



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

MODELO DE ESPACIO VECTORIAL PARA LA REPRESENTACIÓN Y
CLASIFICACIÓN DE ESTRUCTURAS DE PROTEÍNAS

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

RICARDO CORRAL CORRAL

GABRIEL DEL RIO GUERRA

Instituto de Fisiología Celular
Universidad Nacional Autónoma de México

NINA PASTOR COLÓN

Centro de Investigación en Dinámica Celular
Universidad Autónoma del Estado de Morelos

DANIEL ALEJANDRO FERNÁNDEZ VELASCO

Facultad de Medicina
Universidad Nacional Autónoma de México

Ciudad de México. Julio, 2017



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mi mamá, por su gran amor, confianza y apoyo incondicional.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología por otorgarme una beca para apoyar la realización de esta investigación.

Al Programa de apoyo a proyectos de investigación e innovación tecnológica, proyectos IN205911 e IN208014.

Al Dr. Gabriel del Rio Guerra por la guía en la realización de este trabajo, así como por proporcionarme espacio y recursos computacionales en las instalaciones del Instituto de Fisiología Celular, UNAM.

A la Dra. María Teresa Lara Ortiz por su asistencia en la gestión de trámites.

Índice general

1. Introducción	9
1.1. Proteínas, clasificación estructural e inferencia funcional	9
1.2. Espacio de plegados	15
2. Métodos	21
2.1. Aprendizaje automático	21
2.2. Alineamiento de ambientes estructurales	24
2.2.1. Regiones	24
2.2.2. Alineamiento estructural	25
2.3. Ordenamiento de vecinos estructurales	26
2.3.1. Atributos para la representación de secuencias de proteínas	27
3. Resultados	31
3.1. Sistema de Clases de Cúmulos de residuos	31
3.2. Espacio de plegados en 26 dimensiones	33
3.3. Propiedades del espacio de plegados	34
3.3.1. Distancias en el espacio de plegados	34
3.3.2. Separabilidad del espacio de plegados	37
3.4. Predicción de clasificación estructural	39
3.5. Obtención de vecinos estructurales	40

3.6. Predicción de plegados a partir de secuencia	45
3.6.1. Predicción de vectores de clases de cúmulos de residuos desde secuencia	45
3.6.2. Estado actual en la predicción de plegado desde secuencia	49
3.6.3. Evaluación de predictores de plegado de secuencias diver- gentes	53
4. Discusión	61
5. Conclusión	65
Bibliografía	65
Apéndice	73
A. Predicción de topología para objetivos en CASP12	77
B. Machine Learnable Fold Space Representation based on Resi- due Cluster Classes	81

Lista de abreviaciones

AUROC	Area under the Receiver Operating Characteristic
CAFA	Critical Assessment of protein Function Annotation (http://biofunctionprediction.org/)
CASP	Critical Assessment of protein Structure Prediction (http://predictioncenter.org/)
CATH	Class, Architecture, Topology, Homology (http://www.cathdb.info/)
ERF	Extremely Random Forests
HSP	Half-Space Proximal
MDS	Multi-Dimensional Scaling
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank (http://www.rcsb.org/)
RCC	Residue Cluster Class
RMSD	Root-Mean-Square Deviation
RNG	Relative Neighborhood Graph
ROC	Receiver Operating Characteristic
SAS	Structural Alignment Score
SCCRs	Sistema de Clases de Cúmulos de Residuos
SCOP	Structural Classification of Proteins (http://scop.berkeley.edu/)

CAPÍTULO 1

Introducción

1.1. Proteínas, clasificación estructural e inferencia funcional

Las proteínas son las macromoléculas más abundantes en las células y ejercen una gran variedad de actividades biológicas. Por ejemplo, las proteínas realizan actividades enzimáticas acelerando varias reacciones químicas, como las proteasas, que degradan a otras proteínas, o lipasas que aceleran la degradación de triglicéridos. Otras proteínas funcionan como hormonas comunicando y coordinando la respuesta de varias células, de igual forma, los receptores celulares que sensan la presencia de otras proteínas o compuestos no protéicos, son proteínas.

De esta manera, las actividades ejercidas por una gran cantidad de proteínas distintas constituyen la maquinaria celular que sostiene la vida.

Las proteínas están constituídas por secuencias de aminoácidos unidos covalentemente por un enlace peptídico. La secuencia de aminoácidos de una proteína también es referida como estructura primaria de la proteína. Las secuencias pueden ser de longitudes muy variables, pero sólo incorporando veinte tipos de

aminoácidos. Dadas las propiedades fisico-químicas de los aminoácidos incorporados en una secuencia de proteína y las interacciones entre estos aminoácidos y el medio celular, esta secuencia adopta arreglos locales altamente regulares como alfa hélices y hojas beta que se conocen como estructura secundaria. Finalmente, el arreglo global de todos estos elementos da origen a una estructura tridimensional marginalmente estable, o estructura terciaria. Se postula que la capacidad de una proteína de ejercer su actividad biológica radica en su estructura tridimensional (o *plegado*, producto del proceso de *plegamiento* de su secuencia de aminoácidos). Por ello, contar con información de la estructura tridimensional de una proteína puede esclarecer directamente su mecanismo de acción.

Existe una gran cantidad de proteínas de secuencia conocida, pero sin estructura tridimensional determinada. Los procedimientos comúnmente utilizados para la determinación de estructura incluyen difracción de rayos X, resonancia magnética nuclear y microscopía electrónica, experimentos que en general resultan más costosos que experimentos de secuenciación. Es por esta limitación en la disponibilidad de información estructural, que se han desarrollado esfuerzos internacionales para promover el avance en las tareas de predicción de estructura a partir de secuencia, siendo el más grande de estos la Evaluación Crítica de Predicción de Estructura de Proteínas (CASP, por sus siglas en inglés) [1]. Este experimento global se celebra cada dos años desde 1994 y consiste en liberar un conjunto de secuencias para las cuales su información estructural no ha sido dada a conocer. Cada laboratorio envía sus modelos de estructura predichos y se efectúa una sola medición de desempeño sobre todas las predicciones de los participantes. De esta manera, se obtiene una visión objetiva sobre el desempeño de las diversas estrategias de predicción y se promueve la investigación y difusión de las estrategias más prometedoras. En general, las estrategias radican de forma fundamental en comparaciones directas entre la secuencia objetivo y las secuencias conocidas para las cuales se conoce también su estructura. Los métodos predominantes de comparación y medición de similitud en secuencia son los basados en alineamientos [2]. Estos métodos rápidamente pierden utilidad a medida que las secuencias comparadas han divergido evolutivamente, para lo

cuál métodos más sofisticados para identificar homología remota han sido desarrollados, como HMMER [3], que construye un alineamiento entre la secuencia objetivo y un modelo oculto de Markov construido con la información de familias de secuencias completas. Cada celebración del experimento CASP muestra una pequeña mejora, principalmente por la mayor disponibilidad de secuencias con estructura tridimensional determinada experimentalmente [4]. Sin embargo, continuamente se encuentran secuencias con ningún parecido detectable aún por estos métodos más sofisticados a secuencias conocidas y los avances en la predicción de estructura sobre estas secuencias *difíciles* siguen siendo pobres [4].

El conocimiento de la estructura tridimensional permite inferir la función molecular de una nueva proteína con mejores expectativas que a partir de su secuencia, ya que la función tiene una relación más directa con la estructura que con la secuencia. Aunque el costo experimental para determinar el conjunto de actividades de una proteína puede ser similar o menor al costo para determinar la estructura tridimensional, la predicción de la función de las proteínas a partir de su estructura tridimensional permite poner a prueba hipótesis sobre la relación que guardan estas dos características de las proteínas. Por ejemplo, la obtención de ejemplos de estructuras de proteína semejantes (también llamados vecinos estructurales) a una nueva estructura de proteína, permite generar hipótesis sobre la función de dicha estructura nueva, a partir de las funciones conocidas de sus vecinos.

La transferencia funcional puede hacerse también sólo a partir de información de secuencia, sin embargo, la diversidad en secuencia es mayor que la diversidad estructural, ya que varias familias de secuencias pueden converger en un solo tipo de plegado. Esto impone una limitación a los métodos basados en comparación de secuencia, pues es necesario conocer ejemplos de toda la diversidad asociada a un plegado. Es esta dificultad (mencionada previamente) la que también limita la predicción de la estructura.

Un esfuerzo similar al experimento CASP se ha desarrollado recientemente para evaluar la capacidad de los métodos de predicción de función a partir de la

secuencia de las proteínas. El experimento de Evaluación Crítica de Anotación Funcional (CAFA, por sus siglas en inglés) celebró su primera edición en 2010. La segunda y más reciente edición finalizó en 2014, poniendo de manifiesto que los métodos de aprendizaje de máquina, que integran múltiples tipos de datos, tienden a tener un mejor desempeño que la transferencia de función por similitud local en secuencia [5].

A pesar de estos esfuerzos científicos comunitarios, la relación entre la estructura primaria de una proteína y su plegado tridimensional, así como la capacidad de esta estructura terciaria para ejercer alguna actividad biológica siguen sin comprenderse. Sin embargo, continuamente se generan más datos de ejemplos particulares de estructuras y sus roles funcionales. En particular, la creciente cantidad de estructuras resueltas depositadas en el banco de datos de proteínas [6] (PDB por sus siglas en inglés) ha revelado que la gran mayoría de las estructuras de cadenas completas, se pliegan usando sub estructuras repetidas, también llamadas dominios. El número de nuevas clases de dominios encontrados en nuevas estructuras resueltas cada vez es menor; esto sugiere que podría haber un número finito de clases de dominios en los que converjan todas las secuencias (ver Figura 1.1).

Los dos principales esfuerzos para definir estos dominios, así como de efectuar una clasificación estructural y evolutiva de los mismos, son las bases de datos CATH [7] y SCOP [8]. Tanto CATH como SCOP son esquemas de clasificación jerárquicos y comienzan por un nivel estructural global que depende del contenido de estructura secundaria, pasando por niveles de clasificación estructural más finos, hasta llegar a sub clasificaciones basadas en parentescos evolutivos medidos por alineamientos de secuencia. Los principales niveles de clasificación estructural en CATH son Clase, Arquitectura y Topología, que hacen referencia al contenido de estructura secundaria, a la organización global de estos elementos de estructura secundaria, así como a sus detalles específicos en su orientación y cercanía, respetivamente (ver Figura 1.2).

El nivel Homología en CATH agrupa dominios de proteínas que se piensa que poseen algún ancestro evolutivo común y por lo tanto se consideran homólogos. En SCOP, los niveles de clasificación principales son Clase, Plegado, Superfamilia y Familia. Los dos primeros hacen referencia a aspectos estructurales, y los

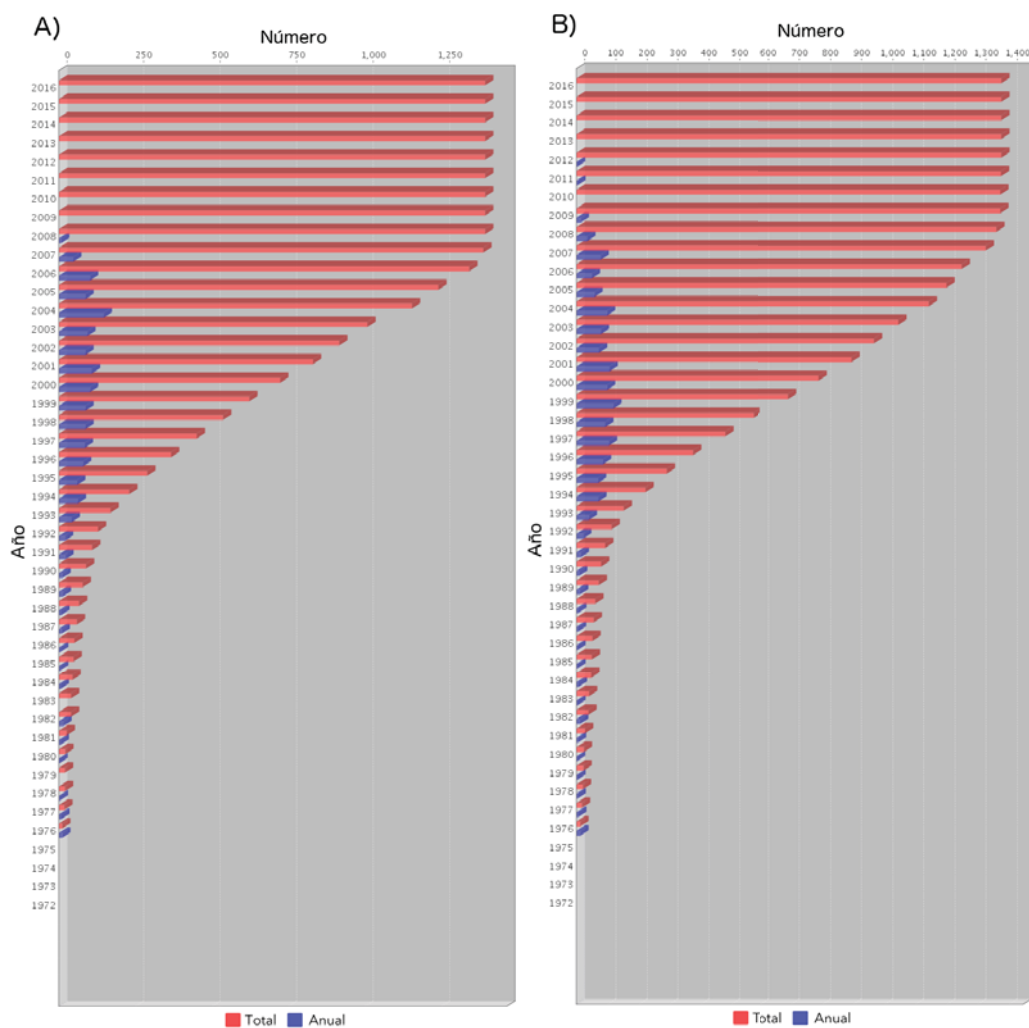


Figura 1.1: **Crecimiento de clasificaciones únicas de dominios por año**

La figura muestra la cantidad de nuevas clases de dominios determinados por año en azul, y el número total de clases de dominio acumuladas en rojo para las bases de datos SCOP 1.75 (panel A) y CATH 4.0 (panel B). Esta figura fue obtenida del sitio web de estadísticas del PDB (<http://www.rcsb.org/>).

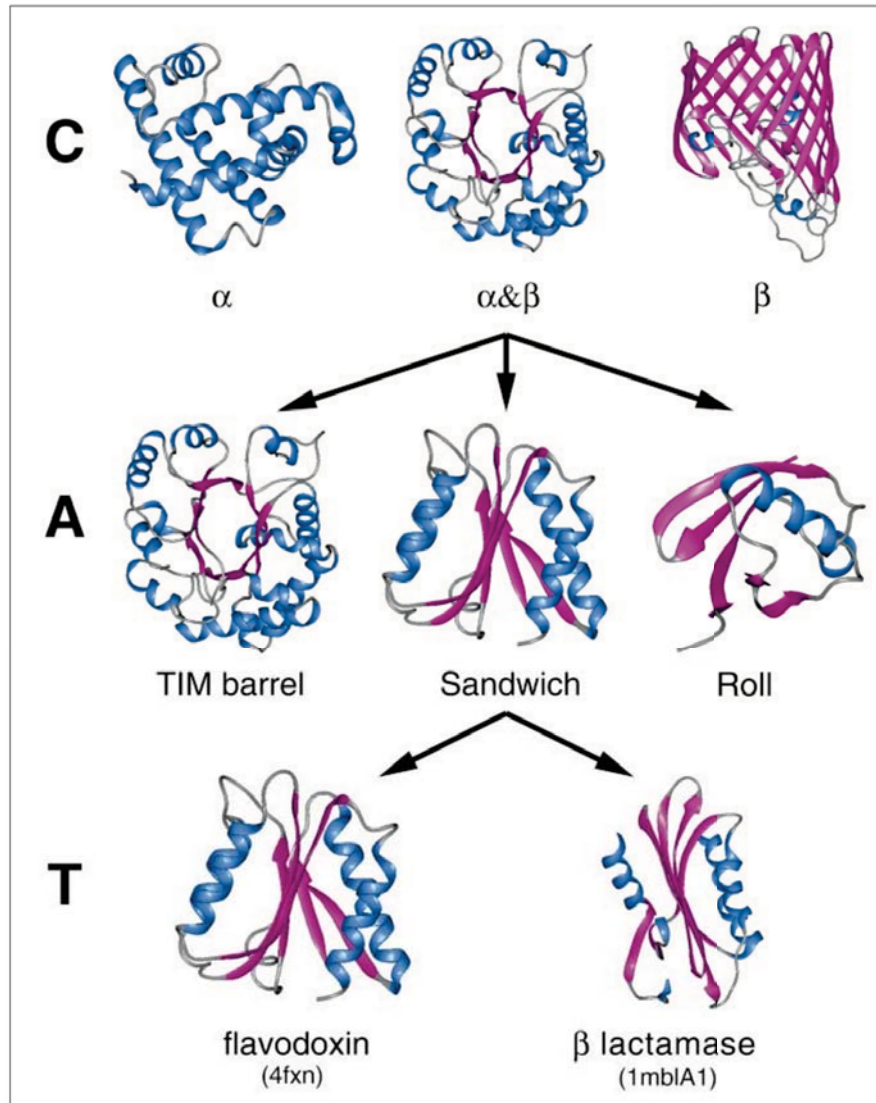


Figura 1.2: Niveles jerárquicos de clasificación estructural de dominios en CATH. La figura muestra ejemplos de dominios bajo las clases jerárquicas de Clase (C), Arquitectura (A) y Topología (T). Figura obtenida del artículo *CATH - a hierarchic classification of protein domain structures (1997)* [9]

dos últimos incluyen información de secuencia. Aunque CATH y SCOP poseen una gran similitud entre sus criterios de definición de dominios y clasificación, poseen algunas discrepancias que han sido estudiadas, dando origen a propuestas de nuevos esquemas de clasificación unificados [10], pero no han logrado una adopción global por la comunidad científica.

1.2. Espacio de plegados

Para abordar conceptualmente todos estos aspectos complejos de las proteínas y permitir una visión global de la diversidad estructural, se ha hecho referencia en la literatura a un espacio abstracto, denominado espacio de plegados, donde es posible representar cualquier plegado de una proteína. Este espacio de plegados se ha convertido en un marco conceptual en el que se unifican ideas acerca de la estructura, función y evolución de proteínas [11]. Por ejemplo, se ha debatido si el espacio de plegados es discreto o continuo [12, 13, 14], es decir, si hay cambios radicales entre *clases* de plegados, o si hay una sucesión de plegados muy semejantes conectando cualquier par de *clases*. Los resultados de esta discusión muestran que es posible transitar entre cualquier par de clases de plegados, a través de plegados intermedios con alta similitud, reforzando también la idea de que ya se han descubierto ejemplos de todas las clases estructurales posibles [13].

Para inferir las proximidades de plegados en este espacio, se han usado medidas de similitud como la desviación cuadrada media entre posiciones atómicas (RMSD, por sus siglas en inglés) [15]; en este caso, se asume que la proximidad de plegados es la única propiedad relevante para explicar función y evolución de proteínas.

Usando el programa DALI [16] para obtener una medida de similitud, Holm mostró proyecciones en dos dimensiones para explorar estructuras vecinas en el espacio de plegados [17]. DALI también ha sido utilizado como medida de similitud para visualizar la distribución de clases y uso de plegados entre dos especies bacterianas [18] y para explorar la asignación funcional basada en esta

representación del espacio de plegados [19]

Las representaciones explícitas del espacio de plegados basadas en medidas de similitud estructural tienen dos importantes limitaciones. Primero, el tiempo necesario para ejecutar una comparación para cada par de estructuras es restrictivo; por ejemplo, 25,000 horas de procesamiento fueron necesitadas para calcular similitudes entre 1,898 estructuras de proteínas [19]. Segundo, la posición de cada punto depende fuertemente del conjunto de estructuras analizadas, de manera que la inclusión de una sola estructura adicional puede desplazar a todas las posiciones previamente asignadas; por lo tanto, hay tantas representaciones del espacio de plegados como conjuntos de datos de estructuras de proteínas, incluso adoptando una única medida de similitud. También se ha mostrado que la forma cualitativa del espacio de plegados es altamente dependiente de la medida de similitud escogida y de los parámetros de escalamiento seleccionados [20].

