKAVNEGLSVALTI AVDSSSS VI SALKAVNEGLSVALTI AVDSSS VI SALKAVNEGLSVALTI AVDSSSS VI SALKAVNEGLSVALTI AVDSSS VI SALKAVNEGLSVALTI AVDSS VI SALKAVNEGLSVALTI AVDSS
1355		-	RGSQ RGSQ RGSQ RGSQ RGSQ RGSQ RGSQ RGSQ	RGSQ RGSQ RGSQ RGSQ RGSQ RGSA	AEKA AEKA AEKA AEKA AEKA AAKA AAKA AEKA	VEKA AEKA AEKA AEKA AEKA AEKA AEKA AEKA
1356	I BALWWYNA I BALW	ADITS	AULIS FGAUSS RATTYPELI ALIS FGAUSS RATTYPELI ALIS FGAUSS RATTYPELI APLIS ROSODYRALTYPELI APLIS ROSODYRALTYPELI APLIS ROSODYRALTYPELI APLIS ROSODYRALTYPELI APLIS ROSODYRALTYPELI APLIS ROSODYRALTYPELI APLIS APGOQ S RASTYPELI	APLISEGSQOYALITVELT APLISEGSQOYALITVAELT APLISEGSQOYALITVELT APLISEGSQOYALITVELT APLISEGSQOYALITVELT APLISEGSQSFAMSYPELT APLISEGAGSFAMSYPELT	APVIS APVIS APVIS APVIS APVIS APVIS	APVI APVI APVI APVI APVI APVI APVI APVI
1357	LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LERING LE	TOT	MIGF MVGF MVGF MVGF MVGF MVGF MVGF MVGF MV	MVGF MPGF MVGF MVGF MVGF	I S S L L S S L	LATY LATY LATY LASY LASY LASY LSSY LSSY
1358	0.1.1.0 0.1.1.0 0.1.1.0 0.1.1.0 0.1.1.0 0.1.1.0 0.1.1.0 0.1.1.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.0 0.0.00 0.0.00 0.0.000000	DIUD	RLHF RLHF RLHF RLHF RLHF RLHF	RLHF1 RLHF1 RLHF1 RLHF1 RLHF1 RLHF1	KIHFA KIHFA KIHFI KIHFI KIHFA KIHFA	RIHFI RIHFI RIHFI RIHFI RIHFI RIHFI RIHFI
1359	60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 60501 6050000000000	-		UPEP UPEP UPEP	LVPYP LVPYP LVPYP LVPYP LVPYP LVPYP LVPYP	LVPYP LVPYP LVPYP LVPYP LVPYP LVPYP
1360	(167) (167) (167) (167) (167) (168) (168) (168) (168) (168) (168) (178) (168) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178) (178)	TAVA	LAVN LAVN LAVN LAVN LAVN LAVN LAVN LAVN	LAVN LAVN LAVN LAVN LAVN LAVN	FQTNI FQTNI FQTNI FQTNI FQTNI FQTNI	FQTNI FQTNI FQTNI FQTNI FQTNI FQTNI FQTNI
1361	610P 610P 610P 610P 610P 610P 610P 610P	ad Ins	SDLRK ADLRK ADLRK SDLRK SDLRK SDLRK SDLRK SDLRK	SDLRK ADLRK SDLRK SDLRK SDLRK SDLRK	HTVUV HTVUV HTVUV HTVUV HTVUV HTVUV HTVUV HTVUV HTVUV	HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA HINGA
1362	ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEH ISDEHI	COT N	NTDS NTDS NTDS NTDS NTDS NTDS NTDS NTDS	N109 N109 N109 N109 N109 N109 N109	GALN GALN GALN GALN GALN GALN GALN GALN	6ALN 6ALN 6ALN 6ALN 6ALN 6ALN 6ALN 6ALN
1363	FWE V FWE V FW	A MISCUTTON DE AGAIN NEDI DET UNVI IDE DEI DESENTI	MANOT CLAIR DE GLASDILRA, NWL PFREI PUCA MISOUT CLAIR DE CLASDILRA, NWUPF PLEI FURDA ALISOUT CLAR POLINGURAL, NWUPF PLEI FURDA ALISOUT FURDA ALISOUT CLAR POLINGURAL, NWUPF PLEI FURDA ALISOUT FURDA A	A TINSOFTICIL REPORT NSDIRKALAWANT IP FIRE FRAVA ALTISOFTICIL REPORT NSDIRKALAWANT PRALIN FRAVA ANN SOFTICIL REPORT NSDIRKALAWANT IP FRALIN FRAVA ANN SOFTICIL REPORT NSDIRKALAWANT IP FRALIN FRAVA ATTSOFTICIL REPORT NSDIRKALAWANT IP FRALIN FRAVA	WHAT HIRD AND DETERMINED THE STORE THE STORE S	A 14 HI ATYAYA NA DATAWA TAWA TAWA TAWA TAWA TAWA TAWA T
	0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K 0165K	- UND	60110 60111 60111 60111 60111 60111 60111	GVTC GVTT GVTC GVTC GVTC G1TT G1TT	8117 8177 8177 8177 8177 8177 8177 8177	8117 8177 8177 8177 8177 8177 8177 8177
1364	90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90050 90		AAM ATM ATM ATM ATM ATM ATM AVMS GVMS	ATM ATM: AAM: ATM: ATM: TVM: TVM: TVM:	140 141 141 141 141 141 141 141 141 141	