Una representación alternativa del espacio de plegados puede ser construida asignando la posición de cada vector considerando únicamente atributos estructurales de la proteína que representa. En lugar de considerar únicamente la proximidad entre los plegados, cada estructura de proteína es representada por un punto (vector posición) en un espacio de dimensión fija (e.g., un espacio euclidiano en n dimensiones). Las posiciones de estos vectores pueden ser ajustadas de manera que las distancias relativas entre ellos aproximen alguna medida de similitud estructural de las proteínas que representan. Por ejemplo, usando la puntuación de similitud del programa de alineamiento estructural secuencial (SSAP) entre cada par de estructuras como medida de similitud [21], seguido de escalamiento multidimensional (MDS) [22]. De esta manera, la posición de una nueva estructura puede ser asignada sin alterar las posiciones existentes. Un ejemplo de representación de un modelo de espacio vectorial es FragBag [23] que representa estructuras de proteínas en 400 dimensiones; aquí, cada componente del vector representando un plegado de proteína corresponde al número total de ocurrencias de un fragmento de secuencia de proteína particular. Otro ejemplo de espacio vectorial se muestra en el trabajo de Fain et. al., donde se utilizan invariantes en teoría de nudos como valores de cada componente para vectores en un espacio de 30 dimensiones [24]. En estos casos, se asume que estructuras

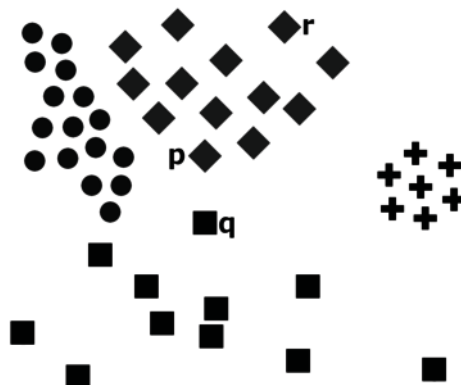


Figura 1.3: **Proximidad en el espacio no siempre implica pertenencia a la misma clase.** La figura ilustra casos en donde la proximidad en un espacio no implica pertenencia a una clase. Los puntos representados por cruces definen una clase donde las distancias entre cada par de sus miembros son menores que la distancia a cualquier otro punto perteneciente a otra clase, una suposición hecha al usar las puntuaciones de similitud. Una excepción a esta suposición es representada por los miembros de la clase rombo p y r , cuya distancia es mayor que la distancia entre p y q , que pertenecen a clases distintas. La clase círculo posee una distribución de miembros densa y regular, la clase rombo es regular y dispersa, mientras que la clase cuadrado es muy dispersa e irregular. En cualquier caso, las clases están claramente segregadas.

cercanas deberían pertenecer a la misma clase estructural, supuesto que no es necesariamente correcto como se mostrará más adelante.

Una vez que un método de construcción del espacio de plegados es elegido, la distancia métrica inducida por el espacio puede ser usada como una medida de similitud y se espera que esté en concordancia directa con medidas de similitud estructural (tales como RMSD, GDT, TMscore, etc.), pero superando los problemas de estas medidas como consecuencia de no ser métricas [25].

En una representación del espacio de plegados, se espera que las estructuras cercanas entre sí pertenezcan a un mismo plegado. Sin embargo, puede ocurrir que algunas proteínas con plegados diferentes se encuentren más cercanas que

proteínas bajo el mismo plegado. (ver Figura 1.3), de manera que puede haber confusión al distinguir la pertenencia a las clases en este espacio si únicamente se consideran las medidas de similitud.

En resumen, el espacio de plegados como marco conceptual ha facilitado la visión global de la diversidad estructural y su intrínseca relación con la diversidad funcional y sus orígenes evolutivos. Se puede hacer referencia a este espacio de plegados mediante representaciones abstractas o explícitas.

En las representaciones abstractas no se asigna una posición a las estructuras y el único concepto relevante es la cercanía vista como una manifestación de la similitud estructural. Esta similitud puede ser cuantitativa cuando se hace uso de alguna puntuación de similitud por algún método computacional basado en alineamiento estructural, o bien cualitativa, cuando la referencia a la similitud entre dos estructuras refleja solo la opinión humana basada en la experiencia.

En las representaciones explícitas, cada estructura es identificada con una posición en un espacio vectorial. Existen dos estrategias generales para asignar posiciones a cada estructura. La primera estrategia consiste en obtener una puntuación de la similitud entre cada par de estructuras analizadas y obtener las posiciones mediante escalamiento multidimensional de manera que las distancias entre las posiciones asignadas aproximen a las puntuaciones de similitud. La segunda estrategia es independiente de medidas de similitud y la asignación de la posición de una estructura sólo depende de sus características estructurales. Esta distinción entre representaciones del espacio de plegados se resume en la Figura 1.4.

En este trabajo se proponen las Clases de Cúmulos de Residuos como un modelo de representación de estructuras de proteínas individuales, así como una estrategia para la construcción explícita del espacio de plegados. Se exploran características de esta representación y su concordancia con esquemas de clasificación actuales y se abordan varias aplicaciones basadas en aprendizaje de máquinas. Finalmente, se explora una estrategia para el uso de este espacio como paso intermedio para la predicción de plegado de nuevas secuencias.

Representación abstracta		Representación explícita	
Cualitativa	Cuantitativa (RMSD)	RMSD/MDS	Basada en estructura

Figura 1.4: **Representaciones del espacio de plegados** Una representación abstracta cualitativa es cualquier referencia conceptual al espacio de plegados; la representación es implícita en la referencia a las proteínas y su cercanía en el espacio de plegados. Este mismo tipo de referencia abstracta se convierte en cuantitativa cuando los enunciados sobre cercanías están basados en cualquier medición explícita a partir de datos experimentales de la estructura de proteínas representadas. Las representaciones explícitas asignan una posición fija a cada estructura de proteína en un espacio vectorial. Estas posiciones pueden asignarse de tal manera que la distancia entre ellas aproxime valores de mediciones de similitud estructural, o bien, estas posiciones pueden ser obtenidas para cada proteína considerando únicamente su propia información estructural, independientemente de las demás estructuras conocidas.

Hipótesis

La representación del espacio de plegados basada en clases de cúmulos de residuos segrega a las estructuras de proteínas conforme a sus características estructurales.

Objetivos

Objetivo general Crear una representación del espacio de plegados que segregue las estructuras de proteínas.

Objetivos específicos

- Probar separabilidad del espacio de plegados bajo esquemas de clasificación establecidos.
- Identificar fronteras de decisión en el espacio de plegados para clasificación automática.
- Utilizar esta representación como medida de similitud para obtención de vecinos estructurales.

CAPÍTULO 2

Métodos

2.1. Aprendizaje automático

El aprendizaje automático es una rama de la inteligencia artificial cuyo objetivo es descifrar relaciones existentes sobre un conjunto de observaciones. El problema de clasificación en aprendizaje automático consiste en determinar la relación existente entre los atributos de un objeto y su *clase*. Formalmente, si se tiene un objeto x perteneciente a una clase y , se dice que existe una función

$$f(x) = y$$

que relaciona estas dos entidades. El problema de clasificación, cuando f es desconocida, es encontrar una función g tal que

$$g \approx f$$

La construcción de la función g se realiza únicamente a partir de datos observados, sin ningún conocimiento o suposiciones previas del comportamiento de f . En este trabajo, el objetivo principal es aproximar la relación entre una estructura de proteína y su clasificación estructural.

Existen varios modelos para representar g , cada uno con distintos parámetros internos Θ . El entrenamiento o aprendizaje de g_Θ consiste en encontrar el conjunto de parámetros Θ tal que $L(y, g_\Theta(x))$ sea mínima. Donde L es alguna función de costos que cuantifica la diferencia entre las estimaciones de g_Θ y los valores observados de y . La función L puede o no ser usada explícitamente en la búsqueda de los parámetros Θ . Los datos utilizados para entrenar a g se denominan datos de entrenamiento.

Un objeto x se representa mediante un número fijo de n atributos. Algunos modelos restringen la naturaleza de los atributos de x . Esta sección se concentra en atributos numéricos reales, es decir $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

Árboles de decisión

Un árbol de decisión es un modelo g que consiste en un conjunto de reglas jerárquicas a cerca de los valores en cada uno de los atributos en x . Cada regla es representada por un nodo que establece un valor de corte en un atributo dentro del rango de valores observados. Cada nodo separa el conjunto de datos de entrenamiento actual en dos grupos; el que contiene los objetos con un valor menor al valor de corte en dicho atributo, y el grupo de objetos con valores mayores al valor de corte. Nodos con reglas adicionales se pueden seguir añadiendo a cada subgrupo. Existen varios métodos para construir árboles de decisión, por ejemplo, los primeros métodos ID3 [26] y C4.5 [27] desarrollados por Ross Quinlan fueron los primeros métodos generadores de árboles de decisión reportados. La idea de estos métodos es seleccionar en cada paso uno de los atributos para elegir el valor de corte basado en la idea de ganancia de información. La ganancia de información de un atributo es la diferencia entre la entropía de los datos actuales de entrenamiento y la entropía inducida al seleccionar dicho atributo. La entropía cuantifica la uniformidad del número de ejemplos distribuidos en cada clase. Dos medidas de entropía utilizadas típicamente son la medida de entropía de Shanon y el índice Gini [28].

Un ejemplo de un árbol de decisión para estimar la Clase en CATH de estructuras de proteínas se encuentra en la Figura 2.1.

El desempeño de un árbol de decisión depende de los atributos y valores de corte seleccionados. Estos valores están sesgados a desempeñarse bien con los

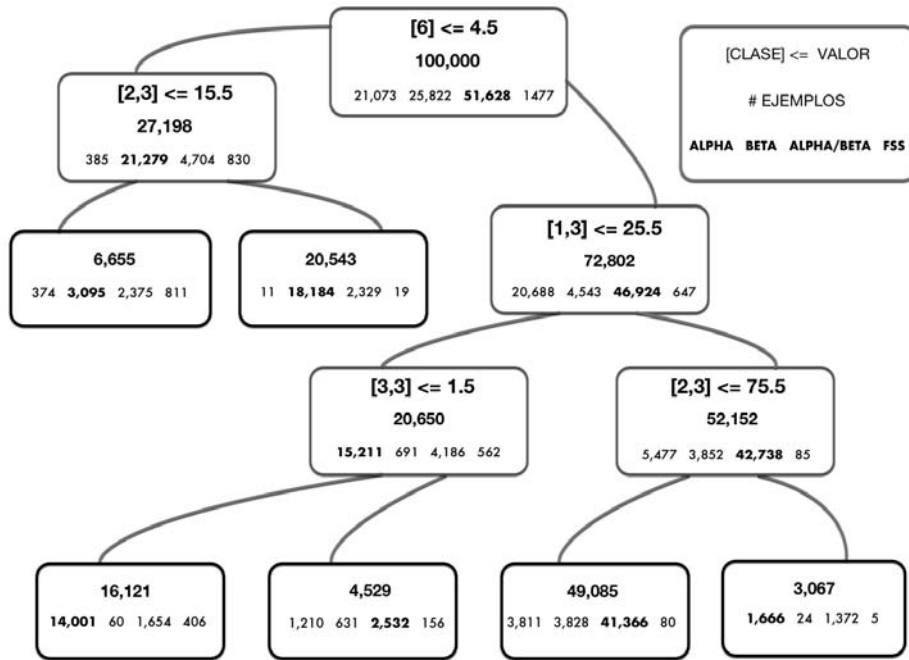


Figura 2.1: **Árbol de decisión para clasificación estructural de proteínas**

En la parte superior de cada nodo no terminal, se muestra una regla a cerca del valor en cada uno de los 26 atributos de clases de cúmulos de residuos. La parte media contiene el número de ejemplos totales que cumplen las reglas de los nodos previos. En la parte inferior se muestra la cantidad de ejemplos observados en cada clase en el conjunto de entrenamiento. La clase estimada en cada nodo corresponde al grupo con mayor cantidad de ejemplos.

datos de entrenamiento, por lo que suponen un problema sobre datos nuevos. Adicionalmente, se ha observado que pueden generarse diversos árboles de decisión con desempeño semejante. Estas ideas han dado origen a los ensambles de árboles de decisión, o *bosques*, donde varios árboles son entrenados independientemente y la tarea de clasificación consiste en obtener el consenso de dicho conjunto de árboles [29]. Empríricamente se ha observado que la heterogeneidad de los atributos y valores de corte escogidos por cada árbol ofrece un mejor desempeño y la mejor manera de asegurar dicha heterogeneidad es elegir conjuntos de atributos y valores de corte aleatorios. Esta estrategia es referida como bosque de árboles extremadamente aleatorios. Una descripción detallada de este método puede ser encontrada en [30].

2.2. Alineamiento de ambientes estructurales

El alineamiento estructural consiste en relacionar regiones estructurales entre un par de proteínas. En esta sección se muestra una definición de región estructural libre de parámetros y una estrategia para asignar dichas regiones de una estructura de proteína con otra.

2.2.1. Regiones

Un residuo r puede pertenecer a más de un cúmulo de residuos (ver Apéndice B, sección 2.2.1.). Todos aquellos residuos que pertenezcan a un cúmulo que contenga a r son adyacentes a r . Se estudiará la adyacencia de residuos sobre una sola clase de cúmulos \vec{c} .

$$ady_{\vec{c}}(r) = \{r_j : \{r, r_j\} \subset RC_{\vec{c}}, \forall RC_{\vec{c}} \in \mathfrak{R}\}$$

Donde \mathfrak{R} es un sistema de clases de cúmulos de residuos definido en el Apéndice B, sección 2.2.1.

Sea $V(\mathfrak{R})$ el conjunto de residuos en \mathfrak{R} . $U_{\vec{c}} \subseteq V(\mathfrak{R})$ es una región de clase \vec{c} en \mathfrak{R} si para cualquier $X \subset U_{\vec{c}}$ existen $r_i \in X$, $r_j \in X - U_{\vec{c}}$ tal que $r_i \in ady_{\vec{c}}(r_j)$. Sea la relación \sim entre los residuos r_i y r_j si $r_i, r_j \in U_{\vec{c}}$. Entonces \sim es una relación de equivalencia.

La clase de equivalencia $[r]$ de un residuo de aminoácido r esta dada por

$$[r]_{\vec{c}} = \{r_j : r \sim r_j\}$$

Se tiene que para cualquier par $r_i, r_j \in V(\mathfrak{R})$

$$[r_i]_{\vec{c}} = [r_j]_{\vec{c}}$$

ó bien

$$[r_i]_{\vec{c}} \cap [r_j]_{\vec{c}} = \emptyset$$

Con esto se tiene fragmentada toda la proteína en regiones

Conjunto factor

Dada la relación de equivalencia \sim en $V(\mathfrak{R})$, entonces el conjunto factor $V(\mathfrak{R})/\sim$ es el conjunto que consiste de todas las clases de equivalencia de \sim .

Regiones isomorfas

Dado un par de proteínas P_1, P_2 , con sistemas $\mathfrak{R}_1, \mathfrak{R}_2$, se dice que las regiones $\mathcal{U}_i \in V(\mathfrak{R}_1)/\sim, \mathcal{U}_j \in V(\mathfrak{R}_2)/\sim$ son isomorfas (denotado $\mathcal{U}_i \simeq \mathcal{U}_j$) si existe una biyección $f : \mathcal{U}_i \rightarrow \mathcal{U}_j$ tal que para cada par de residuos $a_1, a_2 \in \mathcal{U}_i$ que están en contacto en P_1 , entonces $f(a_1), f(a_2)$ están en contacto en P_2

2.2.2. Alineamiento estructural

Se implementa un método de programación dinámica (principio de optimalidad de Bellman [31]) para alinear residuos, si éstos se encuentran en ambientes estructuralmente equivalentes. Se define un ambiente $\langle r \rangle$ de un residuo r como

$$\langle r \rangle = \bigcup_{\vec{c}_i} [r]_{\vec{c}_i}$$

Se dice que los residuos r_i y r_j tienen ambientes estructuralmente equivalentes si $\langle r_i \rangle \simeq \langle r_j \rangle$. Dadas dos proteínas P_1, P_2 se definen sus conjuntos de residuos con ambientes estructuralmente equivalentes $R_1 = \{r \in P_1 : \exists r' \in P_2, \langle r \rangle \simeq \langle r' \rangle\}$ y $R_2 = \{r \in P_2 : \exists r' \in P_1, \langle r \rangle \simeq \langle r' \rangle\}$. Si $|R_1| = m$ y $|R_2| = n$

Se trata de maximizar la función objetivo:

$$\phi(i, j) = \text{Max} \begin{cases} \phi(i-1, j-1) + 1 & : \langle r_i \rangle \simeq \langle r_j \rangle \\ \phi(i-1, j) \\ \phi(i, j-1) \end{cases}$$

con $\phi(0, j) = \phi(i, 0) = 0$.

El número máximo de residuos alineados es $\phi(m, n)$.

2.3. Ordenamiento de vecinos estructurales

Dada una estructura de proteína p y una colección de estructuras P , la tarea de obtención de vecinos estructurales consiste en obtener una permutación de las estructuras en P ordenadas de acuerdo a la semejanza estructural con p . La permutación inducida depende de una función de puntuación $f : p \times P \rightarrow \mathbb{R}$.

En este trabajo, tanto p y cualquier $q \in P$ son representadas por sus 26 atributos inducidos por sus clases de cúmulos de residuos. Teniendo esta representación se siguen dos estrategias para definir f y obtener el ordenamiento en P dada una estructura *query* p .

Enfoque no supervisado

En el enfoque no supervisado, la medida de similitud entre p y $q \in P$ se define por su distancia eculideana.

Esto es:

$$f(p, q) = \sum_{i=1}^{26} \sqrt{(p_i - q_i)^2}$$

Enfoque supervisado

En este enfoque, es necesario un clasificador estructural $C : P \rightarrow \mathbb{R}^n$ sobre n clases estructurales. C arroja la probabilidad de pertenencia de una estructura a cada una de las n clases.

Se determina la clase i más probable de la estructura *query* q y se ordenan las estructuras de acuerdo a su probabilidad de pertenencia a dicha clase. Esto es, la función de similitud se define como:

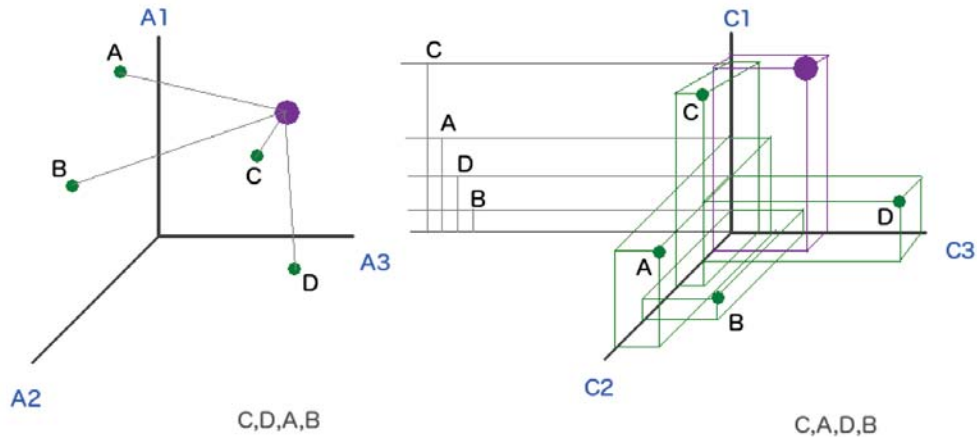


Figura 2.2: **Estrategias de obtención de vecinos estructurales** El enfoque no supervisado (panel izquierdo) consiste en ordenar los vecinos estructurales de acuerdo a la distancia entre las estructuras en el espacio de atributos A1, A2 y A3. El ordenamiento para la estructura *query* en morado es C,D,A,B. El enfoque supervisado (panel derecho) consiste en ordenar las estructuras de acuerdo a la probabilidad de pertenencia a la clase predicha para la estructura *query*. En este caso, las estructuras se representan en el espacio de clases predichas C1, C2 y C3. El ordenamiento final para la estructura *query* es C,A,D,B , que corresponden al ordenamiento en la proyección de las estructuras sobre C1.

$$f(p, q) = C(p)_{\arg \max_i C(q)_i}$$

La Figura 2.2 muestra un esquema comparando ambas estrategias.

2.3.1. Atributos para la representación de secuencias de proteínas

Varios conjuntos de atributos han sido propuestos para capturar las propiedades de las secuencias de proteínas, algunos de naturaleza biofísica como peso molecular medio de los aminoácidos, carga e hidrofobicidad [32] y otras de

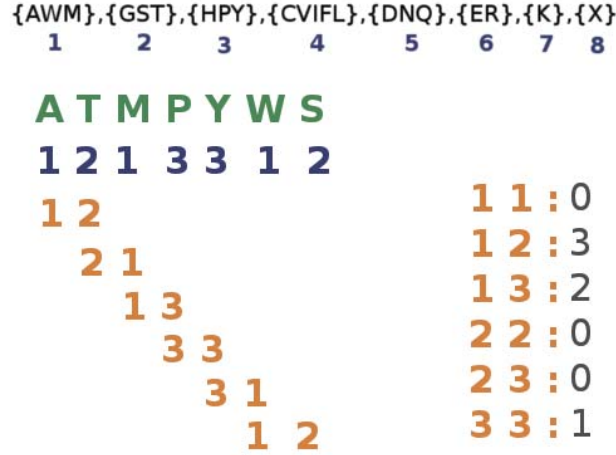


Figura 2.3: **Obtención de atributos desde secuencia** La secuencia se traduce a un alfabeto reducido (números en azul). Luego, se obtienen tuplas de tamaño 2. Todas las permutaciones de una tupla se hacen equivalentes, por ejemplo 12 y 21. Se muestra el conteo de tuplas en el extremo inferior derecho.

naturaleza combinatoria, como la frecuencia de aparición de los aminoácidos o k -péptidos, es decir, fragmentos de secuencias de longitud k [33]. Un inconveniente del uso de k -péptidos es la cantidad exponencial de atributos resultantes, por ejemplo, el número de 2-péptidos posibles son 400, el número de 3-péptidos son 8000 y en general, se necesitan 20^k atributos para representar k -péptidos. Para aliviar el crecimiento en el número de atributos, en este trabajo se consideran equivalentes todas las permutaciones de un k -péptido particular. De esta manera, el número de atributos necesarios para representar cualquier k -péptido es igual al número de combinaciones con reemplazo $\binom{20}{k}$, con

$$\binom{\binom{20}{k}}{k} = \binom{20+k-1}{k} = \frac{(k+19)!}{k!(19!)}$$

Adicionalmente, se utiliza una reducción del alfabeto de aminoácidos. Esta estrategia ha sido utilizada para la construcción de k -péptidos y consiste en agrupar varios aminoácidos y hacerlos equivalentes entre sí [34]. En este trabajo se considera la agrupación F_Ic4 como es utilizada en [35]. La Figura 2.3 muestra un ejemplo de obtención de atributos de una secuencia.

Método para la predicción de vectores estructurales desde secuencia

Predicción de vectores estructurales

El procedimiento para generar un vector en 26 dimensiones desde una secuencia se basa en crear tareas de clasificación independientes para cada componente (posición en el vector).

Dada una colección de proteínas con estructura conocida, los pasos principales para entrenar un predictor de vectores desde secuencia son los siguientes:

1. Crear una base de atributos S para las secuencias
2. Crear una base de vectores V para las estructuras
3. Entrenar un clasificador sobre los atributos en S para predecir cada posición i en los vectores en V (ver Figura 2.4)

Este procedimiento contiene varios parámetros que afectan las predicciones finales. En el paso 1, los parámetros involucrados son el tamaño máximo de tupla considerada, típicamente 3, así como el tipo de mapeo para reducir el tamaño del alfabeto original de 20 a algún número menor (el cual depende del mismo mapeo). Este mapeo es necesario para lidiar con la explosión combinatoria del número resultante de atributos.

En el paso 3, se tiene el clasificador para predecir los valores en cada componente como otro parámetro. Además, cada clasificador contiene parámetros internos adicionales.

Es importante considerar el ajustar todos estos parámetros dependiendo del objetivo de la predicción de los vectores.

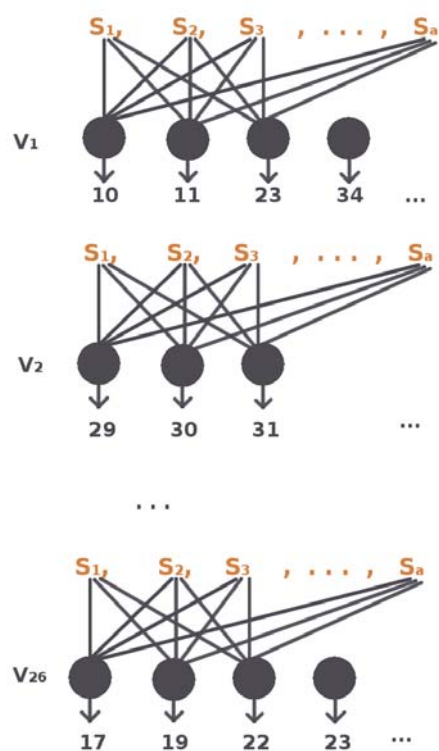


Figura 2.4: **Predicción de vectores en 26 dimensiones desde secuencia.** La predicción del valor de cada uno de los 26 componentes en el espacio de plegados depende de los a atributos de secuencia.

CAPÍTULO 3

Resultados

3.1. Sistema de Clases de Cúmulos de residuos

Para representar cualquier estructura de proteína se utilizó una generalización del concepto de contacto entre un par de residuos a cualquier número de residuos. Estos subconjuntos de residuos en contacto simultáneo son llamados cúmulos de residuos y son distinguidos por su clase, que depende de la contigüidad de los residuos en la estructura primaria; a estos cúmulos también se les conoce como *clanes maximales* (del inglés *maximal clique*). Al conjunto total de cúmulos de residuos en una estructura de proteína se le llama Sistema de Clases de Cúmulos de Residuos (SCCRs). Ya que ningún cúmulo en este sistema es subconjunto de ningún otro, se dice que posee la propiedad de Sperner [36]. Una descripción detallada sobre la definición y construcción de dicho sistema se encuentra en el Apéndice B, sección 2.2.

Alineamiento de ambientes estructurales

El SCCR puede ser utilizado para definir regiones estructurales; conjuntos de residuos pertenecientes a varios cúmulos (2.2.1). El tamaño máximo observa-

do de estas regiones es de 13 aminoácidos, lo que permite hacer comparaciones de isomorfismos basados en algoritmos de teoría de gráficas (ver Métodos 2.2.1). Cada residuo define una región, y en este caso se denomina ambiente estructural de dicho residuo. Por lo tanto, residuos de distintas estructuras pueden ser alineados uno a uno cuando poseen ambientes estructurales equivalentes (isomorfos). Un ejemplo de dos estructuras y sus ambientes estructuralmente equivalentes alineados se muestra en la Figura 3.1.



Figura 3.1: **Visualización de alineamiento estructural basado en regiones.** Alineamiento entre la triosa fosfato isomerasa de *Saccharomyces cerevisiae* (7TIM cadena A, izquierda) y *Homo sapiens* (2VOM cadena A, derecha). Las regiones del mismo color representan ambientes estructuralmente equivalentes.

Aunque las estructuras comparten varios fragmentos con alta similitud, sólo algunos ambientes resultan isomorfos debido a la sensibilidad de los ambientes estructurales a pequeñas diferencias en las cadenas laterales. Los ambientes isomorfos no necesariamente tienen alguna relación con la contigüidad o el orden de sus residuos en su estructura primaria. Algunos ejemplos de secuencias con ambientes isomorfos son los siguientes:

- TIIAAAQAG
KKEMDDAA
- NLLKELG_A
GVADAHF_M
- NKLQAGGY

características estructurales, sin embargo, el conteo de clases de cúmulos sí es capaz de reflejar propiedades estructurales globales y locales, como se mostrará más adelante.

Para inspeccionar visualmente la estructura del espacio de plegados, se utilizó un enfoque no supervisado de reducción de dimensionalidad (Laplacian-Eigenmaps [38]) sobre el conjunto CATCHOP (Apéndice B 2.1.1). La proyección en dos dimensiones se muestra en el Apéndice B Figura 8. En esta representación, los dominios principalmente α y principalmente β se observan en direcciones opuestas, mientras que los dominios $\alpha\beta$ se encuentran entre estos grupos. Para aprender a diferenciar las clases estructurales de las proteínas es necesario evaluar cuantitativamente la separación de estas clases, determinar sus fronteras y determinar si existen distancias características entre estructuras semejantes.

3.3. Propiedades del espacio de plegados

3.3.1. Distancias en el espacio de plegados

Para averiguar si las distancias en esta representación explícita del espacio de plegados en 26 dimensiones reflejan directamente la pertenencia de las estructuras de dominios a una misma clase estructural, se calcularon todas las distancias entre pares de dominios bajo cada nivel de clasificación en CATH. La Figura 3.3 y la Figura 3.4 muestran histogramas de distancias Manhattan y Euclidianas entre todos los pares de dominios bajo los niveles de clasificación Clase (C), Arquitectura (CA), Topología (CAT) y Homología (CATH). Aunque las distancias tienden a ser menores a mayores niveles de clasificación, se observa un gran solapamiento entre las distancias a estos distintos niveles.

Para cuantificar los valores esperados promedio para pares de dominios miembros de una misma clase, se obtuvieron los intervalos de confianza bayesianos para cada nivel de clasificación. Los intervalos de confianza para los niveles C, CA, CAT y CATH son (236.69, 236.81), (189.07, 189.23), (175.05, 175.38) y (126.07, 126.65) respectivamente para la distancia Manhattan, y (88.50, 88.55), (69.86, 69.92), (63.58, 63.70) y (44.84, 45.05) para la distancia Euclídeana (ver Cuadro 3.1). Las distancias promedio son estrictamente mayores en niveles de

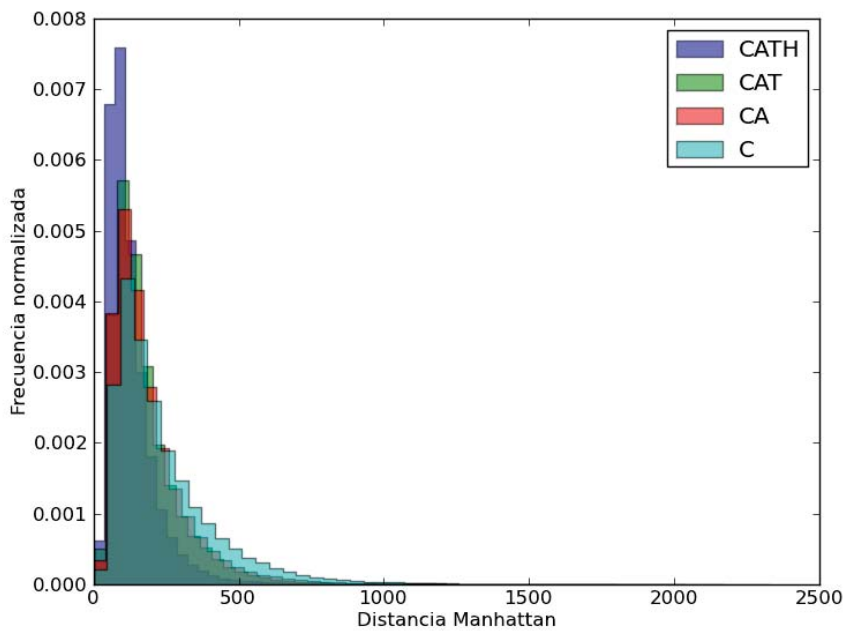


Figura 3.3: **Histogramas de distancias Manhattan** La figura muestra histogramas de distancias Manhattan entre cada par de dominios bajo una misma clasificación a los niveles de Clase (C), Arquitectura (CA), Topología (CAT) y Homología (CATH)

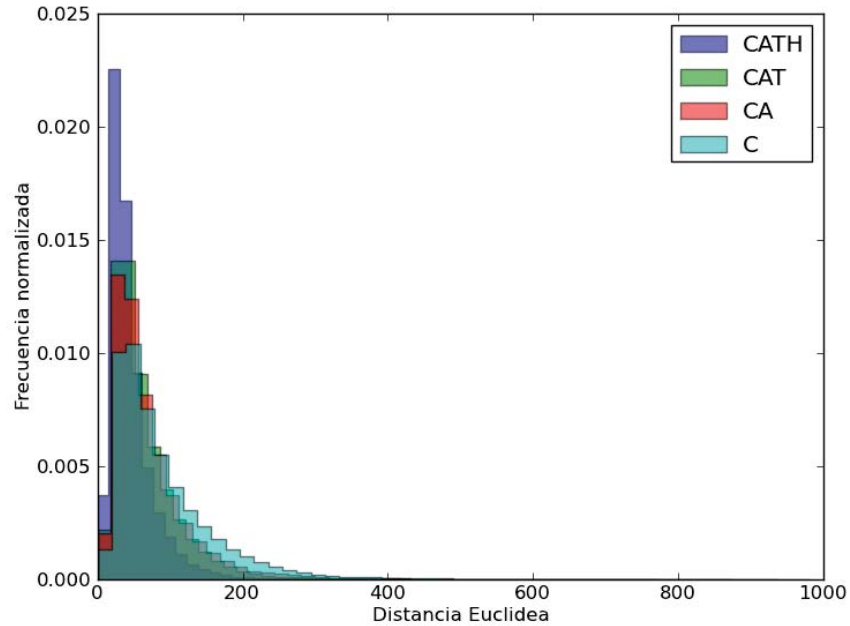


Figura 3.4: **Histogramas de distancias euclidianas** La figura muestra histogramas de distancias Euclidianas entre cada par de dominios bajo una misma clasificación a los niveles de Clase (C), Arquitectura (CA), Topología (CAT) y Homología (CATH)

clasificación más amplios con respecto a niveles más profundos, sin embargo, en todos los casos la desviación en un nivel sobrelapa con la media de los demás, de manera que no es posible determinar la pertenencia de un par de dominios a una misma clase dada únicamente con su distancia.

Distancias entre diversas conformaciones

La sensibilidad del espacio de plegados a similitudes estructurales locales fue analizada usando pares de conformaciones en ensambles de NMR. Estas estructuras pertenecen a una misma proteína sometida a restricciones físicas y varían generalmente solo en algunos puntos a lo largo de la cadena principal. Se calcularon las distancias coseno en el espacio de plegados y la puntuación TM-

Nivel	I.C. media $p = 1$	I.C. d.e. $p = 1$	I.C media $p = 2$	I.C. d.e. $p = 2$
CATH	(126.07, 126.65)	(87.89, 88.30)	(44.84, 45.05)	(32.48, 32.64)
CAT	(175.05, 175.38)	(126.68, 126.91)	(63.58, 63.70)	(47.04, 47.13)
CA	(189.07, 189.23)	(136.87, 136.99)	(69.86, 69.92)	(52.68, 52.72)
C	(236.69, 236.81)	(172.81, 172.89)	(88.50, 88.55)	(67.31, 67.35)

Cuadro 3.1: Intervalos de confianza bayesianos El cuadro muestra los intervalos de confianza bayesianos para la media y desviación estandar (d.e.) para las distancias Manhattan ($p = 1$) y Euclideana ($p = 2$) para los pares de dominios bajo cada nivel de clasificación Clase (C), Arquitectura (CA), Topología (CAT) y Homología (CATH)

score [39] basada en RMSD utilizando [40] entre cada par de conformaciones en 230 ensambles de NMR (mismo conjunto de datos evaluado en [23]). La Figura 3.5 muestra el resumen de este experimento; cada punto representa una comparación entre un par de conformaciones de un mismo ensamble, posicionado según su similitud medida por TM-align y la distancia coseno en el espacio de plegados. En esta figura se observa que en una gran cantidad de casos, la distancia entre pares de conformaciones es menor a la distancia coseno media de 0.058 observada entre dominios pertenecientes a una misma topología.

Un ejemplo de comparación entre la similitud reflejada por TM-score y la distancia euclideana en el espacio de plegados entre conformaciones de un mismo ensamble y estructuras diferentes bajo una misma topología se muestra en la Figura 3.6.

3.3.2. Separabilidad del espacio de plegados

Para determinar si los dominios en el espacio de plegados pueden ser segregados en clases estructurales por algún tipo de frontera de decisión se realizó la prueba de serparabilidad usando el estadístico *cut edge weight statistic* [41]. El proceso de prueba de separabilidad con este estadístico, consistió en la prueba de hipótesis de que el estadístico obtenido bajo la asignación de clases según CATH a nivel de Topología es distinto al estadístico bajo la hipótesis nula cons-

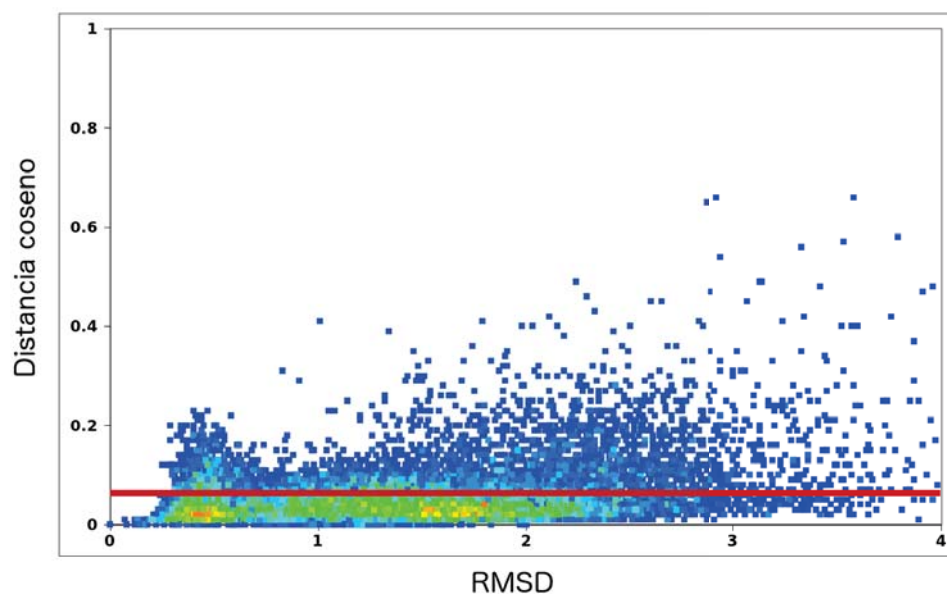


Figura 3.5: **Relación entre RMSD y distancia coseno entre conformaciones de una misma proteína.** Cada punto representa una comparación entre un par de conformaciones de un mismo ensamble de NMR, su posición en el eje de abscisas es la puntuación TM-score y su posición en el eje de ordenadas es la distancia coseno en el espacio de plegados. El color indica el número de pares encontrados con mediciones equivalentes. La línea roja muestra el intervalo de confianza bayesiano (0.055, 0.0613) para la media de distancias coseno entre dominios miembro de una misma topología.