1365	WITT DI COCCAVICA MULTI SCENE DI SCOVICET SDQ . LEN NVYNKA NY MIL DI COCCAVICA MULTI SCENE DI SCOVICET SDQ . LEN NVYNKA MULADOCCOSO (SCIALTENT VI DEL DI SCOVICET SDQ . LEN NVYNKA MULADOCCOSO (SCIALTENT VI DEL DI SCOVICET SDQ . LEN NVYNKA MULADOCCOSO (SCIALTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIALTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCOVICE SDD . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCIATT (SSID) . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCIATT (SSID) . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCIATT (SSID) . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCIATT (SSID) . LEN NVYNKA MULADOCCOSO (SCIATTENT VI DEL DI SCIATT (SSID) . LEN NVYNKA MULADOCCOSO (SCIATTENT (SSID) DI SCIATT (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . SSITT (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . SSITT (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . SSITT (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . SSITT (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . SSITT (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . LEN NVYNKA MULADOCCOSO (SSID) . LEN NC MULATI (SSID) . LEN NVYNKA MULADO					
1366	REI , WHI GOGGONG (GAKTWAY SDEIGD FIGTTHEDSDIG , LEATINWYNAR REI , WYWLI FIGGGONG UGAKTWAY SDEIGD FIGTTHEDSDIG , LEATINWYNAR REI , HYNGGGONG UGAKTWAY SDEIGD FIGTTHEDSDIG , LEATINWYNAR REI , HYNGGGONG UGAKTWAY YDDIGD FIGTTHEDSDIG , LEATINWYNAR REI , HYNGGGONG UGAKTWAY YDDIGD FIGTTHEDSDIG , LEATINWYNAR REI , HYNGGGONG UGAKTWAY YDDIGD FIGTTHEDSDIG , LEALWYYNAR REI , HYNGGGONG UGAKTWAY YDDIGD FIGTTHACHADDIG , LEALWYYNAR REI , YNHGGONG UGAKTWAY YDDIGD FIGTTAK SCHODD FITTH FIGT REI , SHIYUGGONG UGAKTWAY YDDIG FIGTTAK SCHODD FITTH FIGT REI , SHIYUGGONG UGAKTWAY YDDIG FIGTTAK SCHO					
1367	~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~ ~	- 0				
1368						
1369	017. 017. 017. 017. 014. 014. 014. 014. 014. 014. 014. 014					

1

V S

1370 Figure E10: Multiple sequence alignment of tubulin- $\alpha$  and tubulin- $\beta$  sequences. 15 sequences on the left are from 1371 tubulin- $\beta$  and on the right are from tubulin- $\alpha$ . The sequences are randomly selected from the train set of the protein families. 1372 AFS(Tubulin- $\alpha$ ) amino acids are in green and AFS(Tubulin- $\beta$ ) in red. The intensity of the color is proportional to the 1373 Shapley value  $\phi(i)$  of the amino acid *i* (See Figure E6c) 1374

1422 1423		P19374 h2b	Q76FE5 h2b	P78567 h2b	Q7GBK0 h2b	060814 h2b	Q27894 h2b		P0C1H3 h2b	Q1E5N0 h2b	Q76FE9 h2b	Q7Z2G1 h2b	P02289 h2b	P54346 h2b A2WKT1h2h	P04909 h2a	Q00728 h2a	P48003 h2a	P06897 h2a	Q4HTT1 h2a	P691411h2a	USHGA4 nza	0873G4 h2a	CARFI 12/11/24	A2Y5G8 h2a	Q71U19 h2a	45DWF1 h2a	Q7L7L0 h2a A1C5F11h2a	_
1421													_															1.
1419 1420														MAPKKDEKPATAEAGAEAPAKAEAKPKAEKAGKKAKKEPAK MAPKKDEKPATAEAPAKAEAKPKAEKAGKKAKKEPAK														ŀ
1418								-				1	-	AEAGA							ł	-	-					
1417	. MAPKVAEKKPSLAGKAPAGKAPAEK	MP PKVSG		MAP KAADKK PASKAPATASKAPEKKI	MP P KAA S	MP E PAK S	MPDPAKS.	MPEPAKS.	MP E P AK S	MAP KAAEKK PVEKT PAVKKPKAEKK	-		MPR S P S K S S P R K G S P R K G S P R K G S P I	EAPAKAEAKPKAEKAGKKA	MALFL				-	-	ł	-						ŀ
1416	AEKKI	SG		ADKK.	A S	KS.	KS.	KS.	K S	AEKK	-		SKSS.	AEAK	NULL		:		÷	÷	÷	÷	÷	-		:	÷	
1415	PSLAC			PASKA						PVEK	-		PRKG	PKAEL	LAON				ł	-	-	-	:	-				
1414	3KAPA		MS	4 P A T A	-	-	-	-		TPAVK			S P R K C	KAGKK	ULAL!			-	-	-	ł	-		-				
1413	GKAP.	K	AEYG	SKAP.	K	A	Α.	Α.	A.	KPKA.	÷		SPRK	AKKE			-	:	÷	÷	ł	÷	÷	-		:	÷	
1412	AEKKE	KAAKKA	QRQQF	EKKDA	GAKKA	PAPKK	PAPKK	PAPKK	5.4		MP.		GSPKE	PAKK <i>i</i> da		MP				-	ł			MA				-
1410	KEAGKKT.	KAGK.	MSAEYGQRQQPGGRGGRSS	GK	KGAKKAASKAK	. APAPKKGSKKA VTK	.APAPKKGSKKAVTK	APAPKKGSKKAVTK	APAPKKGSKKAVTK	VPTSKE	PATPAKRA.	MASQRLAP	KRGGKGAKRAGKG	KA AKEPKG.	MSGRGKOG	GGKG.			-	MTGGGKS.GG.		MSGRGKG	MS GGKG GK	GGKG		MSGRGKGA		
1409 1410	Г		GRSS.	K	S	Α	Α	Α	A	T d A.	. T P A	ASQR	AKRA	- AKE.	MSGR	K S I G	MSGR	MS SG	MAGR	MTGG	MSAK	MSGR	-MS G	K S S G	MS GC	MSGR	MS.	- MA
1408	T	AQK		Г	KAK.	VTK	VTK	VTK	VTK	SKE	KRA	LAP	GKG			GKAGS	GKQG.	GKSGG	GKAI.	GKS.C	GKTG.	GKG	0.0X0	SKSS6	GKG	GKGA.		MAGRGK.