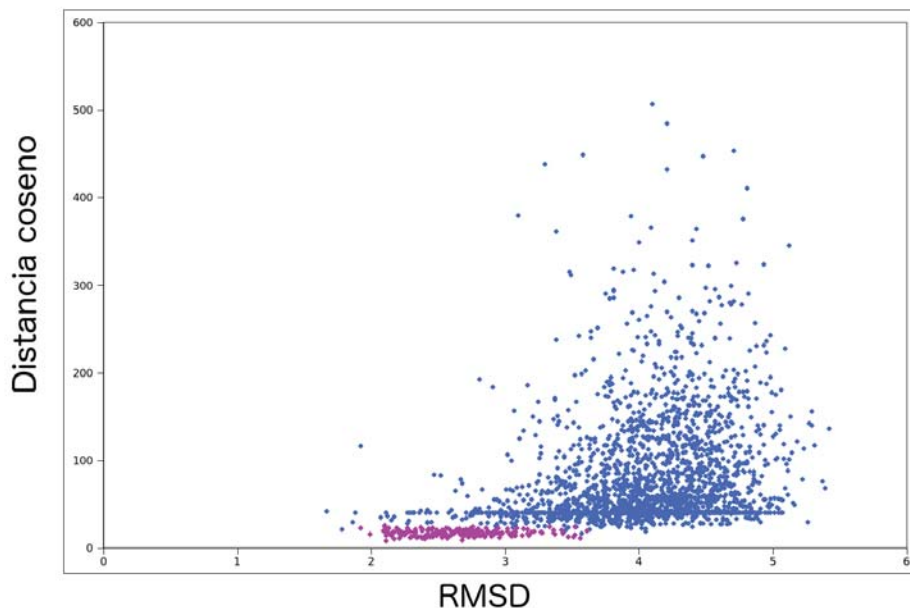


Figura 3.6: **Relación entre RMSD y distancia coseno entre un ensamble de NMR y dominios bajo la misma topología.** Cada punto representa una comparación entre un par de conformaciones de un mismo ensamble de NMR, su posición en el eje de abscisas es la puntuación TM-score y su posición en el eje de ordenadas es la distancia euclídeana en el espacio de plegados. En rosa se muestran pares de conformaciones del ensamble de NMR para el factor de elongación TFIIS. En azul se muestran pares de estructuras de dominios bajo la misma Topología.

truida asignando una clasificación al azar. El resultado de la prueba ($p = 0$), muestra que los dominios pertenecientes a distintas topologías son separables.

3.4. Predicción de clasificación estructural

Para hacer uso de la separabilidad de esta representación del espacio de plegados, se entrenó un clasificador basado en un bosque extremadamente aleatorio (ERF, por sus siglas en inglés) para asignar clasificaciones automáticas por clase, arquitectura y topología, de acuerdo a la definición de CATH. Se realizó

una validación cruzada estratificada a 10 iteraciones para evaluar la capacidad del clasificador ERF para aprender reglas generales del conjunto de datos CATHALL1 obteniendo eficacias de 96.93 %, 88.3 % y 84.94 % para los niveles de Clase, Arquitectura y Topología respectivamente. También se entrenó el clasificador ERF sobre todo el conjunto de datos obteniendo eficacias mayores al 99 %, (Apéndice B Tabla 1), indicando que prevalecen algunas discrepancias entre las clases predichas y las clases anotadas en CATH (Apéndice B Figura 7). La capacidad predictiva fue evaluada de la misma forma sobre dominios según SCOP con resultados semejantes (Apéndice B Tabla 3).

3.5. Obtención de vecinos estructurales

Se ha reportado que ContactLib [42] obtiene el mejor desempeño en encontrar vecinos estructurales al compararse contra el mejor método libre de alineamientos FragBag [43]. Por otro lado, FragBag ha reportado mejoras sobre seis métodos de alineamiento estructural ampliamente utilizados, incluyendo SGM, PRIDE, BLAST, STRUCTAL, CE and SSM. FragBag fue probado en un subconjunto de dominios en SCOP 1.75B, referido como SCOP30. Tres puntos de corte SAS20, SAS35 y SAS50 identifican tres tipos de vecindades a 2.0, 3.5 y 5Å respectivamente. SAS20 es el más sensible y SAS50 el más relajado. La concordancia entre el criterio SAS y asignación de superfamilia según SCOP es baja (Apéndice B Tabla 4). Esta discrepancia puede explicarse con el ejemplo de la la Figura 3.7, que muestra un ejemplo de estructuras alineables localmente pero con arquitecturas distintas. Por lo tanto, solo los dominios en cada vecindad SAS pertenecientes a la misma superfamilia definida por SCOP son considerados vecinos reales. Se calculó el área bajo la curva ROC (AUROC), para probar la capacidad de la representación del espacio de plegados para identificar vecinos reales de acuerdo a los criterios SAS y SCOP; este enfoque fue escogido para resumir los resultados obtenidos para todos los dominios SCOP analizados. Una puntuación AUROC de 1,0 es una puntuación perfecta, y una puntuación de 0,5 es esperada de un método aleatorio. Se calculó el AUROC de cada estructura en SCOP30 (3,295 dominios; ver Apéndice B, sección 2.1.2) usando enfoques supervisados y no supervisados (ver Métodos).



Figura 3.7: **Alta puntuación de alineamiento no implica pertenencia a una misma clase estructural** La figura muestra las estructuras del dominio péptido inhibidor ia3 (d1dpjb_) y el dominio mio-inositol monofosfatasa (d2bjia_), estos pueden ser alineados bajo SAS20, ya que toda la estructura de d1dpjb_ puede ser sobrepuesta con una alfa hélice en d2bjia_, sin embargo la arquitectura global es radicalmente distinta.

En el caso no supervisado, las estructuras en SCOP30 fueron ordenadas a partir de una estructura *query* usando la distancia euclidiana; esta lista ordenada fue comparada con los vecinos reales provistos por los criterios SAS y SCOP para cuantificar la correspondiente AUROC. Las AUROCs medias para este experimento fueron 0.87, 0.85 y 0.84 para SAS 20, 35 y 50 respectivamente. Por comparación, nótese que la mejor puntuación obtenida en este experimento por FragBag es 0.747 para SAS20 [23].

Un ejemplo de estructuras vecinas obtenidas mediante este criterio se muestra en la Figura 3.8. Aquí se muestran las cadenas de proteína contenidas en el PDB cercanas a la estructura de la levansacarasa de *Bacillus subtilis* (PDB ID 1OYG, cadena A). Se observa que las estructuras obtenidas pertenecen a la misma proteína en presencia de diferentes ligandos, a la proteína homóloga en

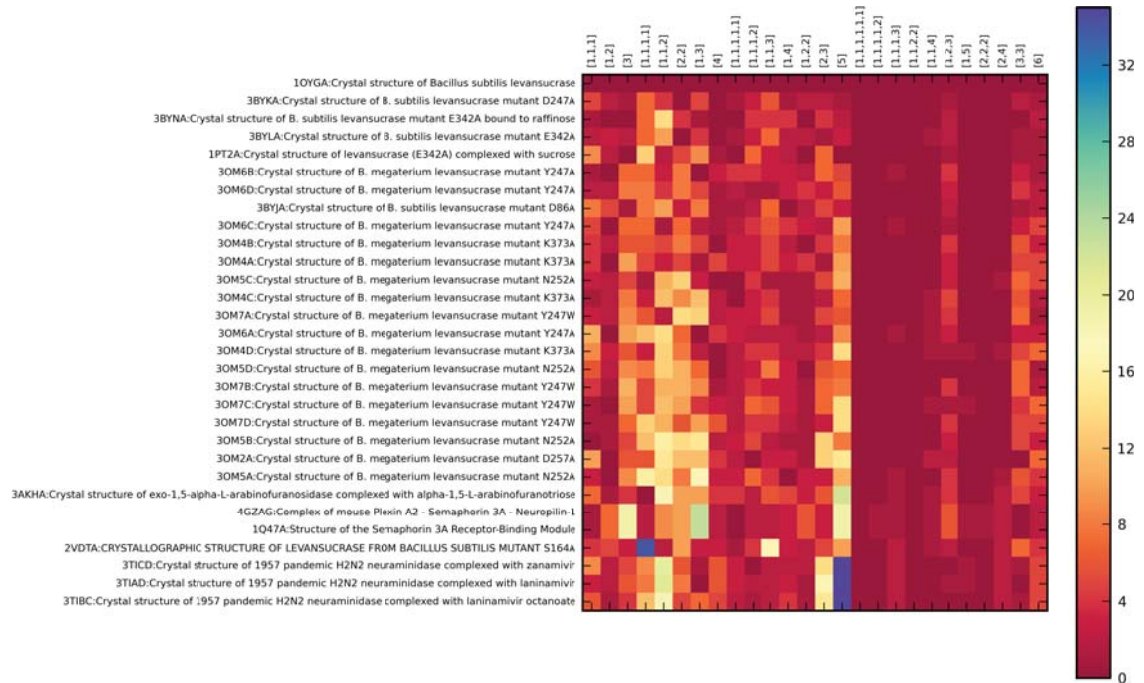


Figura 3.8: Estructuras en el PDB cercanas a la levansacarasa de *Bacillus subtilis* (1OYG cadena A). Se muestra en código de colores las diferencias en los valores de cada uno de los 26 componentes de la estructura *query* y las estructuras cercanas. La primer estructura encontrada es la misma estructura *query*, de manera que la diferencia en todas las componentes es 0.

B. megaterium y algunas mutantes de estas. Adicionalmente se observa algo de cercanía con estructuras de semaforinas y neuraminidasas.

Estas distintas familias estructurales pueden ser identificadas por agrupamiento jerárquico como se muestra en el dendrograma de la Figura 3.9.

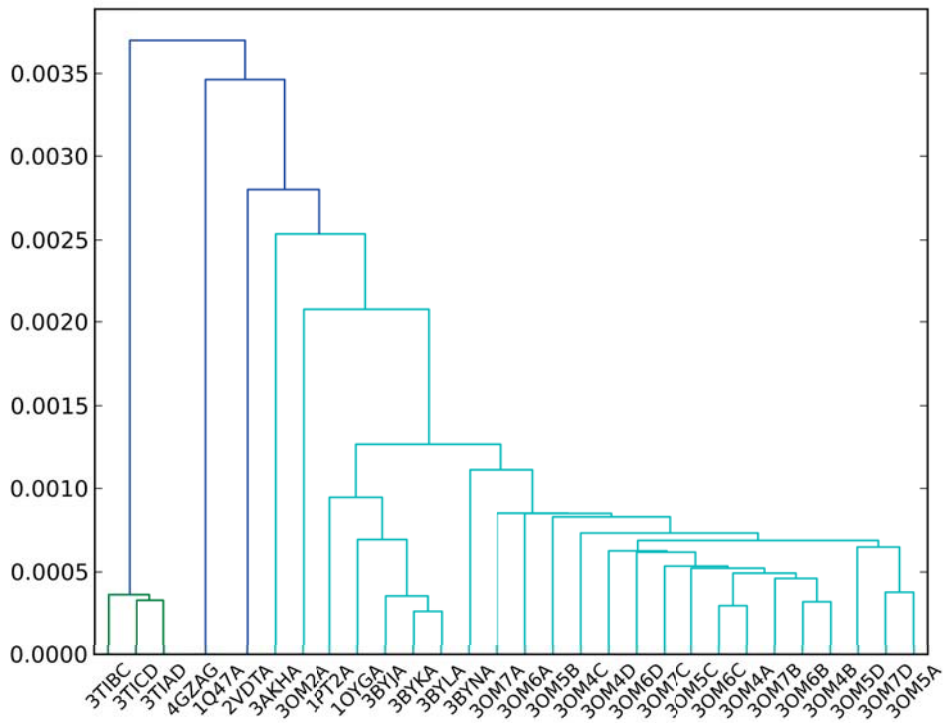


Figura 3.9: **Denograma de vecinos cercanos a la levansacarasa de *Bacillus subtilis*.** Dendrograma construido por agrupamiento jerárquico de las estructuras cercanas a la estructura 1OYG cadena A. Se observa la segregación de familias estructurales de levansacarosas (cian), neuraminidasas (verde) y semaforinas (azul).

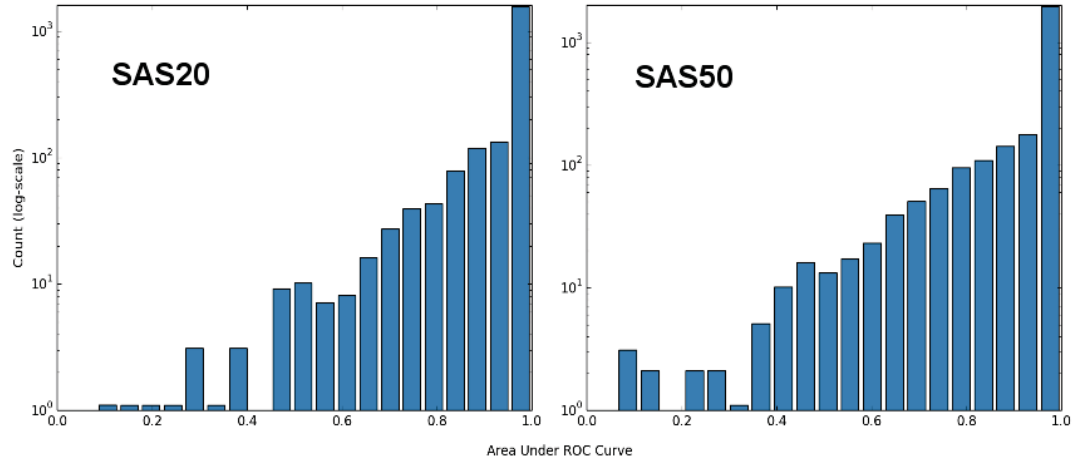


Figura 3.10: **Histogramas de puntuaciones AUROC** Histogramas en escala logarítmica de las puntuaciones AUROC obtenidas en la tarea de obtención de vecinos estructurales a los niveles SAS20 y SAS50.

En el análisis supervisado, AUROCs fueron obtenidas entrenando un clasificador ERF con un subconjunto de SCOP excluyendo las estructuras en SCOP30 (ver Apéndice B sección 2.1.2). Entonces, para una estructura query se predice su clase, y el ordenamiento del resto de los dominios es determinado por las log-probabilidades de pertenecer a la clase predicha. Las puntuaciones AUROC obtenidas con este enfoque para SAS 20, 35 y 50 son 0.950, 0.940 y 0.935, mientras ContactLib alcanza 0.956, 0.918 y 0.906. ContactLib reporta que el 75 % de sus puntuaciones son mayores a 0.935, mientras que con este enfoque, la puntuación AUROC más baja para el 75 % de las estructuras es 0.958. Para SAS35 y SAS50, 75 % de las puntuaciones AUROC son mayores a 0.939 y 0.929 respectivamente. Con este enfoque, la puntuación AUROC mediana para cualquier nivel SAS es mayor a 0.99 (véase Apéndice B Tabla 4 para un resumen de estas puntuaciones). La Figura 3.10 muestra estas puntuaciones AUROC en escala logarítmica para los niveles SAS20 y SAS50.

3.6. Predicción de plegados a partir de secuencia

En este apartado se propone una representación de secuencias de proteínas mediante un número fijo de atributos y se explora un enfoque de aprendizaje profundo que incorpora información de secuencia y atributos estructurales para la predicción de plegados. En particular, este estudio se centra en evaluar la predicción de plegado a partir de secuencias muy divergentes, es decir, secuencias con poca o nula similitud detectable a secuencias conocidas.

3.6.1. Predicción de vectores de clases de cúmulos de residuos desde secuencia

Como se ha mencionado previamente, los enfoques basados en similitud en secuencia para predicción de plegado dependen de la disponibilidad de suficientes secuencias y estructuras semejantes. En consecuencia, la disponibilidad de datos de secuencias y estructuras conocidas pero con poca semejanza a una secuencia de interés no puede ser incorporada por estos métodos.

En esta sección se presenta una representación del espacio de secuencias independiente de alineamiento y parcialmente independiente del orden (ver Métodos 2.3.1). Con esta representación se explora la capacidad de predecir correctamente el plegado de una proteína a partir de su secuencia mediante la estimación del vector en 26 dimensiones que la representa (SCCRs; ver sección 2.2).

El poder realizar una *buena aproximación* de este vector a partir de la secuencia de una proteína puede ser utilizado directamente para responder algunas de las preguntas que se pueden responder con la estructura completa. Anticipamos los siguientes cuatro escenarios para el resultado de estas predicciones de vectores.

A La posición de los vectores predichos desde secuencia es igual a la de los vectores *reales* derivados directamente de la estructura.

B La posición de los vectores predichos desde secuencia es cercana a la de los vectores *reales* derivados directamente de la estructura. Sin embargo, la

distribución de los vectores mantiene las ventajas de la distribución *real* para aprendizaje automático.

- C La posición de los vectores predichos desde secuencia es muy distinta a la de los vectores *reales* derivados directamente de la estructura. Además, la distribución de los vectores no mantiene las ventajas de la distribución *real* para aprendizaje automático.
- C La posición de los vectores predichos desde secuencia puede o no ser cercana a la de los vectores *reales* derivados directamente de la estructura. Sin embargo, la distribución de los vectores supera las ventajas de la distribución *real* para aprendizaje automático.

La Figura 3.11 muestra una representación de estos cuatro casos. Cada punto representa un vector predicho desde una secuencia, y el color representa alguna clase estructural (conocida únicamente a partir de la estructura explícita tridimensional)

El procedimiento para la predicción de vectores estructurales desde secuencia se describe en Métodos 2.3.1

Evaluación de predicciones de vectores estructurales

Para evaluar la bondad de la distribución de los vectores predichos, se utilizaron estos vectores como atributos para tareas de clasificación estructural. Se utilizaron 168,964 dominios definidos en CATH versión 3.5. Se crearon atributos para las secuencias basados en combinaciones con remplazo para fragmentos de secuencia de longitud 2 y 3 sobre un alfabeto de aminoácidos reducido descrito anteriormente como *F₇C4* [35].

Se entrenó para una clasificación binaria con regresión logística para cada valor posible en cada componente del vector estructural.

Se evaluó la eficacia de las predicciones mediante validación cruzada a diez iteraciones con un ensamble de 25 árboles extremadamente aleatorios.

Esta evaluación se realizó sobre los vectores estructurales reales y sobre los atributos de secuencia crudos.

El Cuadro 3.2 muestra las calificaciones obtenidas por cada conjunto de datos.

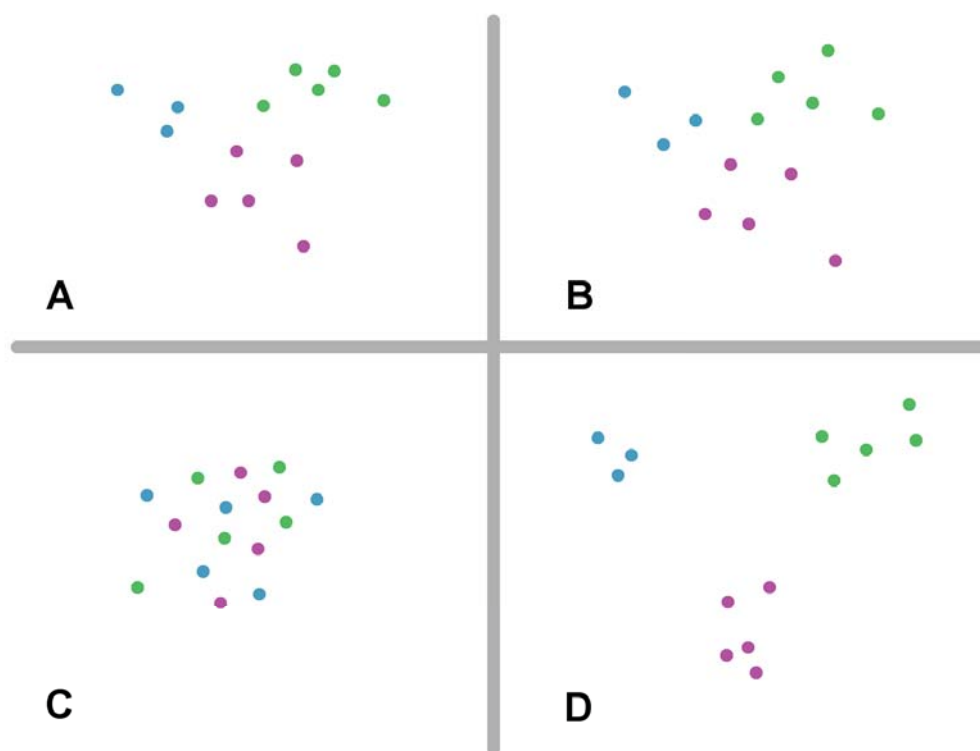


Figura 3.11: Escenarios posibles para la predicción de vectores estructurales a partir de secuencia. A) Predicción perfecta de posiciones / buena separabilidad, B) Predicción buena de posiciones / buena separabilidad, C) Mala predicción de posiciones / mala separabilidad, D) Mala predicción de posiciones / muy buena separabilidad.

Cuadro 3.2: Eficacia de predicciones sobre diferentes tipos de atributos

nivel CATH/dataset	v26d real	v26d predicho	atributos de secuencia
C	0.966	0.968	0.985
CA	0.882	0.950	0.977
CAT	0.838	0.943	0.974

Calificaciones de validación cruzada para predicción de clasificaciones estructurales CATH sobre los vectores estructurales reales, vectores predichos y sobre atributos crudos de secuencia.

Comparación de vectores reales y predichos

Como se mostró en la sección anterior, la distribución de los vectores predichos desde secuencia (vectores predichos) resulta más bondadosa para la tarea de predicción de clasificación estructural, que los vectores obtenidos desde estructura (vectores reales). Sin embargo, este resultado no da información acerca del parecido entre los espacios donde se encuentran estos vectores.

Cada vector está definido por su sentido y magnitud, de manera que una forma de comparar el parecido entre los puntos provenientes de estructura y los predichos desde secuencia, es comparar los cambios en sus magnitudes y su sentido, resumido en el ángulo formado por ambos puntos.

La figura 3.12 muestra los parecidos en estas propiedades dada una selección particular de parámetros. Aquí, cada punto refleja el parecido entre los vectores que representan a un dominio en particular. La posición depende de la magnitud del vector real y predicho que representa a tal dominio. El color es la distancia coseno entre ambos vectores. Si dos vectores a y b están separados por un ángulo θ , la distancia coseno está definida como

$$1 - \cos(\theta) = 1 - \frac{a \cdot b}{|a||b|}$$

Nótese que los vectores con igual magnitud caen sobre la diagonal de esta figura.

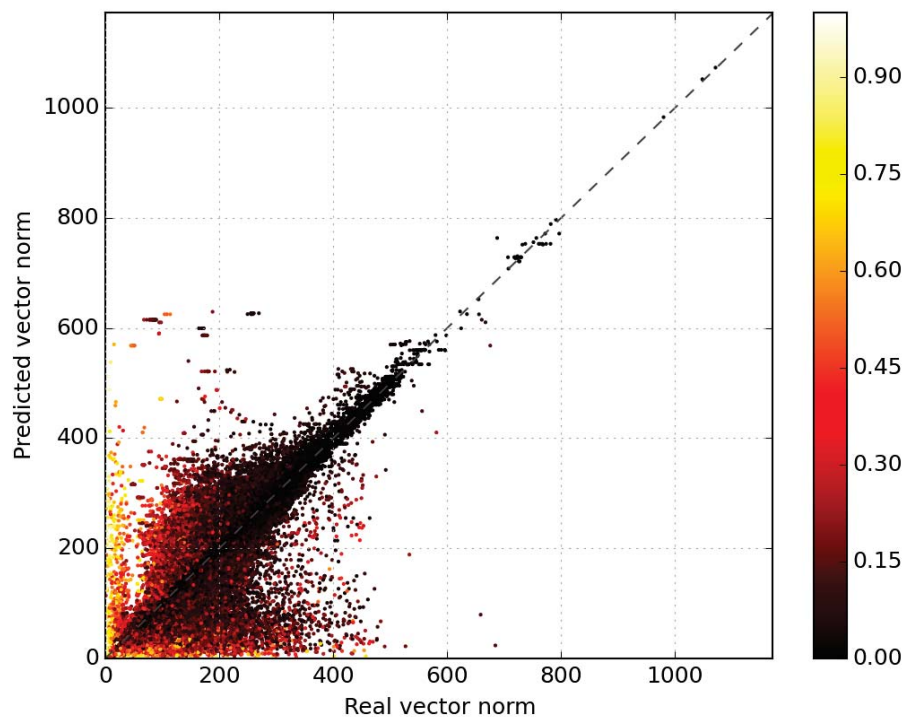


Figura 3.12: Parecido entre las posiciones de los vectores predichos y reales. Cada punto representa un dominio. La posición está dada por las normas de su vector real y su vector predicho. El color está dado por la distancia coseno que los separa. Una distancia 0 implica un ángulo 0, mientras que una distancia de 1 implica que ambos vectores están separados por un ángulo $\frac{\pi}{2}$

3.6.2. Estado actual en la predicción de plegado desde secuencia

La predicción del plegado de una proteína a partir de la información de secuencia permanece como un problema abierto, en el sentido en que se desconocen los detalles que rigen la relación secuencia-estructura. En la práctica, es posible obtener buenas predicciones de plegado a partir de secuencia si se conocen las estructuras resueltas para una familia de secuencias semejantes a

la secuencia de interés. Sin embargo, la predicción de plegado de una secuencia sin parecido claro a alguna proteína de estructura conocida continúa siendo una tarea mucho más complicada con resultados menos satisfactorios. La comunidad de predicción de estructura distingue este par de tareas como predicción de estructura basada en templados (template-based prediction) y como modelado libre (free-modeling prediction).

La evaluación de las predicciones basadas en templados consiste en comparar la estructura real y la predicha mediante algún método de alineamiento estructural, sin embargo, en el caso de las predicciones libres, la estructura predicha no resulta tan parecida a la estructura real. El poder de los métodos automatizados para distinguir características estructurales locales de predicciones estructurales globales resultan no ser discriminativos, de manera que la evaluación de la calidad general de las predicciones requiere inspección manual [44].

En el caso de modelado libre existen dos posibilidades principales; la secuencia no comparte similitud en secuencia detectable con secuencias conocidas, pero adopta un plegado conocido, o bien adopta un nuevo plegado. La distinción entre estos casos también requiere inspección visual.

Estado actual de la eficacia de predicciones en modelado libre

A lo largo de estos experimentos CASP[1] se pueden notar algunas tendencias a cerca del desempeño de los enfoques para la predicción de modelado libre. Primero, se observa desde los experimentos CASP6 y CASP7 (años 2004 y 2006) que la precisión de los métodos medida por la puntuación GDT_TS [45], decrece al aumentar el tamaño de la secuencia objetivo [46], como se muestra en la Figura 3.13.

Por otra parte, el tamaño medio de las secuencias objetivo ha disminuido en cada experimento. Sabiendo que secuencias más pequeñas obtienen mejores puntuaciones, y que en cada nuevo experimento el tamaño medio de los objetivos es menor [47] como se muestra en la Figura 3.14, no es evidente que los métodos de predicción de estructura hayan realmente mejorado como se puede suponer al revisar los resultados en los experimentos más recientes. Por ejemplo, reconociendo que las proteínas pueden poseer varios dominios, es posible que la mejoría en las predicciones de la estructura de proteínas se deba en parte a

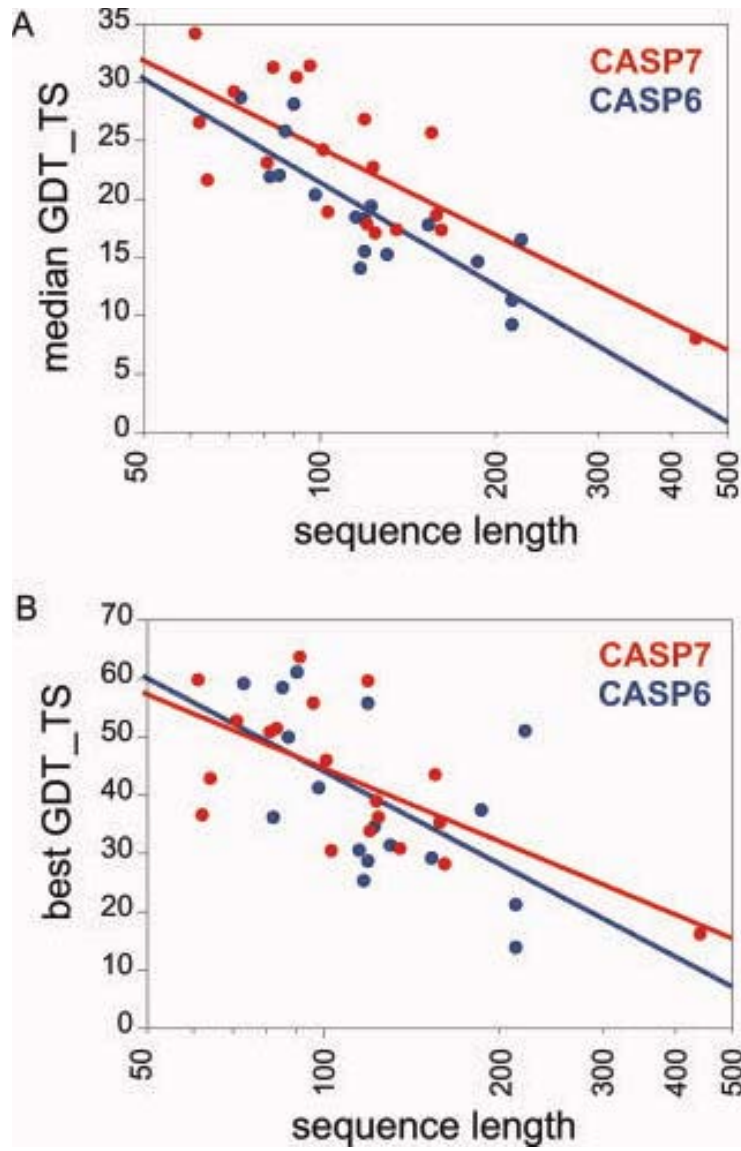


Figura 3.13: Puntuaciones GDT_TS medias como función de la secuencia objetivo. Figura tomada de [46].

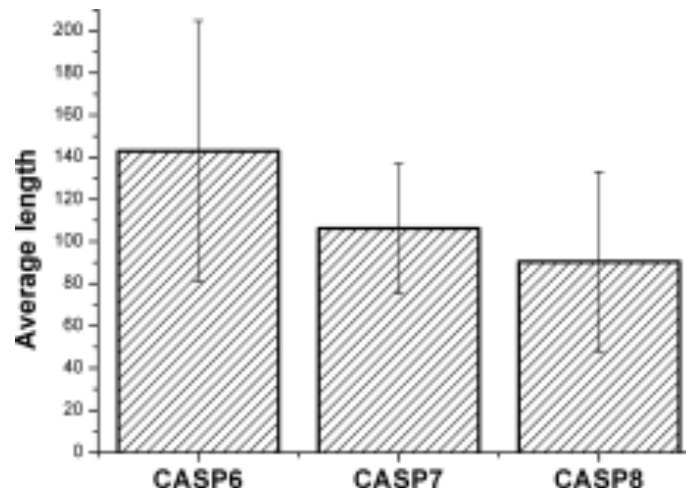


Figura 3.14: Comparación de la longitud de secuencia promedio para los modelos libres de templado en CASP6, CASP7 y CASP8. Figura tomada de [47].

haber eliminado la necesidad de predecir correctamente los dominios presentes en una secuencia.

Una característica general de los modelos libres de templado predichos es su baja calidad en términos físicos. Independientemente de capturar las características globales del plegado real, los modelos tienden a no formar incluso correctamente algunos elementos de estructura secundaria, como los puentes de hidrógeno típicos en las hojas beta [44]. Esto sugiere que es necesario adoptar nuevas estrategias que contribuyan a mejorar las predicciones de la estructura de proteínas, principalmente en el modelado libre de templado.

Dificultad de las secuencias objetivo

Distintas proteínas poseen diferentes niveles de dificultad para su modelado, por lo que cualquier comparación entre la capacidad predictiva puede carecer de sentido sin su relación con esta dificultad [4]. Una medida de dificultad para una secuencia objetivo puede ser la semejanza que posee con el templado más cercano.

La Figura 3.15 muestra la distribución entre la dificultad de los objetivos para todos los experimentos CASP, como una función de la similitud en estructura y

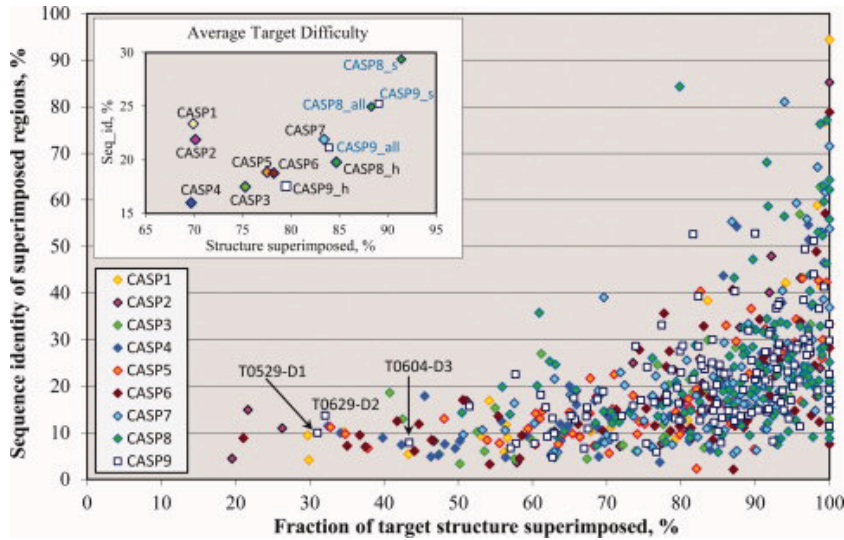


Figura 3.15: Distribución de la dificultad de las secuencias objetivo. Figura tomada de [4].

secuencia entre la estructura experimental de cada objetivo y el mejor templado disponible.

La Figura 3.16 muestra las puntuaciones del mejor modelo predicho para cada objetivo en función de la dificultad.

3.6.3. Evaluación de predictores de plegado de secuencias divergentes

Esta sección aborda una estrategia para la predicción de plegado de secuencias con baja similitud a las secuencias conocidas. El plegado en este contexto es la predicción de la clasificación estructural a nivel de arquitectura o topología según CATH a partir de la secuencia de una proteína.

Como se mostró en la Tabla 3.2, el desempeño de un predictor de clasificaciones estructurales basado solo en atributos de secuencia (espacio de secuencias) resulta superior al predictor sobre el espacio de plegados en 26 dimensiones basado en las clases de cúmulos de residuos (espacio de plegados real). El predictor basado en la estimación del espacio en 26 dimensiones desde atributos de secuencia (espacio de plegados predicho) muestra un desempeño intermedio.

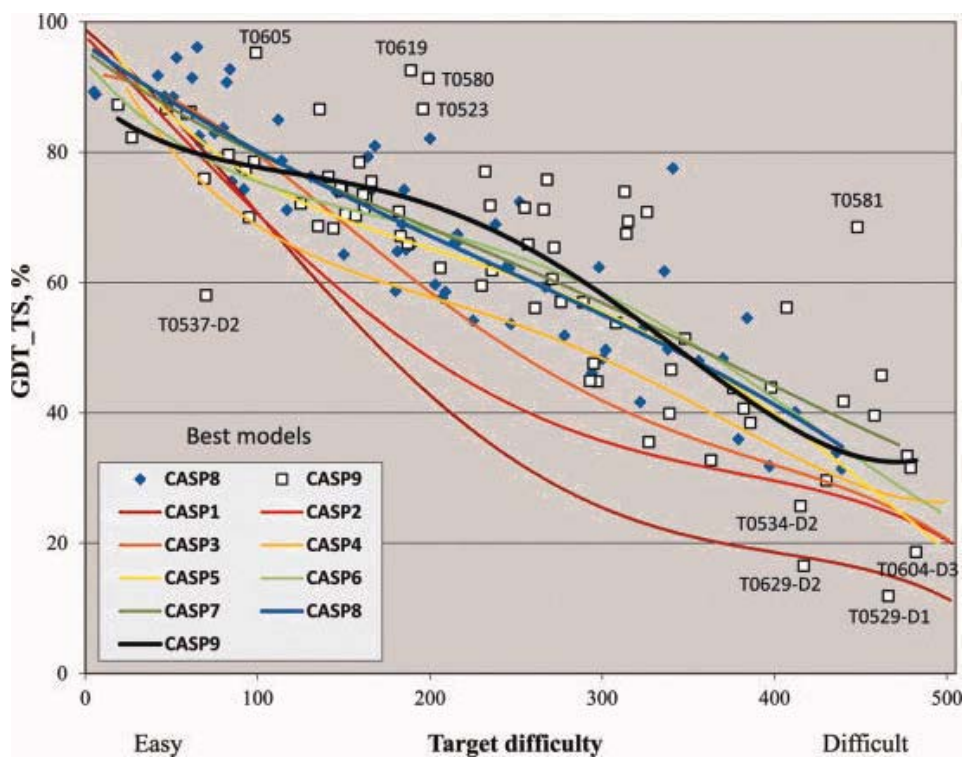


Figura 3.16: Puntuaciones GDT_TS del mejor modelo en función de la dificultad de la secuencia objetivo. Figura tomada de [4].

De estos resultados surgen dos preguntas principales; primero, ¿por qué el desempeño del predictor sobre el espacio de secuencias supera al predictor sobre el espacio de plegados real?, y segundo, ¿para qué considerar entonces el espacio de plegados real y predicho en lugar de sólo considerar el espacio de secuencias? Una respuesta a la primer pregunta puede ser que cada familia resulta en una región bien definida y segregada en el espacio de secuencias, pues sus secuencias miembro tienen gran semejanza, de manera que los vectores de atributos que las representan están bastante cercanos. Aunque cada familia perteneciente a una misma clase estructural se encuentre lejana una de la otra, las reglas para separar cada una independientemente son sencillas y se pueden aprender por separado.

En el caso del espacio de plegados real, las secuencias que adoptan un mismo plegado convergen en una misma región con fronteras de decisión más complicadas. Es decir, la motivación para considerar el espacio de plegados y no únicamente el espacio de secuencias proviene justo de los puntos recién mencionados.

El objetivo de aprender el mapeo del espacio de secuencias al espacio real, es lograr hacer converger familias de secuencias potencialmente distantes a una única región. Las reglas de mapeo se extraen de todas las secuencias hacia todas las clases estructurales, es decir, ninguna información acerca de la pertenencia a alguna clase estructural es tomada en cuenta al aprender este mapeo.

Todo esto implica que la capacidad de mapear correctamente cualquier secuencia a su región en el espacio de plegados correspondiente implica la capacidad de predecir clasificaciones estructurales de secuencias aún sin contar con ejemplos semejantes.

La evaluación de la capacidad predictiva de topología a partir de secuencia se realizó con una familia de secuencia a la vez. Las familias son definidas por su clasificación H en CATH.

En cada evaluación, todas las secuencias pertenecientes a una familia son eliminadas del conjunto de entrenamiento, luego se realizan las predicciones de topología para cada una de las secuencias removidas. La puntuación de eficacia es la fracción de estas secuencias con una predicción correcta sobre el número de secuencias de la familia. La media de eficacia obtenida para el enfoque basado en secuencia (seq) y para el enfoque de aproximación al espacio de plegados

(rcc) es 0.131 y 0.113 respectivamente. Si se considera en cada caso la mejor predicción de seq y rcc, se obtiene una media de 0.16. Muchas de las eficacias son 0.0, es decir, ninguno de los miembros de la familia evaluada fue predicho con su topología correspondiente. La media de las predicciones para seq y rcc excluyendo ceros es de 0.519 y 0.447 respectivamente. La media de las mejores predicciones excluyendo ceros es de 0.55

La Figura 3.17 muestra un histograma comparativo de las eficacias de ambos enfoques. Aunque la distribución global de eficacias es muy similar, las eficacias son muy distintas en cada familia como se muestra en la Figura 3.18. Se observa poca correlación entre las eficacias de ambos enfoques. Los valores de las diferencias entre eficacias se muestran en la Figura 3.19.

Esta estrategia fue utilizada para predecir la topología de las secuencias objetivo del experimento CASP12. Los resultados aún no pueden ser evaluados, pero las predicciones se muestran en el Apéndice A, sección A.