1407		:						-	ł	-	-					SKD S		GK		36	ł			TGKTS	K.		K	
1406	TATG.	NISK	GN	AP SG.	AARS	AQKK.	VQKK.	AQKK.	T QKK.	GGEK	KRVQ.	RKSS.	GRRR.	DGEK	GKA	A	GKA.	AGDA	GAGA.	KASS.	. RKK	. GKA	VGSA	GVDG	VGSS	. GKA	SAGA	GKTS
1405	EKKKR	GDKKK	KKSKK	DKKKE	TDKKK	DGKKE	DGKKE	DGRKE	GDKKE	KGKKK	QEKRE	AAHKS	NVVKE	KDKKK	RAKAK	GKSQK	RAKAK	S SKAC	AKKAT	GKNAC	ASKGI	KGKSE	AKAS	TKK00	EKAST	RAKAK	DKA S 1	GKKA
1404	TATGEKKKRTKARKETYSSYIYKVLKQTHPDTGISTRAMSIMNSFVNDIFERVATEASKL	NI SKGDKKKNRKRKE SYAIYIYKVLKQVHPDTGI SSKAMSIMNSFVNDIFERIASEASRL	GNKKSKKRCRRKESYSMYIYKVLKQVHPDIGISAKAMSIMNSFVNDVFEQLACEAARL	AP SGDKKKR IKARKE TYSSY I YKVLKQVHPDTG I SNRAMS I LNSFVND I FERVATEA SKL	AAR S T DKKK R R R R R S Y S I Y I Y K V L K Q V H P D T G I S S K A M S I M N S F V N D I F E R I A A E A S R L	AQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRL	VQKKDGKKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRL	AQKKDGRKRKRSRKESYSVYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRL	TQKKGDKKRKKSRKESYSIYVYKVLKQVHPDTGISSKAMGIMNSFVNDIFERIAGEASRL	GGEKKGKKKSKKSMETYKIYIFKVLKQVHPDIGISSKAMSITNSFINDIFEKLAGESAKL	KRVQQEKRHHKKRTETESVYIYRVLKQVHPETGVSKKSMSIMNSFINDIFEKIALEASKLVRYNKHTLSSREVQTAVRLLLPGELAKH	RK SSAAHK SHKK PKR SWNVYV SR SLAAIN SHMSMSGR TMK IVN SYVNDVMER I AMEAA SIVRAHKKL IL GAREVQ TAVR LY PPELAKH	GRRRNVVKRRRRRESYGSYIYRVLKQVHPDTGISSRGMSVMNSFVNDVFERIAGEASRL	DGEKKDKKKKKSAVETYKLYIYKVLKQVHPDTGISSKAMSINNSFINDIFEKVATEASKL A sethartabretyssevityvit potudatoissoviakinassivadiebelissessevi	SRSS	SHSA	MSGRGKQG GKARAKAKTRSSRAGLQFPVGKVHRLLKKGV, YSEVGAGAPVYLAAVLEVLTAELELAGNAARDXKKTRTTPRHQLATRNDE ELYKLLGRVTTAQGG	QSRSA.	SRSS.	QSRSS.	[ SNSA	KSRSS		AAGKGKSSGGKSSGGKSSGGKSSGGKSSGGKSSGGKSGGK	SRSA.	<b>SRSA</b>	NSGKGK SAGADKASTSRSAKAGI TEPVGRIHKLIKGS, YAQRVGSGAPVYL ISVLEVI TAEILELAGNARDNKKSRI I PRHQLAIRNDE MAGDAT GTTSGETE AVDOSATE ADDVEDI I DVI VEGT AAADDVVAAATEVETAAADVVAAATEVETAADDVEVEDI I DVIDE AAADDAT	/SKSA
1403	KETYS	KESYA	KESYS	KETYS	RESYS	KESY!	KESY (	KESY!	KESYS	METYL	TETF	KRSW	RESYC	VETY	RAGLC	KAGL	RAGL	KAGL(	KGGL (	KAGL!	KAGL	RAGL	KAGL	RAGIC	KAGL 1	RAGL	KAGL WAGI C	KAGL
1402	SSYLY	AIYIY	SMY I Y	SSYIY	S I Y I Y	<b>VYVY</b>	<b>VYVY</b>	<b>VYVY</b>	S I YVY	KIYIF	SVYIY	NVYVS	GSYIY	KLYIY 2 c v i v	DVG	2FPCG	2F PVG	QFPVG	QFPVG	AFPVG	QFPVG	QFPVG	TFPVC	DEPCG	TFPVG	QFPVG	TFPVG	OFPVC
1401	KVLK.	KVLK	KVLK	KVLK	KVLK	KVLK	KVLK	KVLK	KVLK	5KVLK	<b>YRVLK</b>	SRSLK	YRVLK	XXVLX XXVX	RVHR	RVKR	RVHR	<b>FRIHR</b>	<b>FRIAR</b>	GRVHR	3R I GR	GRIHR	RVHR	RVKR	RIHR	RVHR	SR I HR	<u>ik lak</u>
1400	QTHPI	QVHP	QVHP	QVHP.	QVHP	QVHP.	QVHP.	QUHP	QVHP.	QVHP.	QVHP.	AINS.	QVHP.	QVHP.	LLRK	FLKN	LLRK	LLRK	FLKA	LLRK	YLKK	LLRK		FI KO	LLRK	FLRK	LLRK	1 NTKK
1399	DTGIS	DTGIS	DIGIS	DTGI	DIGI	DTGI	DTGI	DTGI	DTGI	DIGI	ETGVS	HMSMS	DIGI	DTGI	GN YA	NTQNK	GN.YS	GN.YA	GK.YA	GN . YA	GK.YA	GN . Y/	GN.YA	NTONK	GN.YA	GN.YA	GN.Y/	GK.Y.