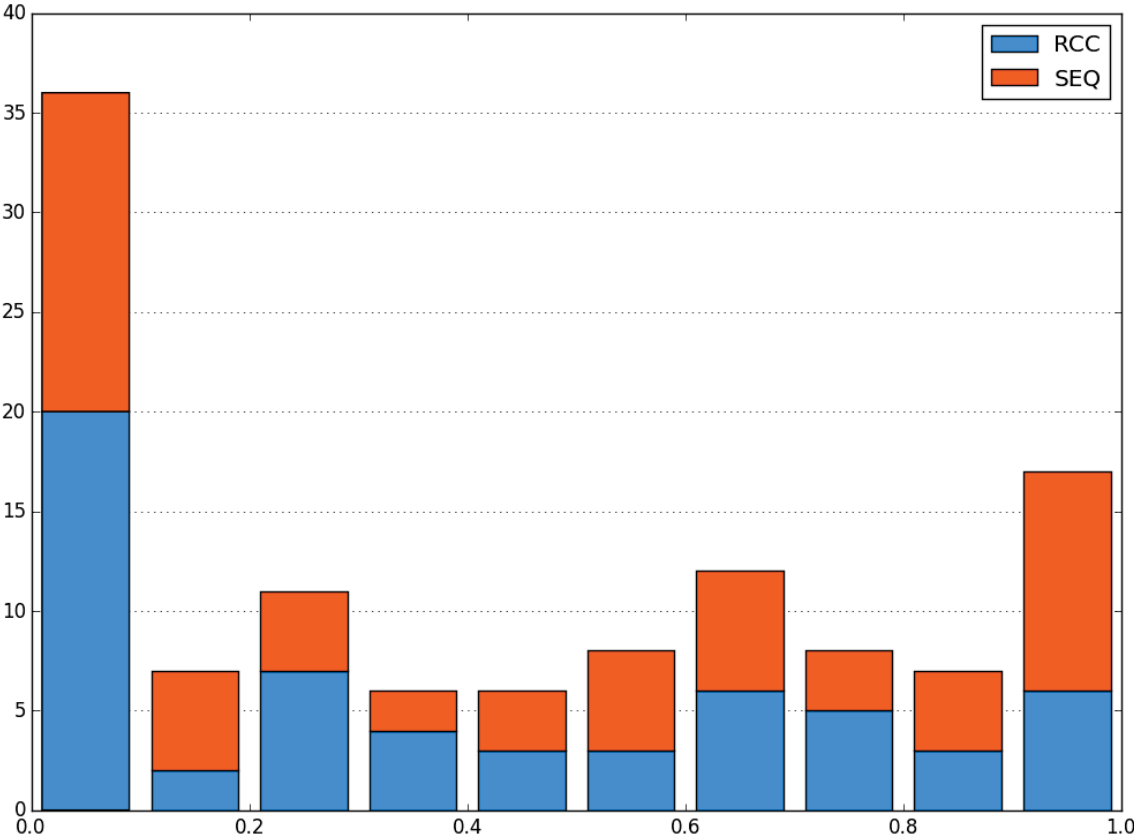


Figura 3.17: **Histograma de eficacias.** Se muestra la cantidad de eficacias en cada rango para los enfoques SEQ y RCC

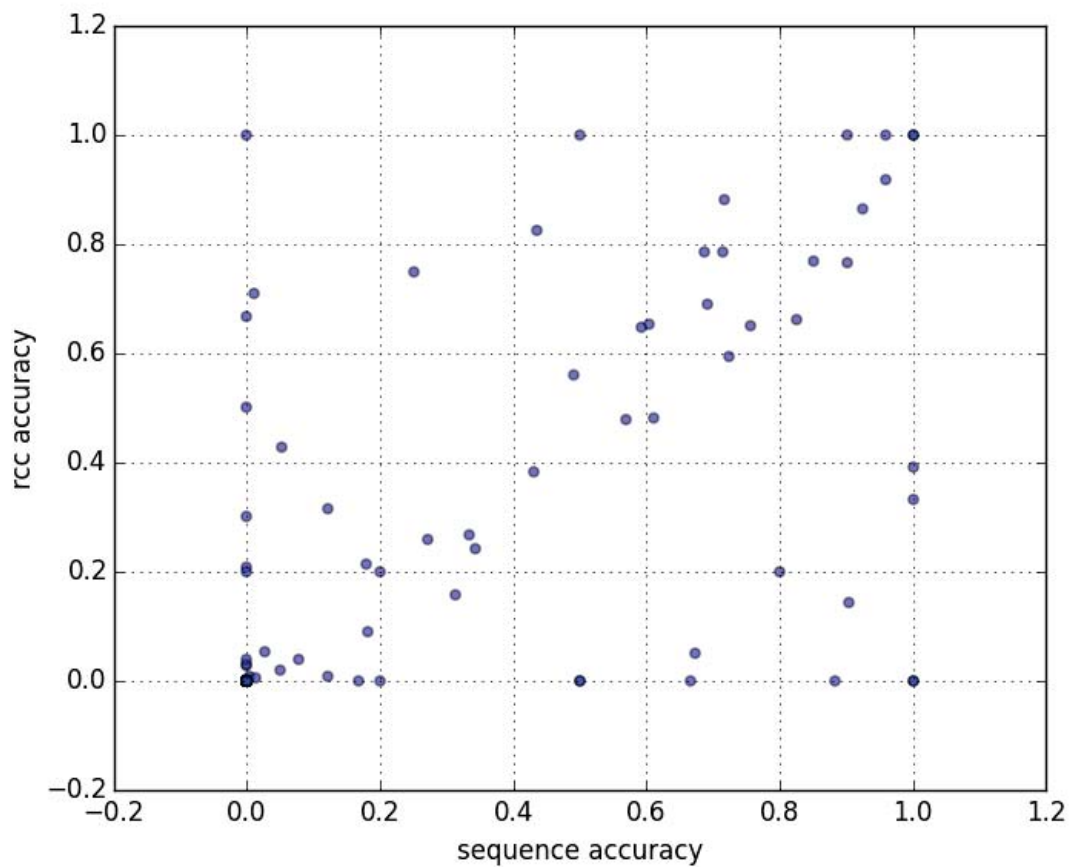


Figura 3.18: Comparación de eficacias de predicción de topología a partir de secuencia (seq) y a partir de predicción del vector en 26 dimensiones (rcc) Cada punto representa una familia definida a nivel de H en CATH. La eficacia es medida como la fracción de sus miembros predichos bajo la topología correcta

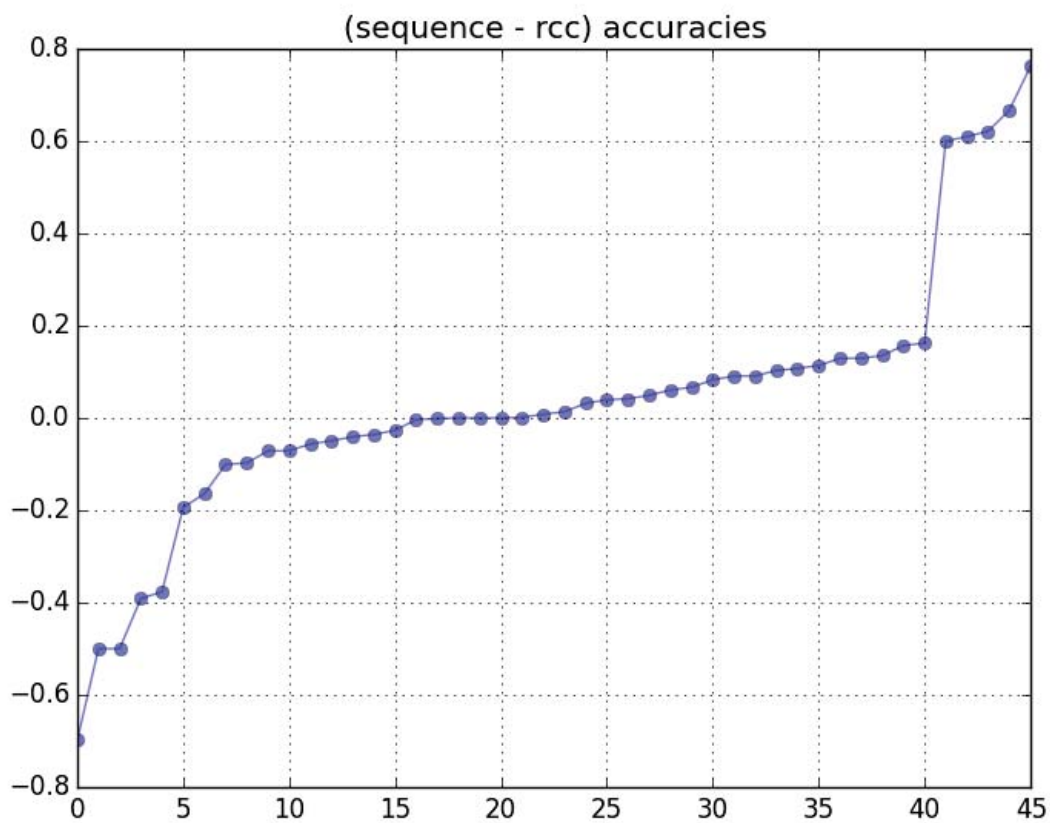


Figura 3.19: **Diferencias entre eficacias de predicción de SEQ y RCC**
 Cada punto corresponde a la evaluación de una familia completa. Las diferencias están ordenadas ascendentemente.

CAPÍTULO 4

Discusión

Representación de estructuras de proteínas mediante cúmulos de residuos

En el presente trabajo se introduce una representación de la estructura de proteínas usando una Familia de Sperner definida por cúmulos de residuos que están en proximidad en el espacio tridimensional. En la Figura 6 Apéndice B se muestra una relación lineal entre el tamaño de la proteína y el número total de tales cúmulos que puede ser asociada con la densidad casi constante de proteínas. La pequeña desviación en el número de cúmulos para proteínas pequeñas también es consistente con la observación de que proteínas pequeñas tienden a tener densidades promedio mayores [37].

Se encontró que las proteínas usualmente no contienen cúmulos mayores a seis residuos, independientemente de la composición de estructura secundaria. Por lo tanto cúmulos de residuos de tamaño tres hasta seis fueron identificados de acuerdo a la contigüidad de residuos en secuencia.

Representación del espacio de plegados

Este Sistema de Clases de Cúmulos de Residuos es usado para construir un modelo de espacio vectorial del espacio de plegados de proteínas en 26 dimensiones (ver Figura 8 Apéndice B). Esta representación de la estructura de proteínas puede ser usada para probar la separabilidad de los plegados de proteínas independientemente de medidas de similitud (ej. desviación cuadrada media) o el uso de cualquier clasificador particular.

Separabilidad del espacio de plegados

Se ha probado que el Cut Edge Weight Statistic es una prueba apropiada para la separabilidad y consecuentemente *aprendibilidad* de cualquier conjunto de datos [41]. De cualquier manera, el uso de esta prueba requiere construir una Gráfica de Vecinos Relativos (RNG por sus siglas en inglés) que consume $O(n^3)$ operaciones, (donde n es el número de estructuras de proteínas analizadas; ver Métodos 2.4 Apéndice B). Aquí se muestra que una gráfica HSP puede ser construida más rápido que una RNG y se demuestra que una HSP siempre contiene a una RNG, por lo tanto, HSP puede ser usada como alternativa para generar la RNG en $O(n^2)$ amortizado. Se probó la separabilidad de este espacio con dominios y clases reportadas en CATH a nivel de topología y se encontró que es posible aprender automáticamente esta estructura del espacio de plegados.

Visualización del espacio de plegados

Para una vista intuitiva del espacio de plegados, la Figura 8 Apéndice B presenta un encajamiento en dos dimensiones coloreados de acuerdo a su Clase. Aunque todos los atributos en 26 dimensiones no pueden ser preservados en tal encajamiento, se puede apreciar que los dominios $\alpha\beta$ se encuentran entre los dominios principalmente α y principalmente β . Seguido de esto se ejecutaron diferentes pruebas para determinar si esta separación reproduce la separación propuesta por CATH.

Clasificación estructural automática

Las puntuaciones de un clasificador ERF para identificar correctamente la clase estructural a la que pertenecen diferentes proteínas (Tabla 1 Apéndice B), muestran un buen desempeño en una prueba de validación cruzada. La puntuación obtenida de 0.969 para predecir la Clase según CATH puede ser comparada directamente con la puntuación 0.857 reportada por [48]. Una puntuación casi perfecta se obtiene cuando se evalúa usando el mismo conjunto de entrenamiento y prueba. Aunque no existe confusión entre este ERF y CATH en clasificar proteínas de las clases principalmente α y principalmente β , hay varias diferencias entre $\alpha\beta$ y principalmente α o principalmente β . Algunos ejemplos de estas discrepancias se presentan en la Figura 7 Apéndice B, donde las clases predichas por el ERF parecen visualmente ser más convenientes que la asignación CATH. En el contexto del espacio de plegados, estas estructuras pueden estar en una región poblada con dominios pertenecientes a la clase predicha, pero que representan un problema con la asignación original dada por CATH, como ha sido mencionado previamente por los mismos desarrolladores de CATH [22].

Método para la obtención de vecinos estructurales

Para la tarea de obtención de vecinos estructurales, se mostró que esta representación en modelo de espacio vectorial del espacio de plegado de proteínas supera a FragBag [43] usando la distancia euclideana. Sin embargo, este enfoque no supera los resultados obtenidos por ContactLib [42]. Esta tendencia en este enfoque puede ser explicada al considerar que estructuras de proteínas bajo la misma clasificación estructural pueden tener fronteras de decisión que la distancia no puede detectar con facilidad, como se notó en la Figura 1.3. Cuando un enfoque supervisado es utilizado para aprender las fronteras de decisión, nuestros resultados superan a ContactLib en términos de la AUROC media y de la distribución de puntuaciones AUROC en general, pero es prácticamente idéntico en el caso de la AUROC media para SAS20. El resultado de que al usar dos enfoques diferentes se obtenga un punto meseta alrededor de 0.96 de AUROC media, sugiere una frontera superior consecuencia de la asignación de Superfamilia en SCOP. En cualquier caso, las puntuaciones de validación cruzada presentadas aquí pueden servir como base para predictores futuros sobre esta

representación del espacio de plegados para aprender asignaciones de CATH y SCOP.

Clasificación estructural automática desde secuencia

Se propone una estrategia de obtención de atributos de secuencia que permitan eficacias de predicción de plegado de mayores a 0.97. Sin embargo, esta capacidad disminuye si se eliminan secuencias semejantes a la secuencia de interés. Para reducir la dependencia de ejemplos directamente semejantes, se propuso la estrategia de estimar la posición de la secuencia en el espacio de plegados. Las eficacias medias en la predicción de plegado desde secuencia o desde la estimación en el espacio de plegados son bastante parecidas, al rededor de 0.5. Los aciertos de ambos enfoques pueden ser sobre las mismas secuencias, sin embargo, en varios casos uno de los enfoques tiene aciertos, mientras el otro no. Esto muestra que la estrategia de transformar el espacio de secuencias en el espacio de plegados permite inferir correctamente la topología de algunas secuencias cuando la información de similitud detectable en las secuencias es insuficiente. Queda por determinar si es posible anticipar cuándo uno de los dos enfoques tendrá mejor desempeño que el otro e idear alguna estrategia que combine ambas predicciones.

Perspectivas tecnológicas y mejora de resultados

Los resultados en este trabajo dependen en gran medida de los algoritmos de aprendizaje de máquina utilizados, pero las estrategias generales propuestas permiten la utilización directa de otros algoritmos. Una familia de algoritmos con gran desarrollo al momento de escribir este trabajo son las redes neuronales artificiales. Estas redes ya están siendo utilizadas para explorar la posibilidad de superar algunos de los resultados reportados aquí.

Además de mejorar los resultados presentados, espero que la representación mostrada aquí sea de utilidad en otras tareas aún no exploradas.

CAPÍTULO 5

Conclusión

En resumen, se mostró una representación del espacio de plegados de proteínas que reproduce aspectos básicos del empacamiento de proteínas. Se describió un algoritmo eficiente para probar la separabilidad de este espacio vectorial y se mostró que nuestra representación del espacio de plegados de proteínas es separable y consecuentemente aprendible por cualquier clasificador. Esta propiedad fue usada para la predicción directa de clasificaciones estructurales y para la tarea de obtener vecinos estructurales de proteínas. Se propuso además un modelo de espacio vectorial para el espacio de secuencias, se mostró que esta representación obtiene buenas eficacias para la tarea de predicción de clasificaciones estructurales, pero no supera las limitaciones de abundancia de secuencias semejantes conocidas. Se exploró la fusión de los espacios de secuencia y de plegados para aliviar este problema con resultados prometedores.

Bibliografía

- [1] J. Moult, J. T. Pedersen, R. Judson, and K. Fidelis, “A large-scale experiment to assess protein structure prediction methods,” Proteins, vol. 23, 1995.
- [2] P. Manohar and S. Singh, “Review of common sequence alignment methods: Clues to enhance reliability,” Current Genomics, vol. 4, pp. 131–146, 2003.
- [3] S. R. Eddy, “Multiple alignment using hidden markov models,” Proc. Third Int. Conf. Intelligent Systems for Molecular Biology, pp. 114–120, 1995.
- [4] J. M. Andriy Kryshchak, Krzysztof Fidelis, “Casp9 results compared to those of previous casp experiments,” Proteins, vol. 79, 2011.
- [5] Y. J. et al., “An expanded evaluation of protein function prediction methods shows an improvement in accuracy,” Genome Biology, vol. 17, no. 184, 2016.
- [6] J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, “The Protein Data Bank,” Nucleic Acids Research, vol. 28, pp. 235–242, 2000.
- [7] F. Pearl, A. Todd, I. Sillitoe, M. Dibley, O. Redfern, T. Lewis, C. Bennett, R. Marsden, A. Grant, and D. Lee, “The CATH domains structure database and related resources Gene3D and DHS provide comprehensive domain

- family information for genome analysis,” Nucleic Acids Research, vol. 33, pp. D247–D251, 2005.
- [8] M. A. G., B. S. E., and C. C. Hubbard T., “Scop: a structural classification of proteins database for the investigation of sequences and structures,” J. Mol. Biol., vol. 30, no. 1, 1995.
- [9] C. Orengo, A. Michie, S. Jones, D. Jones, M. Swindells, and J. Thornton, “Cath - a hierarchic classification of protein domain structures,” Structure, vol. 5, no. 8, pp. 1093 – 1109, 1997.
- [10] G. Csaba, F. Birzele, and R. Zimmer, “Systematic comparison of scop and cath: a new gold standard for protein structure analysis,” BMC Structural Biology, vol. 9, no. 1, pp. 1–11, 2009.
- [11] C. S and B. CL, “Viral capsid proteins are segregated in structural fold space,” PLoS computational biology, vol. 9, 2013.
- [12] B. H. Rachel Kolodny, Donald Petrey, “Protein structure comparison: implications for the nature of ‘fold space’, and structure and function prediction,” Current opinion in structural biology, vol. 16, pp. 393–398, 2006.
- [13] S. Y. L. Jeffrey Skolnick, Adrian K. Arakaki and M. Brylinski, “The continuity of protein structure space is an intrinsic property of proteins,” Proceedings of the National Academy of Sciences of the United States of America, vol. 106, no. 37, pp. 15690–15695, 2009.
- [14] N. V. G. Ruslan I Sadreyev, Bong-Hyun Kim, “Discrete-continuous duality of protein structure space,” Current opinion in structural biology, vol. 19, no. 3, pp. 321–328, 2009.
- [15] M. L. Peter Minary, “Probing protein fold space with a simplified model,” Journal of molecular biology, vol. 375, no. 4, pp. 920–933, 2008.
- [16] H. H and H. L, “Advances and pitfalls of protein structural alignment.,” Curr. Opin. Struct. Biol., pp. 341–348, 2009.
- [17] L. Holm and C. Sander, “Touring protein fold space with dali/fssp,” Nucleic acids research, vol. 26, no. 1, pp. 316–319, 1998.