1397	TRAM	SKAM	AKAM	SNRAM	S SKAM	S SKAM	S SKAM	S SKAM	S SKAM	SKAM	SKKSM	GRTM	S S R GM	SKAM	LERVG	CMR VG.	ERVG	AQRVG	<b>AERVG</b>	AQRVG	AKRVG	AERIG	AQRIG	WRVG	AQRVG	AQRVG	AQRVG TDIG	AERIG
1396	S IMN	NIMI S	INNN S	INTI SI	NINI SI	ININ DI	ININ DI	IG IMN	IG IMN	IL IN	INMI SI	IK I VN	NIN SI	ININI SI	AGAPV	AKAAV	AGAP	AGAP	AGAP	AGAP	AGAP	AGAP	SGAP	AKAAV	SGAPV	AGAP	SGAP	IAGAP \
1395 1396	SFVND	SFVND	SFVND	SFVND	SFVND	SFVND	SFVND	SFUND	SFVND	SFIND	SFIND	SYVNE	SFVND	SFINE SEVND	VYLAA	VYVTA	VYLAA	VYLAA	VYLAA	VYLAA	VYLAA	VYLAA	VYLTA	V T L L V	VYLTS	VYLAA	VYLTS	VYLAA
1394	I F ER	) I F E R	VFEQ	DIFER	) I F E R	DIFER	DIFER	DIFER	DIFER	JIFEK	DIFEK	OVMER	OVFER	NIFEK MITTED	VLEY	VLEY	VLEY	VLEY	<b>VULEY</b>	AVLEY	AVLEY	AVMEY	AVLEY ULEY	VEFV	VLEY	VLEY	SVLEY VI EV	AVLEN
1393	VATE	IASE	LACE	VATE	IAAE	IAGE	IAGE	IAGE	IAGE.	LAGE	IALE	IAME	IAGE	VATE.	LTAE	LTAE	LTAE	LAAE	LAAE	LAAE	LCAE	LAAE	EVLAAEII	TAE'	LTAE	LAAE	LTAE	LIAE
1392						ASRL.	ASRL.	ASRL.	ASRL.	SAKL.	A SKL.	AASI.	ASRL.		ILELA	VLELA	ILEL/	ILELA	VLELA	ILEL!	ILEL/	VLEL4		VLETA	ILELA	ILELA	ILEL/	VLEL
1391	A	Α	Α	A A	Α	Α	Α	-	Α	Α	_			S <	AGNAA	AGNAA	AGNAA	AGNAA	AGNAA	AGNAA	AGNAA	AGNAA	AGNAA	AGNAA AGNAA	AGNAA	AGNAA	AGNAA	AGNAA
1390	ATYTKKSTITSREIQTAVRLILPGELAKH.	AHYNKRSTITSREIQTAVRLLLPGEL	AQYSGRTTLTSREVQTAVRLLLPGELAKH	AAYNKKSTISSREIQTSVRLILPGELAKH	AHYNRRSTITSREIQTAVRLLLPGELAKH	AHYNKRSTITSREIQTAVRLLLPGELAKH	AHYNKRSTITSREIQTAVRLLLPGELAKH	PHYNKRSTITSREIQTAVRLLLPGELAKH	AHYNKRSTITSREIQTAVRLLLPGELAKH.	ARYNKKPTITSREIQTSVRLVLPGELAKH.	/R YNK	/RAHK	CQANRRRTISSREIQTAVRLLLPGELAKH	. SRYNKKPTVTSREIQTAVRLVLPGELAKH	RDNK	KDLK	RDNK	<b>RDNK</b>	VRDNK	ARDNK	ARDNK	ARDNK	ARDNK	KDI K	RDNK	RDNK	ARDNK	ARDNK
1389	KSTIT	RSTI	RTTL	KSTI	RSTI	RSTI	RSTI	RSTI	RSTI	KPTI	KHTL	KLTL	RRTI	KPTV	KTRI	VKR1	KTRI.	KSRI	KTRIV	KTRI.	KSRI	KTRI	KSRI v c n i i	VKRT	KSRI	KTRI	KSRI www.ly	KNRI
1388	I SRE I	TSREI	TSREV	S S R E 1	T SRE 1	TSREI	TSREI	TSREI	TSREI	TSREI	SSREV	GAREV	SSREI	T SRE 1	I PRH	T P R H L	I P R H L	I P R H L	VPRHI	I P R H I	TPRHI	I P R H I	I P R H	TPRHT	I P R H	I P R H	I P R H	VPRH
1387	QTAV	QTAV	QTAV.	QTSV	QTAV	QTAV.	QTAV.	QTAV.	QTAV.	QTSV	γQTAV.	/QTAV.	QTAV.	QTAV	0LAI	0LAI.	QLAI.	QLAI	QLAV	QLAI.	QLAV	QLAI	QLAI		0LAI	QLAV	QLAI	0 AI
1386	RLILF	RLLLF	RLLL	RLIL	RLLL	RLLL	RLLL	RLLL	RLLL	RLVL	RLLL	RLVLI	RLLL	RLVLI	RNDE	RGDE.	RNDE.	RNDE.	RNDE.	RNDE.	RNDE.	RNDE.	RNDD	RGDE	RNDE.	RNDE.	RNDE.	RNDE
1385	PGELA	PGELA	PGELA	PGELA	PGELA	PGELA	PGELA	PGELA	PGELA	PGELA	PGELA	P P E L A	PGELA	PGELA	ELN	ELL	BLD	ELN		ELN	ELV.	BLN			E	ELN	ELN	
1384		AKH.	AKH	AKH	AKH.				AKH				AKH	АКН лич	NKLLG	DTLI.	NKLLG	NKLLG	TKLLG	NKLIG	NKFLA	NKLLS	NKLL(	DTI I	NKLIG	NKLLS	NKLLG	GKLL(
1382	ΤΑ	ΛΑ	Α.	Λ <b>Α</b>	Λ <b>Α</b>	Λ <b>Α</b>	ΛA	Λ <b>Α</b>	Λ <b>Α</b>	Λ <b>Α</b>	AV.	A.	ΛA	ΛΑ Λν	FRVTI	RATI	<b>JRVTI</b>	3GVT I	3GATI	<b>J T V H E</b>	AGVTF	I T V D S	GNVT I	RATI	I T VHE	SGVTI	I TVT	GEVI I
1382	ATGDGTRA.	AVSEGTKA	AVSEGTKA	AVSEGTKA	AV S E GTKA	AVSEGTKA	AVSEGTKA	AVSEGTKA	AV SEGTKA	AVSEGTKA	AVSEGTKA.	AMAEGTKA	AVSEGTKA.	. AVSEGTKA	A0GG	AFGG	AQGG	SQGG	ASGG	AQGG	ASGG	I AQGG	LELAGNAARDNKKSRIIPRHLQLAIRNDD. ELNKLGNVTIAQGGVL	AFGG	AQGG	AQGG	JELAGNAARDNKKSRIIPRHLQLAIRNDEELNKLLGDVTIAQGGV ET AGNAARDNKKSRIIPRHLQLAIRNDEELNKLLGDVTIAQGGV	I ASGG