- [18] C. Z. Jingtong Hou, Gregory E. Sims and S.-H. Kim, “A global representation of the protein fold space,” Proceedings of the National Academy of Sciences of the United States of America, vol. 100, no. 5, pp. 2386–2390, 2003.
- [19] C. Z. Jingtong Hou, Se-Ran Jun and S.-H. Kim, “Global mapping of the protein structure space and application in structure-based inference of protein function.,” Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 10, pp. 3651–3656, 2005.
- [20] R. S. Daniel Asarnow, “The impact of structural diversity and parameterization on maps of the protein universe,” BMC Proceedings, vol. 7, no. S7, pp. 753–6561, 2013.
- [21] W. R. T. Christine A. Orengo, “Ssap: sequential structure alignment program for protein structure comparison,” Methods in enzymology, vol. 266, pp. 617–635, 1996.
- [22] J. M. T. Alex D. Michie, Christine A. Orengo, “Analysis of domain structural class using an automated class assignment protocol,” Journal of molecular biology, vol. 262, no. 2, pp. 168–185, 1996.
- [23] Y. N. Inbal Budowski-Tal and R. Kolodny, “Fragbag, an accurate representation of protein structure, retrieves structural neighbors from the entire pdb quickly and accurately,” Proceedings of the National Academy of Sciences of the United States of America, vol. 107, no. 8, pp. 3481–3486, 2010.
- [24] P. RÅgen and B. Fain, “Automatic classification of protein structure by using gauss integrals,” Proceedings of the National Academy of Sciences of the United States of America, vol. 100, no. 1, pp. 119–124, 2003.
- [25] M. J. Sippl, “On distance and similarity in fold space,” Bioinformatics, vol. 24, no. 6, pp. 872–873, 2008.
- [26] J. R. Quinlan, “Induction of decision trees,” Mach. Learn., vol. 1, no. 1, pp. 81 – 106, 1986.

- [27] J. R. Quinlan, “C4.5: Programs for machine learning.,” Morgan Kaufmann Publishers, 1993.
- [28] C. Strobla, A.-L. Boulesteix, and T. Augustin, “Unbiased split selection for classification trees based on the gini index,” Computational Statistics & Data Analysis, vol. 52, pp. 483–501, 2007.
- [29] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5 – 32, 2001.
- [30] P. Geurts, D. Ernst., and L. Wehenkel, “Extremely randomized trees,” Machine Learning, vol. 63, no. 1, pp. 3 – 42, 2006.
- [31] R. Bellman, Dynamic Programming. Princeton University Press, 2003.
- [32] J. R. Bock and D. A. Gough, “Predicting protein-protein interactions from primary structure,” Bioinformatics, vol. 17, no. 5, pp. 455–460, 2001.
- [33] C. Leslie and R. Kuang, “Fast string kernels using inexact matching for protein sequences,” Journal of Machine Learning Research, vol. 5, pp. 1435–1455, 2004.
- [34] C. Etchebest, C. Benros, A. Bornot, A.-C. Camproux, and A. G. de Brevern, “A reduced amino acid alphabet for understanding and designing protein adaptation to mutation.,” Eur Biophys J, vol. 36, pp. 1059–1069, 2007.
- [35] S.-F. Alejandro, P.-R. Ernesto, and S. Lorenzo, “Protein homology detection and fold inference through multiple alignment entropy profiles.,” Proteins, vol. 70, pp. 248–256, 2008.
- [36] D. Lubell, “A short proof of Sperner’s lemma,” Journal of Combinatorial Theory, vol. 1, no. 2, p. 299, 1966.
- [37] H. Fischer, I. Polikarpov, and A. Craievich, “Average protein density is a molecular-weight-dependent function,” Protein Science, vol. 13, no. 10, pp. 2825–2828, 2004.
- [38] M. Belkin and P. Niyogi, “Convergence of laplacian eigenmaps,” preprint, short version NIPS 2008, 2008.

- [39] Y. Zhang and J. Skolnick, “Scoring function for automated assessment of protein structure template quality,” Proteins, vol. 57, pp. 702–710, 2004.
- [40] Y. Zhang and J. Skolnick, “Tm-align: A protein structure alignment algorithm based on tm-score,” Nucleic Acids Research, vol. 33, pp. 2302–2309, 2005.
- [41] D. Zighed, S. Lallich, and F. Muhlenbach, “A statistical approach for separability of classes,” Applied Stochastic Models in Business and Industry, vol. 21, no. 2, pp. 187–197, 2005.
- [42] X. C. *et al.*, “Fingerprinting protein structures effectively and efficiently,” Bioinformatics, vol. 7, 2014.
- [43] B.-T. *et al.*, “Fragbag, an accurate representation of protein structure, retrieves structural neighbors from the entire pdb quickly and accurately,” Proceedings of the National Academy of Sciences of the United States of America, vol. 8, 2010.
- [44] L. Kinch, S. Y. Shi, Q. Cong, H. Cheng, Y. Liao, and N. V. Grishin, “Casp9 assessment of free modeling target predictions,” Proteins, vol. 79, 2011.
- [45] Z. A, V. C, M. J, and F. K, “Processing and analysis of casp3 protein structure predictions,” Proteins, vol. Suppl 3:22, no. 9, 1999.
- [46] R. Jauch, H. C. Yeo, P. R. Kolatkar, and N. D. Clarke, “Assessment of casp7 structure predictions for template free targets,” Proteins, vol. 69, 2007.
- [47] M. Ben-David, O. Noivirt-Brik, A. Paz, J. Prilusky, and J. L. Sussman, “Assessment of casp8 structure predictions for template free targets,” Proteins, vol. 77, 2009.
- [48] S. Shivashankar, S. Srivathsan, B. Ravindran, and A. V. Tendulkar, “Multi-view methods for protein structure comparison using latent dirichlet allocation,” Bioinformatics, vol. 27, no. 13, pp. i61–i68, 2011.

Apéndice

[Apéndice]

APÉNDICE **A**

Predicción de topología para objetivos en CASP12

Cuadro A.1: Predicción de topología CASP12. Parte I

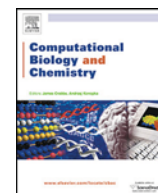
target	seq	rcc	target	seq	rcc	target	seq	rcc
T0859	2.60.40	2.60.40	T0860	2.60.40	2.60.40	T0861	3.40.50	3.20.20
T0862	3.40.50	3.40.50	T0863	1.10.630	3.20.20	T0864	3.40.50	3.40.50
T0865	1.10.10	1.20.5	T0866	2.170.170	2.60.40	T0867	2.60.40	2.30.42
T0868	3.40.50	3.40.50	T0869	1.10.533	3.40.50	T0868	3.40.50	3.40.50
T0869	1.10.533	3.40.50	T0870	1.20.5	3.40.50	T0871	3.20.20	3.20.20
T0872	2.30.30	1.10.10	T0873	3.20.20	3.40.50	T0874	2.60.40	3.40.50
T0875	2.60.40	2.60.40	T0876	3.40.50	2.60.40	T0877	3.40.50	2.60.120
T0878	3.40.50	1.10.630	T0879	3.40.50	2.60.120	T0880	3.40.50	2.60.40
T0881	2.60.40	2.60.120	T0882	3.30.70	3.10.20	T0883	2.60.120	2.60.120
T0884	2.60.40	2.10.70	T0885	3.30.390	1.10.10	T0886	3.40.50	2.160.20
T0887	3.40.50	1.10.490	T0888	2.60.40	2.60.40	T0889	3.40.50	3.40.50
T0890	3.30.565	3.40.50	T0891	3.40.50	3.40.50	T0892	3.40.50	3.40.50
T0893	3.40.50	3.20.20	T0894	3.40.50	3.40.50	T0895	2.60.120	2.60.40
T0896	3.20.20	3.20.20	T0897	3.20.20	3.40.50	T0898	3.40.50	1.10.357
T0899	3.20.20	3.20.20	T0900	2.60.40	2.60.40	T0901	3.20.20	3.40.50
T0902	2.60.120	3.20.20	T0903	3.20.20	3.20.20	T0904	3.40.50	3.20.20
T0905	3.40.50	3.20.20	T0906	3.40.50	1.10.630	TR862	3.40.50	3.40.50
T0907	2.120.10	3.40.50	T0908	2.60.40	2.60.40	TR872	2.30.30	1.10.10
T0909	3.40.50	2.60.120	T0910	3.20.20	3.20.20	TR879	3.40.50	2.60.120
T0911	3.40.50	1.10.620	TR868	3.40.50	3.40.50	T0912	3.20.20	3.20.20
TR869	1.10.533	3.40.50	T0913	1.10.630	3.20.20	TR870	1.20.5	3.40.50
Ts886	3.40.50	2.160.20	T0914	3.40.50	3.40.50	T0915	2.60.40	2.60.120
TR866	2.170.170	2.60.40	T0916	3.40.50	3.40.50	T0917	2.140.10	2.120.10
TR882	3.30.70	3.10.20	T0918	3.20.20	3.20.20	TR877	3.40.50	2.60.120
T0919	3.20.20	2.160.20	TR876	3.40.50	2.60.40	T0920	3.40.50	3.20.20
TR874	2.60.40	3.40.50	Ts894	3.40.50	3.40.50	Ts895	2.60.120	2.60.40
T0921	2.60.40	2.40.70	T0922	2.40.10	2.30.30	TR875	2.60.40	2.60.40
T0923	3.20.20	3.20.20	TR881	2.60.40	2.60.120	T0924	1.10.10	1.10.287
TR884	2.60.40	2.10.70	TR885	3.30.390	1.10.10	T0925	2.30.42	1.10.10
T0926	1.20.5	1.10.30	T0927	3.30.70	1.10.287	T0928	3.20.20	3.20.20
T0929	2.60.40	2.60.40	TR887	3.40.50	1.10.490	Ts866	2.170.170	2.60.40

Cuadro A.2: Predicción de topología CASP12. Parte II

target	seq	rcc	target	seq	rcc	target	seq	rcc
T0929	2.60.40	2.60.40	TR887	3.40.50	1.10.490	Ts866	2.170.170	2.60.40
T0930	1.10.630	3.20.20	T0931	3.40.50	1.10.490	T0932	3.40.50	3.40.50
T0933	3.40.50	2.160.20	T0934	3.20.20	3.20.20	TR890	3.30.565	3.40.50
TR891	3.40.50	3.40.50	T0935	2.60.120	2.60.40	T0936	1.10.357	3.40.50
TR893	3.40.50	3.20.20	T0937	3.20.20	3.20.20	T0938	3.20.20	3.40.50
TR896	3.20.20	3.20.20	T0939	3.40.50	1.10.520	T0940	3.30.930	3.20.20
TR901	3.20.20	3.40.50	T0941	1.10.630	3.20.20	TR905	3.40.50	3.20.20
T0942	3.20.20	3.20.20	TR909	3.40.50	2.60.120	T0943	3.20.20	3.20.20
T0944	3.20.20	3.40.50	TR898	3.40.50	1.10.357	T0945	3.20.20	3.40.50
T0946	3.20.20	2.120.10	TR912	3.20.20	3.20.20	TR913	1.10.630	3.20.20
T0947	3.40.50	3.40.50	T0948	3.40.50	3.40.50	TR928	3.20.20	3.20.20
TR921	2.60.40	2.40.70	TR922	2.40.10	2.30.30	TR594	3.40.50	3.40.50
TR894	3.40.50	3.40.50	TR895	2.60.120	2.60.40	TR694	3.20.20	2.160.20
Ts909	3.40.50	2.60.120	TR910	3.20.20	3.20.20	TR520	3.40.50	3.20.20
TR920	3.40.50	3.20.20	Ts899	3.20.20	3.20.20	TR917	2.140.10	2.120.10
TR945	3.20.20	3.40.50	TR948	3.40.50	3.40.50	Ts942	3.20.20	3.20.20
TR942	3.20.20	3.20.20	Ts901	3.20.20	3.40.50	TR947	3.40.50	3.40.50
Ts947	3.40.50	3.40.50	TR944	3.20.20	3.40.50	Ts896	3.20.20	3.20.20
Tx894	3.40.50	3.40.50	Tx895	2.60.120	2.60.40	Ts941	1.10.630	3.20.20
Tx892	3.40.50	3.40.50						

APÉNDICE B

Machine Learnable Fold Space Representation based on Residue Cluster Classes



Research Article

Machine Learnable Fold Space Representation based on Residue Cluster Classes

Ricardo Corral-Corral^a, Edgar Chavez^b, Gabriel Del Rio^{a,*}^a Department of Biochemistry and Structural Biology, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México D. F., México^b Centro de Investigación Científica y de Educación Superior de Ensenada, México

ARTICLE INFO

Article history:

Received 12 November 2014

Received in revised form 17 July 2015

Accepted 25 July 2015

Available online 30 July 2015

ABSTRACT

Motivation: Protein fold space is a conceptual framework where all possible protein folds exist and ideas about protein structure, function and evolution may be analyzed. Classification of protein folds in this space is commonly achieved by using similarity indexes and/or machine learning approaches, each with different limitations.

Results: We propose a method for constructing a compact vector space model of protein fold space by representing each protein structure by its residues local contacts. We developed an efficient method to statistically test for the separability of points in a space and showed that our protein fold space representation is learnable by any machine-learning algorithm.

Availability: An API is freely available at <https://code.google.com/p/pyrcc/>.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

All possible protein folds are assumed to occupy an abstract space referred to as fold space. This fold space has become a conceptual framework to unify ideas about protein structures with protein function and protein evolution (Cheng and Brooks, 2013). For instance, it is debated whether this space is discrete or continuous (Kolodny et al., 2006; Skolnick et al., 2009; Sadreyev et al., 2009). Relevant to our study is the common use of protein similarity measures (e.g., root mean square deviation or RMSD) aimed to infer their proximities in this space (Minary and Levitt, 2008). In this case, the inference derived from such measurements assumes that the proximity of protein folds is the only relevant property to explain protein fold evolution and function.

Instead of focusing only on the proximity of protein folds, vector space models have been used to expand the protein fold space representation. In this space, each protein structure is represented in a fixed dimension space (e.g., euclidean space) by a point (position vector); adjusting the positions of these vectors by approximating their relative distances to protein similarity measures may derive these position vectors. For example, using sequential structure alignment program scores between each pair of structures as a protein similarity measure (Orengo and Taylor, 1996), followed

by multi dimensional scaling allowed the assignment of positions vectors representing each protein structure (Michie et al., 1996).

Using DALI as similarity measure, Holm (Holm and Sander, 1998) showed two-dimensional projections to explore protein neighbors in fold space. DALI has also been used as similarity measure to visualize class distribution and fold usages between two bacterial species (Hou et al., 2003) and to explore protein function assignment based on position on this fold space representation (Hou et al., 2005).

Fold space constructions based on protein structure similarities have two important limitations. First, the time needed to run a structural comparison for each pair of structures is restrictive: 25000 central processing unit hours were needed for calculating similarities between 1898 protein structures (Hou et al., 2005). Second, the position of each point depends heavily on the set of structures being analyzed, and in such case the inclusion of a single new structure can displace all previously assigned positions; thus, there are as many fold space representations as different protein structures data sets, even adopting a unique similarity score.

An alternative fold space representation may be built by assigning the position of the vector considering only the structural features of the protein it represents. In this way, a new structure can find its location in this fold space without altering the existing ones. One implementation of a vector space model is FragBag (Budowski-Tal et al., 2010), which represents protein structures in 400 dimensions; here, each component in the vector representing a protein fold corresponds to the number of occurrences of a

* Corresponding author. Tel.: +44 000 0000000; fax: +44 000 0000000.
E-mail address: gdelrio@ifc.unam.mx (G. Del Rio).

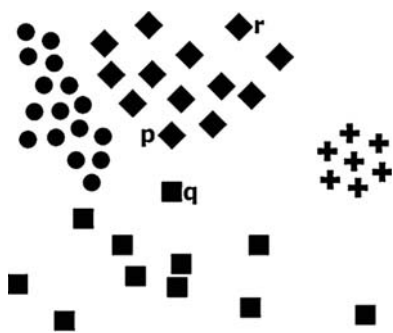


Fig. 1. Proximity in space does not always imply membership to a class. The figure illustrates cases where proximity in a space does not imply membership to a class. Cross points define a class where distances between every pair of its members should be less than the distance to any point belonging to other class, an assumption made when using similarity scores. An exception to this assumption is represented by the rhombus class members p and r , which distance is larger than for member class p to a square class member q . This is true for any distribution within the classes. To show this, circle class has a dense and regular distribution, rhombus class is regular and sparse, while square is very sparse and irregular. In any case, each class is clearly segregated. Please note that the different fold classes may be learned using machine-learning classifiers, but here we illustrate that using similarity measures as the only criterion to distinguish class membership may induce errors.

particular contiguous protein sequence fragment. Another approach uses knot invariants as values in each component for vector points in 30 dimensions (Rogen and Fain, 2003). In these cases, it is assumed that proximal protein structures should belong to the same structural class, assumption that is not necessarily correct as we will argue below.

Once a protein fold space construction is chosen, a metric distance induced by the space can be used as a measure of similarity and it is expected to be in agreement with direct structural measures (such as RMSD, GDT, TMscore, etc), but overcoming the problems noted above about these scores not being a metric (Sippl, 2008).

In such space, a given set of protein structures that are considered to have the same fold may be close in this space representation. Yet, it may occur that some proteins with different folds may be closer than proteins with the same fold (see Figure 1). Thus, confusion may be induced at distinguishing class membership in this fold space when only similarity measures are considered.

To address this problem, the boundaries between proteins with different folds may be obtained using empirical data and machine learning algorithms, which naturally segregate protein structures sharing common features. In these models it is also possible to evaluate the separability of this space independently of any machine-learning algorithm using a statistical test (Zighed et al., 2002). Therefore, it is possible to generate a protein fold space representation independent on the protein similarity measure used and to test for the separability of this space independently of any particular classification algorithm. This protein fold space may then be used to analyze protein structure-function relation and protein evolution without the limitations previously noted of current protein fold space construction approaches.

In this work, we propose a compact (low dimensional) fold space representation based on Residue Cluster Classes (RCCs), a Sperner family that includes all sets of residues in simultaneous contact. We also present an efficient computational method useful to test for the separability of this fold space representation. As a proof of principle, we analyzed the CATH classification and automatically detected conflicts in CATH. Furthermore, we show that our method improves state of the art protein structure neighbor retrieval methods. To facilitate the construction of protein folds represented by RCCs, we present an API available at <https://code.google.com/p/pyrcc/>.

2. Materials and Methods

2.1. Datasets

2.1.1. CATH datasets

CATHALL1 set includes all domains in CATH release v3.5 that were parsable with our API and consists of 168964 domains. CATHALL2 set includes all domains in CATH release v4.0 and contains 235858 domains. CATCHOP was obtained from a random sample of CATHALL1 considering only six domains per topology; topologies with less than six members were excluded rendering a total of 5220 domains (see supplemental Table S1 for a complete list).

2.1.2. SCOP datasets

The SCOP30 dataset was provided by the authors of ContactLib (Xuefeng Cui et al., 2014) and contains 3295 SCOP domains. SCOP30 contains 2639, 3232 and 3290 neighbours at SAS levels 20,35 and 50. Yet, only 2049, 2620 and 2722 domains have at least one neighbor in SAS 20, 35 and 50, respectively. The SCOPtrain1 is a random sample of 136300 SCOP 1.75B. SCOPtrain2 contains all 203025 domains from SCOP release v2.5. The SCOPtrainAUC includes 109310 SCOPtrain domains absent in SAS50 group, and only belonging to a class in SCOP30. If a class (at any level) contains more than 2000 domains, 2000 domains were chosen randomly to represent that class.

2.2. Construction of Residue Cluster Classes

2.2.1. Definitions

Residue Neighbourhood ($N_\epsilon(r)$). Let P be a protein with residues $R = r_1, r_2, \dots, r_n$. The system τ_{prim} is defined as:

$$\tau_{prim} = \{\{r_i, r_j\} : \text{there exists a bond between } r_i \text{ and } r_j\}$$

Given a metric $d: R \times R \rightarrow [0, \infty)$ and a cut-off distance ϵ , the neighbourhood $N_\epsilon(r)$ of a residue r is given by:

$$N_\epsilon(r) = \{x \in R : \exists a \in A(x), b \in A(r); d(a, b) \leq \epsilon\}$$

Where $A(r)$ is the set of non-hydrogen atoms of residue r . Thus, $N_\epsilon(r)$ is the set of all residues near r , i.e., they are at no more than a distance ϵ from r .

Residue Cluster (RC). A residue cluster on P is a subset $A \subseteq 2^R$ such that $A \subseteq N_\epsilon(a)$ for all $a \in A$. If $|A| = k$, then A is a k -Residue Cluster, ${}_kRC$

Residue Cluster Class (RCC). A class over a RC is defined by the primary structure on τ_{prim} . A pair of residues r_i, r_j are contiguous if $\{r_i, r_j\} \in \tau_{prim}$ and a set \bar{s}_L of L residues form a segment in τ_{prim} if it contains $(L-1)$ contiguous residues. By convention, $\bar{s}_1 = \{r\}$ for any $r \in R$.