1380		KA.	KA.	KA	KA	КА	KA	KA	KA.	KA.	KA	KA	KA	SRYNKKPTVTSREIQTAVRLYLPGELAKHAVSEGTKAVTKFTSG AANVVVCTTSIDETATVDFTTDCEFAAV	or de la regelación de la compañía de	W GGKGK I GGKAG SKD AGK OK SHS AKAGLOF PC GKVKFLKVNI ONMRVGAKAVYVTAVLETVI TEVLE AGNAAKD KVKKI TPKH () AI KODE ELD TLI . KATI AF GGVLPK I NA LLKVE GKKKKNK TA	VL PN	NSSGCSGGKAGDASSKAQSSSAAGOFPVGRIHKLEKGN.YAQRVGAGPVYLAAVEFLAAELEELAGNAARDNKKSRIIPRHLQLAIRDEEINKLIGGVIS 1000000000000000000000000000000000000	MAGREAI GAGAAKAT SASSKGG QFVGRIARFIKAG, YAERVGGAPYYLAAV EYLAEV ELAGAARDKKTRIVRHIQLAVRDEE TKLIGGAT IA SGGYDYIHOHCI PKKAGSKASHADDDD	KASSGKNAQSRSSKAGLAF PVGRVERLIKKGS, YAQRVGAGAPVYLAAVLEVLAAE I LELAGNAARDNKKTR I I PRELQLAI RNDE ELNKLIGHVTI AQGGVL PVI HQNLI PKKTGTKPGKNASQE		GKAKGKSKSKSKSKSKSKSKSKSKSKSKSKSKSKSKSKSK	VGSAARAQUTEPORTELEVGRVERLIERGGA. YAQRIGSGAPVYTIAVLEYAAELLELAGNAARDNKKSRITERELQLAIRNDD., ELNKLIGNYTAQGVPYTHOULEKKSAAPSAGE			GKARAKAKSSSARAGTOFPVGRVHRFLKKGN. VAQRVGAGPVVLAAVEFVAAELGELAGNAARDNKKTRTPRHEQLAVRNDE ELNKLLSGVTLAQGGVLPNTQAVLLPKKTSKAST.	NT PN	ELGKLLGEVILLASGGVLPNLHAVLLPKKTKGGKGEETA
1379				÷					-			-	-		IOAV	INRAL	IQAVI	IQSEI	ІНѺНІ	INQHI	INNHI	IQAVI	PNIHQNLI	INRAT	I H Q S I	IQAVI	I HQNI	IHAV
1378 1379	VAKY	VTKY	VTKY	VTKY.	VIKY	VTKY	VTKYTSSK	.VTKYT SAK	VTKY	VTKFTSA.	VTKYTSS	VSNSCR.	VTKYTTSR	VTKF VTVV	LPKK	LLKV	LPKK	LPAK	LPKK	LPKK	LPKK	LPKK	LPKK	LIKV	LPSK	LPKK	LPKK	LPKK
1377	VAKYSTFDN	TSSK	VTKYT S SK.	VTKYSSSTK	VTKYTQSK	VTKYT SAK.	TSSK	T SAK.	VTKYT S SK.	TSA.	TSS.	CR.	TTSR	VTKFTSG.	TESH	EQKKI	PNIQAVLLPKKTETHHKAKGK	SGKPI	AGSSI	TGTKI	SQLK	S S QK	PKKSAKPSASQEL	PHINRALLENNANDASCEL.	K S I K (	TSKA	PNIHQNLLPKKSGKGDKASQEI DNIHAVLTPVTVGGVGFFTA	IXGG
1376							1					-	-		IKAKC	CNKT E	IKAKG	KAGG	ASHA	GKNA	<b>VGTA</b>	K	ASQE	I A K A I	ASQE	λK	NASQ	GEEI

Figure E11: Multiple sequence alignment of histone H2A and histone H2B sequences. 15 sequences on the left are from histone H2B and on the right are from histone H2B. The sequences are randomly selected from the train set of the protein families. AFS(Histone H2A) amino acids are in green and AFS(Histone H2B) in red. The intensity of the color is proportional to the Shapley value  $\phi(i)$  of the amino acid *i* (See Figure E6e)

1.420	
1430	TILRDAQLKS KEQDDEORKS KEQDDEORKS KEQDDEORKS KEQDDEORKS KEQDREORKS KEQDREORKS KEQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELQDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREORKS KELDREOR
1431	TL RDAGLK CLOBEDRY KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCODEND KCO
1432	
1433	HIN SS
1434	<ul> <li>DAVPK S, LINCT L, RDAGLIK</li> <li>DAVPOS IS KEKLODRE DAR ORKEN</li> <li>DAVPOS IS KEKLODRE DAR ORKEN</li> <li>DAVPOS IS KEKLODRE DAR ORKEN</li> <li>DAVOS STRET LIG RL LODR DOR ORKEN</li> <li>DANO S STRET LIG RL LODA DOR ORKEN</li> <li>DANO S STRET LIG RL LODA DOR ORKEN</li> <li>DANO S STRET LIG RL RODA LOD OR ORKEN</li> <li>STRET R S READ LIG RL RODA LOD OR ORKEN</li> <li>STRET R S READ LIG RL RODA LOD OR ORKEN</li> <li>STRET R S READ LID R FOR DOR LID REAL RULL RULL RURGER I NON LOD OR OR OF DOR OF</li></ul>
1435	
1436	<pre>VILD TRANSTACTH DATVERS LINGT LRADAGLK ITTEE FCND DATVERS LINGT LRADAGLK VITE FCND DATVERS LINGT LRADAGLK VITE FCND ADTVERS LINGT LADAGLK FCD TWDERS DATVERS LINGT LED REAL DATVERS DATVERS LINGT LADAGLK FCD TWDERS DATVERS LINGT LED REAL DATVERS DATVERS LINGT LED LED REAL DATVERS DATVERS LINGT LED LED REAL DATVERS DATVERS LINGT LED LED LED LA DATVERS DATVERS LINGT LED LED LA LINER RAT TS FOR NUCHTI RE LINGT LED LA LINER RAT TS FOR LANDY FILL REAL LA LINER RAT TS FOR LANDY FILL REAL LANDY LINE LINGT LED LA LINER RAT TS FOR LANDY FILL REAL FILL REAL LANDY LINE LINGT LED LA LINER RAT TS FOR LANDY FILL REAL FILL LANG LINE LANDY LINE LINE LINE LA LINER RAT LANDY FILL REAL FILL REAL FILL REAL FILL LAND LA LINER RAT LANDY FILL REAL FILL REAL FILL REAL FILL LAND LA LINER RAT LANDY LINE REAL FILL REAL FILL REAL FILL LAND LA LINER RAT LANDY LINE REAL FILL REAL FIL</pre>
1437	CVH. CVH. STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE: STLE
1438	<pre>Discrete control /pre>
1439	
1440	<pre>FFLD FFLD FFLD FFLD FFLD FFLD FFLD FFLD</pre>
1441	FIFEEE PYELD TRNNDACYH FIFEEE PYELD TRNNDACYH FIFEEE PYLFEE CND JACH FIFEEE PYLFEE SYD JACH FIFEEE FILLONNDE JACH FIFEEE FILLONNDE JACH FIFEEE FILLONNDE JACH FIFEEE JACH FIFE SYD JACH FIFEEE FILLONNDE JACH FIFEEE FILLONNDE JACH FIFEEE FILLONNDE JACH FIFEE SYD JACH FIFEE SYD FYND TRN SYD JYD DYNEE JLLYND AND YND G DYNEE SYD FWWY IS TS GAE FILLEFES GAF PYWY IS TS GAE FILLEFES GAF PYWY IS TS GAE FILEFES GAF PYWY
1442	
1443	VET 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1444	NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
1445	ODNDLEY CDDDGKS CDDDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLRF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDLFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODDFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF ODFF OD
1446	POLFODNLRSTLFFEE AUNTHIDDLRSTLSTFFEE AUNTHIDDLRSTLSTFFEE ENVERDDLRSTFSTFFEE FEATURESTFFEE ENVERDDLRSTFSTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFEE FEATURESTFFFFEE FEATURESTFFFFEE FEATURESTFFFFEE FEATURESTFFFFEE FEATURESTFFFFEE FEATURESTFFFFEE FEATURESTFFFFFEE FEATURESTFFFFFFEE FEATURESTFFFFFFEE FEATURESTFFFFFFEE FEATURESTFFFFFFFEE FEATURESTFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
1447	PERO I FONDLE I TE LE FLEE W I FELD TRNNDAUM ANARY FIDDDLA, I LE I FEEE W I FELD TRNDAUM ANARY FIDDDLA, I LE I FEEE W I FELT CND. J DE ACTUM FIDDDLA I LE I FEEE W I FELT CND. J DE ACTUM FIDDDLA I LE I FEEE W I FET CN WDD VEED. ACTUM FIDDLA I LE I FEEE W I FEE TO TO WDD FIL E SO FONDLA I FN I FEE W I FEE TO TO WDD AT ACTUM FIDDLA I LE I FEEE W I FEE TO TO WDD AT ACTUM FIDDLA I LE I FEEE W I FEE TO TO WDD AT ACTUM FIDDLA I LE I FEEE W I FEE TO TO WDD AT ACTUM FIDDLA I LE I FEEE W I FEE TO TO WDD AT ACTUM FIDDLA I LE I FEEE W I FEE TO WD W I FE ACTUM FIDDLA I LE I FEEE W I FEE TO WD W I FE ACTUM FIDDLA I LE I FEEE W I FEE TO WD W I FE ACTUM FIDDLA I LE I FEEE W I FEE TO WD W I FE ACTUM FIDDLA I AND FEE I LE I TO WOULD AT ACTUM FIDDLA I AND FEE I LE I TO WOULD AT ACTUM FIDDLA I AND FEE I LE I TO WOULD AT ACTUM FIDDLA I AND FEE I LI LE R SA F FEI I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI LE R SA F FEE I ACTUM FIDDLE A AND FEE I LI R R SA F FEE I ACTUM FIDDLE A AND FEE I LI R R SA F FEE I ACTUM FIDDLE A AND FEE I LI R R SA F FEE I ACTUM FIDDLE A AND FEE I LI R R SA F FEE I ACTUM FIDDLE A AND FEE I LI R R SA F FEE I ACTUM FIDDLE A AND FEE I LI R R SA F FEE I ACTUM FIE I R SA F FEE I SA F FEE I ANA F FEE I ACTUM FIDDLE A AND FEE I LI R R SA F FEE I ACTUM FIE I R R R R R SA F FEE I ANA F FE I A R FEE I ACTUM FIE I R R R R R SA F FE I R R R R SA F FE I R R R SA F FE I R R R R R R R R R R R R R R R R R R
1448	
1449	MEKLRKMI VEKRMKI LIEKLRKE ALIEKLRKE ALIEKLRKE VEKLRKE VEKLRKE VEKLKKVE NEKLKKVE NEKLKKS NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKKE NEKLKKE NEKLKKE NEKLKKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLKE NEKLK
1450	
1451	A A A A A A A A A A A A A A A A A A A
1452	
1453	HEHYN . EGFRGAVS SYNY HOLWN . KS FRQUVS SYN HOLWN . KS FRQUVS SYN HOLWN . KS FRQUVS SYN HOLFN . KS FRQUVS SYN HEINN . KS FRQUVS SYN
1454	<ul> <li>J. J. EG FR QAU</li> <li>J. K. S. FR QWU</li> <li>J. K. S. S. M. J. FK QWU</li> <li>J. K. S. S. M. J. FK QWU</li> <li>J. K. S. S. M. J. FK QWU</li> <li>J. K. J. S. S. M. J. FK QWU</li> <li>J. K. J. S. S. M. M. J. J. S. S. M. M. J. J. J. S. S. M. M. J. J. J. S. S. M. M. M. J. J. J. S. S. M. M. M. J. J. S. S. M. M. J. J. J. S. S. M. M. J. J. J. S. S. M. M. M. J. J. J. S. S. M. M. M. J. J. J. S. S. M. M. J. J. S. S. M. M. J. J. S. S. M. M. J. J. J. S. S. S. M. M. J. J. J. J. S. S. S. M. M. J. J. J. S. /li></ul>
1455	
1456	HEHYN EG HECWN KS GOPNN KS HOLNN KS HOLN KS HOLPN
1457	
1458	010 SHERNY 010 SHERNY 010 SHERNY 010 SHERNY 010 SHERNY 010 SHERN 010 SH
1459	
1460	<ul> <li></li></ul>
1461	
1462	
1463	0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF2 0D1D1CF2 0D1D1CF2 0D1D1CF2 0D1D1CF2 0D1D1CF2 0D1D1CF2 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1D1CF1 0D1
1464	S F PÓLLIL C F L. C T ONL LO C F L. C F OLLNS SYL. C F OLLNS SYL. S F N SYL SYL. C F OLLNS SYL. C F OLLN
1465	KØNKG S F ØD LDLG PL DGG I QL Ø I S KØNKG T ON LUG SL
1466	
1467	
1468	FFDVD61 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD01 FFEAD0
1469	NDDL NDDL NDDL NDDL NDDL NDDL NDDL NDDL
1470	
1471	
1472	MMAA WMAA WMAA WMAA WMAA WMAA WMAA WMAA
1473	
1474	
1475	MALYPELASEMÁN MALYPELASEMÁN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MALYPELSENAN MAL
1476	
1477	P791821/IIT         P29889/IIT           P26889/IIT         P26889/IIT           P2428/IIT         P14428/IIT           P34428/IIT         P34428/IIT           P34428/IIT         P34428/IIT           P34428/IIT         P34428/IIT           P34458/IIT         P3458/IIT           P3458/IIT         P3458/IIT           P3458/IIT         P3458/IIT           P3458/IIT         P44648/IIT           P3458/IIT         P44648/IIT           P1458/IIT         P44648/IIT           P1458/IIT         P1458/IIT
1478	P75 P76 P76 P76 P76 P76 P76 P76 P76 P76 P76
1/170	

1479

Figure E12: Multiple sequence alignment of interleukin-1  $\alpha$  and interleukin-1  $\beta$  sequences. 15 sequences on the left are 1480 from interleukin-1  $\beta$  and on the right are from interleukin-1  $\alpha$ . The sequences are randomly selected from the train set of 1481 the protein families. AFS (Interleukin-1  $\alpha$ ) amino acids are in green and AFS (Interleukin-1  $\beta$ ) in red. The intensity of the 1482 color is proportional to the Shapley value  $\phi(i)$  of the amino acid *i* (See Figure E6d) 1483

## 1485 E.4. Marginal contribution feature importance (MCI) (Catav et al., 2021) for AFS

<sup>1486</sup> <sub>1487</sub> For a feature i, its MCI score is defined as,

 $MCI(i) = \max_{S \subseteq N \setminus \{i\}} v(S \cup \{i\}) - v(S),$ 

Here,  $v(\cdot)$  is the same as that defined in Section 2.2. We compare the amino acids with the top-d (d = size of AFS) MCI scores to the AFS in Table E6. MCI is computed using the same approximation scheme as in Appendix Section C Algorithm 1 with appropriate modifications.

Table E6: AFS comparison with the amino acids having the top-d MCI (Catav et al., 2021) scores. Here, d is the size of AFS for the respective dataset. The amino acids that differ in the two sets are in **bold and underlined**, with their counts mentioned in the rightmost column. For 8 of 15 datasets, AFS and top-d MCI sets are the same, while only for two datasets do they differ in two amino acids. For all 15 datasets, at least the top-3 MCI amino acids are in AFS. For 11 of these datasets, at least the top-5 MCI amino acids are in AFS.

Paralog pair	top- $d$ MCI amino acids (rank-1 $\rightarrow$ rank- $d$ )	AFS	Difference count
Lysozyme C (74) and $\alpha$ -Lactalbumin (22)	$\{I,A,D,G,R,F,N,E,W,L\}$	$\{I, A, D, N, G, R, E, F, L, W\}$	0
Trypsin (66) and Chymotrypsin (17)	$\{Y,W,T,A,K,V,\underline{I}\}$	$\{Y, W, T, A, V, K, \underline{\boldsymbol{P}}\}$	1
Tubulin- $\alpha$ (117) and Tubulin- $\beta$ (191)	$\{Q, M, K, H, F, I, N, A, Y, C\}$	$\{M,Q,K,N,F,I,H,A,C,Y\}$	0
Histone H2A (180) and Histone H2B (177)	$\{L,G,K,S,M,T,N,F,Y\}$	$\{L,G,S,M,K,N,T,Y,F\}$	0
Interleukin-1 $\alpha$ (16) and Interleukin-1 $\beta$ (25)	$\{G,C,T,V,Q,S,A,\underline{I},P\}$	$\{C,G,T,S,V,Q,A,\underline{\boldsymbol{N}},P\}$	1
Cytochrome P450 CYP3 (32) and CYP51 (32)	$\{H, F, G, K, A, P, N\}$	$\{H, F, G, K, A, P, N\}$	0
Globins			
Myoglobin (107) and Hemoglobin- $\alpha$ (303)	$\{V,Y,E,K,S,G,W,I,C,P\}$	$\{E, S, Y, V, K, P, I, G, C, W\}$	0
Myoglobin (107) and Hemoglobin- $\beta$ (285)	$\{V,K,E,C,W,N,F,Y,M,I\}$	$\{K, V, C, E, W, N, F, M, Y, I\}$	0
Hemoglobin- $\alpha$ (303) and Hemoglobin- $\beta$ (285)	$\{W,S,N,P,\underline{\boldsymbol{V}}\}$	$\{W, P, N, S, \underline{\boldsymbol{G}}\}$	1
GPCRs			
Rhodopsin-like (181) and Glutamate-like (89)	$\{D, E, Q, G, L, \underline{I}\}$	$\{D, Q, E, G, \underline{M}, L\}$	1
Secretin-like (90) and Glutamate-like (89)	$\{W,H,Y,V,D\}$	$\{W, H, Y, V, D\}$	0
Rhodopsin-like (181) and Secretin-like (90)	$\{W, E, H, Q, S, M, V, A\}$	$\{W, E, M, S, V, H, Q, A\}$	0
Rhodopsin-like GPCRs			
Aminergic receptors (186) and Lipid receptors (113)	$\{L, E, P, \underline{K}, F, D, \underline{I}\}$	$\{L, P, E, \underline{\boldsymbol{W}}, F, \underline{\boldsymbol{M}}, D\}$	2
Aminergic receptors (186) and Peptide receptors (367)	$\{L, E, K, F, M, \underline{\boldsymbol{H}}, R, D\}$	$\{L, F, E, M, K, D, \underline{V}, R\}$	1
Lipid receptors (113) and Peptide receptors (367)	$\{R,G,P,\underline{K},I,V,\underline{T}\}$	$\{P, R, G, I, \underline{W}, \underline{S}, V\}$	2