Let be γ the family of all segments in τ_{prim} . C is a set cover of a k -Residue Cluster ${}_kRC$ if

$$C = \{\bar{s}_{L_\alpha} : \alpha \in A, \bar{s}_{L_\alpha} \in \gamma\}$$

such that

$${}_kRC = \bigcup_{\alpha \in A} \bar{s}_{L_\alpha}$$

and

$$c_i \cap c_j = \emptyset, \forall c_i, c_j \in C$$

Therefore, $C = \bar{s}_{L_1}, \bar{s}_{L_2}, \dots, \bar{s}_{L_A}$ is a covering if

$$\sum_{i=1}^A L_i = k$$

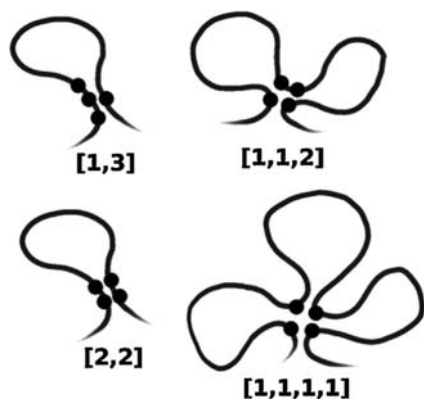


Fig. 2. Class names for residue clusters of same size. The curve represents a protein backbone, and black circles represent residues in a cluster being considered. Four distinct classes ([1,3], [1,1,2], [2,2] and [1,1,1,1]) arising from ${}_4\text{RC}$'s are shown. The class name is defined by the number of residues found in each segment that defines the RCC (e.g., ${}_4\text{RCC}$). For convention, numbers are written in non-decreasing order.

and can be uniquely identified by the vector class $\vec{C} = (L_1, L_2, \dots, L_A)$ with $L_i \leq L_j$ for any i, j such that $i \leq j$.

Lets define the k -Residue Cluster Class ${}_k\text{RC}_{\vec{C}}$ as the set of all ${}_k\text{RC}$ of class \vec{C} . The residue cluster class system \mathfrak{R} is a Sperner family (Lubell, 1996). It is defined as the set of all residue clusters for all k, \vec{C} such that none is subset of some other, that is:

$$\mathfrak{R} = \{W \in R : \exists X \in {}_k\text{RC}_{\vec{C}}, X \subseteq W\}$$

Let $M(\vec{c}_i)$ be the number of occurrences of a residue cluster class \vec{c}_i in a system \mathfrak{R} .

The vector point $\vec{\mathfrak{R}}$ of \mathfrak{R} is defined by

$$\vec{\mathfrak{R}} = \sum_{i \in I} M(\vec{c}_i) v_i$$

with $v_i \in B$, where $B = \{v_1, v_2, \dots, v_{|I|}\}$ is the standard orthonormal basis for $\mathbb{R}^{|I|}$ and I is an index set over the set of cluster classes considered.

Thus, an RCC construction is illustrated in Figure 2 where different classes are assigned to a ${}_k\text{RC}_{\vec{C}}$ according to sequence contiguity.

2.3. Construction of vector points for domain structures

Clusters were computed using graph theoretical algorithms. First, a graph G is constructed with a $V(G)$ vertex set that represents the residues set for a given protein, and $E(G)$ edges set is the set of all residue clusters of size two. This graph is equivalent to a residue contact map and is constructed with the same program used in (Cusack et al., 2007): any two residues are paired if they share at least one atom pair at 5Å or less in the three dimensional representation of a folded protein. As any clique in G (i.e. a complete subgraph) is consistent with our cluster definition, all maximal cliques (cliques that are not a subgraph of any other clique) are considered as clusters, and classes are assigned directly from sequence contiguity. Maximal cliques were obtained using the algorithm previously described (Tomita et al., 2006) as implemented in NetworkX (Hagberg et al., 2008).

Thus, for any given protein there are 3 ${}_3\text{RCC}$ ([1,1,1], [1,2], [3]); see Figure 2 for the nomenclature used to refer to these RCCs), 5 ${}_4\text{RCC}$ ([1,1,1,1], [1,1,2], [2,2], [1,3], [4]), 7 ${}_5\text{RCC}$ ([1,1,1,1,1], [1,1,1,2], [1,2,2], [1,1,3], [2,3], [1,4], [5]) and 11 ${}_6\text{RCC}$ ([1,1,1,1,1,1], [1,1,1,1,2], [1,1,2,2], [2,2,2], [1,1,1,3], [1,2,3], [3,3], [1,1,4], [2,4], [1,5], [6]); thus, there are 26 RCCs for any given protein. To represent these in a vector, we used the frequency of occurrence of each ${}_k\text{RCC}$ as the value for each dimension of this vector. Figure 3 illustrates an example of the construction of a vector using a self-avoiding walk over a



Fig. 3. Schematic construction of an RCC vector. The figure represents a protein three-dimensional structure in a lattice, where each residue is represented by a circle and the dotted lines represent residue contacts (tertiary contacts). A self-avoiding walk (SAW) showed as a straight line visiting all points of a grid represents the protein backbone. Considering four points in any square as a cluster (an RCC of size 4), there can be five classes of clusters showed at the bottom. Numbers in each cluster class are the number of occurrences of that cluster formed by the SAW. This same idea can be extended to real proteins to quantify clusters of size 3, 4, 5 and 6 to generate the final vector representation of a protein.

lattice grid. An API was made for cluster calculations from protein structures and can be accessed via <https://code.google.com/p/pyrcc/>.

2.4. Calculation of the Cut Edge Weight statistic

The Cut Edge Weight statistic (Zighed et al., 2002) was used to test the hypothesis that a given class distribution of vector points (i.e., protein domains in CATH) is random. The construction of the statistic is limited by the construction of the relative neighbourhood graph (RNG) of the points. A naive algorithm takes $O(n^3)$ and lower bounds have been reported only for particular cases in low dimensions. Here we include a demonstration that a RNG is always a subgraph of the Half-Space Proximal (HSP) graph (see below).

To obtain the RNG, the HSP graph H of the set of 26-dimension vectors was obtained (see below for HSP construction procedure). Then, for each edge (a, b) in H we remove the edge if the following condition was satisfied:

$$\max\{d(a, c), d(b, c)\} < d(a, b)$$

where $d(i, j)$ is the euclidean distance between points i and j .

This condition was applied for any c in H . After this procedure, the resulting graph was the RNG used for statistics computation as described in Zighed et al. (2002).

A detailed description of HSP and a proof for containing RNG is described below. The source code to compute this statistical test can be obtained from the authors upon request.

2.5. Relative Neighborhood Graph as a subgraph of the Half-Space Proximal graph

The Half Space Proximal (HSP) is a local test for building a geometric spanner of bounded dilation over a set of points in the space, it was introduced in Chavez et al. (2006). Assume there is a set V of points loaded with a metric. To fix ideas think in the plane and the euclidean distance between points, in our case the set of points are 26 dimensional vectors representing protein structures, although these could be any set of objects in an abstract metric space. All the construction will be based on the distance metric. For each point u in V we compute its HSP neighbours as follows.

For a point u in V take its nearest element v and add an edge from u to v ; remove all the elements that are closer to v than to u . The

region of points closer to v than to u is called the forbidden region from the point u with respect to v . From the remaining points, those points not in the forbidden region, take the nearest point to u and repeat until all points in V belong to some forbidden region. In the end, we will have a directed graph with vertex set V and the edges found with the above procedure. In this paper we are interested in the HSP as a super graph of the Relative Neighbourhood Graph (RNG), faster to build, as described below.

For the same set V two points u, v will share an edge in the RNG if there is not a point $z \in V$ such that z is in the intersection of the circles centered in u, v , respectively, with radius $|uv|$. Then, RNG and the HSP are related and we will show that $\text{RNG} \subseteq \text{HSP}$.

Lemma 1. *If there is not an edge from the point u to v , then there exist a point z such that it connects to u and z is in the intersection of the two circles centered at u and v with radius $|uv|$.*

Proof. If there is not an edge from the point u to v , then, v is in a forbidden region of u and some point z . The point z connects to u and is closer to u than from v , so, z is in the circle centered at u with radius $|uv|$. As v is in the forbidden region of u with respect to z , then v is closer to z than to u , so, z is in the circle centered at v with radius $|uv|$.

One way to characterize the RNG is by observing that two points p and q will share an edge whenever there is not a third point r that is closer to both p and q than they are to each other.

Lemma 2. *If an edge is in the RNG then, it is in the HSP.*

Proof. Lets suppose that two points p and q are not connected by the HSP, then, there exist a point z that is in the intersection of the two circles centered at p and q with radius $|pq|$, then the two point p, q are not connected in the RNG.

A brute force approach to build the RNG built over a set of n points require checking $O(n^2)$ pairs to see if they share an edge; furthermore, each pair requires $O(n)$ points to be tested for inclusion in the intersection. This yields a total complexity of $O(n^3)$ operations. On the other hand, all the HSP neighbours of a point u can be computed in an amortized $O(n)$ operations. The total complexity of building the HSP is then $O(n^2)$. The total number of HSP neighbours of any point u is bounded, and depends only on the dimension of the space, and it is independent of the size of the point set. This implies that the number of edges in the HSP is only linear on n .

Using the above observations to test if an edge is in the RNG graph, we only need to test the linear number of edges of the HSP. Each test cost $O(n)$ operations, and consist in checking if the intersection is empty. Even with brute force, the total time for building the RNG would be $O(n^2)$ instead of $O(n^3)$ using other construction algorithms.

Figure 4 illustrates the inclusion of the RNG in the HSP. Dotted edges in the figure are those to be removed from the HSP. To exemplify the improvement on computing time by our approach, Figure 5 shows execution times to calculate RNG, HSP and RNG from HSP for two randomly generated datasets.

2.6. Learner details

A learner including an ensemble of 250 extremely randomized trees (ERT) was trained for cross-validation tests, predicting the Class, Architecture and Topology in CATH classification, and Class, Topology and Superfamily in SCOP. Bootstrap samples were not used for building the trees, and a minimum of 3 nodes were required to remain after each split. In each tree, Gini index was used for deciding attributes for splitting.

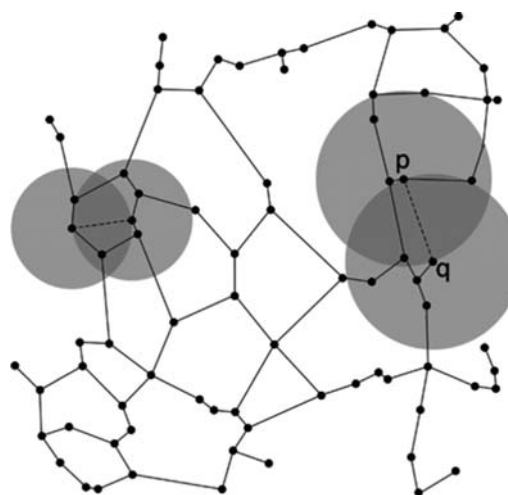


Fig. 4. Schematic representation about the differences of RNG and HSP. Light gray circles are traced with radius equals to the distance $|pq|$ between any pair of points in space (p and q in the figure); the dark gray regions (also known as lunes) test for the presence of any point in space within those light gray circles. If there exist a point z that is in the intersection of the two circles centered at p and q with radius $|pq|$, then the two points p, q are not connected in the RNG.

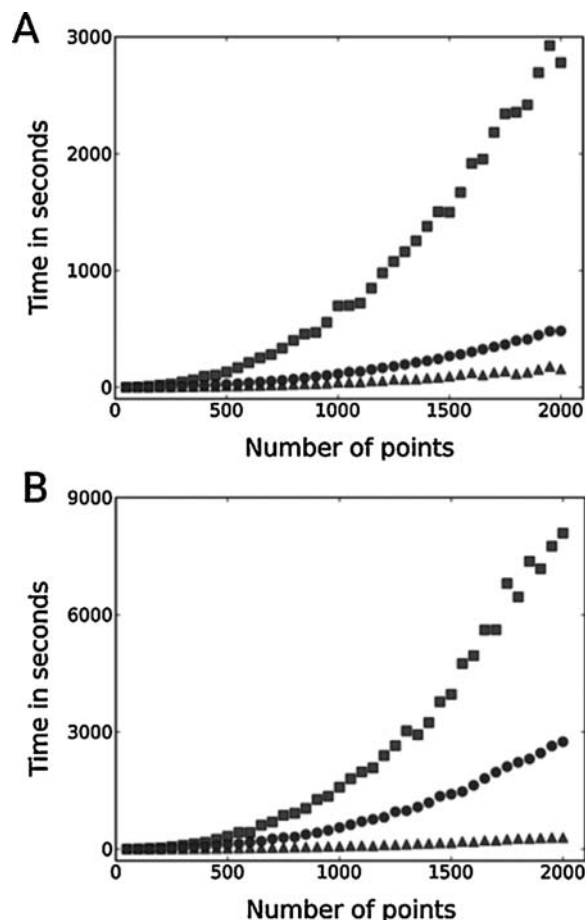


Fig. 5. Execution times to obtain RNG from random datasets. The points in the plot represent the computing time to generate a RNG (squares), HSP (triangles) and RNG from HSP (circles) over random point sets of different sizes in 2 dimensions (A) and 50 dimensions (B) (see Methods). RNG from HSP is clearly more efficient than brute force RNG, although this relative efficiency gains decreases as the number of dimensions increases, since all edges in HSP are checked for inclusion of RNG, and the number of edges in HSP tends to be larger on higher dimensions. These data sets were obtained using the random number generator model (without a seed number) implemented in the random standard package in Python 2.7.

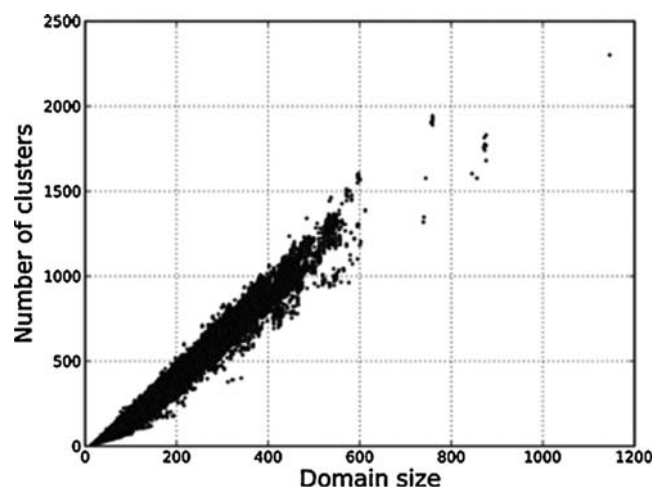


Fig. 6. Relationship between domain size and number of clusters in each domain. The number of residues found in every protein domain of the CATCHOP dataset is plotted against the total number of RCCs found in the corresponding domain. Note the scarcity of data in domains larger than 600 residues in CATH.

Same setup but with 300 trees was used for AUROC analysis. These classifiers were implemented using sklearn (Pedregosa et al., 2011).

To learn the structural classification from CATH, we used the CATHALL1 and CATHALL2 sets; for SCOP, we used the SCOPtrain1 and SCOPtrain2 sets.

3. Results

3.1. Defining vector dimensionality

The CATHALL1 set (see methods) was used to identify RCCs. Figure 6 shows a linear relationship between protein size and number of RCCs; RCCs larger than 10 residues were not observed. To define the largest RCC to be used for protein structure representation, we identified the frequency of occurrence of the largest RCCs in the CATHALL1 set (see supplemental Figure S1). The fraction of domains having RCCs of size six and seven account for the 91.61% and 7.92% respectively; furthermore, they are represented in 94.51% and 48.13% of topologies. Thus, RCCs larger than 6 residue were rare and found in only very few topologies. For the rest of our study we used RCCs of size 3 to 6 that include a total of 26 different RCCs, which were used to build our vector space model.

3.2. Separability of fold space

Next, we tested for the separability of our vector space; while learnability is addressed in Vapnik-Chervonenkis learning theory, this is learner dependent and do not evaluate the practical separability of a particular dataset (Holden and Niranjana, 1995). To quantify the separability (and thus, learnability) of our fold space representation, we used the Cut Edge Weight statistic (Zighed et al., 2002). In this procedure, the null hypothesis is that data is randomly distributed and cannot be classified. The implementation of this test is however computationally expensive, and thus we used the CATCHOP set (see Methods). Using this test in CATCHOP set, we observed that the null hypothesis is false ($p=0$) and consequently domains belonging to different topologies are separable.

3.3. Learning of fold space for structural classification

To use the separability of our fold space representation, we trained an extremely random forest (ERF) classifier to automatically

Table 1
CATH Cross-Validation scores

Classification level	Cross-Validation accuracy	standard-error	full train accuracy
C	0.9693	0.0014	0.9987
CA	0.8830	0.0019	0.9967
CAT	0.8494	0.0022	0.9984

Performance on 10-fold cross-validation over CATH domains. Mean of accuracies for each of the 10 iterations and corresponding standard errors are shown. Full train accuracy was obtained by training and testing on the same full CATHALL1 dataset.

Table 2
Confusion matrix of Classes over CATH (v3.5) domains.

True/Predicted	mainly- α	mainly- β	$\alpha\beta$	few secondary structure
mainly- α	36936	0	23	0
mainly- β	0	39944	38	1
$\alpha\beta$	42	74	89310	0
fss	19	7	17	2553

In each row (i) and column (j), the number of domains belonging to class i, predicted as j are shown. A perfect prediction has only numbers in the main diagonal. Note that no discrepancies exist between mainly- α and mainly- β .

assign classifications by class, architecture and topology according to the CATH definition. We performed 10-fold stratified cross-validation test to assess the ability of our ERF classifier to learn general rules from the CATHALL1 set (see Methods section); in this case, 10% (1/10) of the CATHALL1 domains were randomly chosen as part of the training samples (the rest was part of the test set) and this random procedure was performed 10 times; please note that in this way we do not generate specific training sets avoiding bias in the test. We also trained our ERF classifier over the entire data set. Performance predictions for these two approaches are reported in Table 1. We repeated this procedure for a new release of CATH included in CATHALL2 set and the results are reported in supplemental Table S2.

The discrepancies between the ERF classifier using the full data in CATHALL1 as the training set and the CATH classification are summarized in the Confusion matrix presented in Table 2. Some examples of these discrepancies are presented in Figure 7 (see supplemental Tables S3 for a complete list).

Classification performance was evaluated in the same way over SCOP domains (See Table 3 for results obtained with SCOPtrain1 set

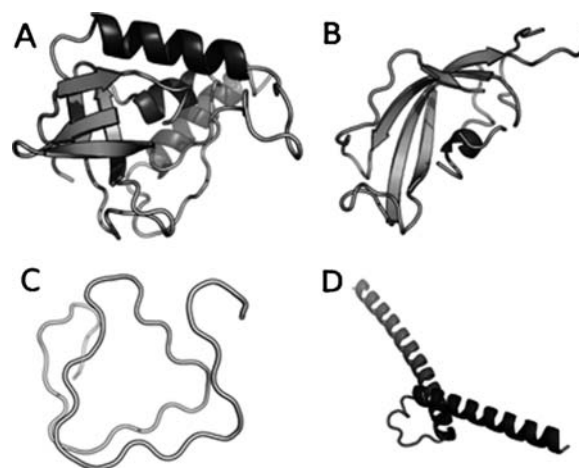


Fig. 7. Examples of domain classification discrepancies between ERF and CATH. A) 1a3uA00: predicted $\alpha\beta$, in CATH mainly- β ; B) 1w26A03: predicted mainly- β , in CATH $\alpha\beta$; C) 1h6wA02: predicted few secondary structure, in CATH mainly- β ; D) 1nkpD00: predicted mainly- α , in CATH few secondary structure.

Table 3
SCOP Cross Validation scores

Classification level	accuracy
Class	0.9243
Topology	0.8632
Superfamily	0.8781
Family	0.9097

Performance on 10-fold cross-validation over SCOPtrain1 domains. Final accuracies are the mean of accuracies for each of the 10 iterations.

and Supplemental Table S4 for results obtained with SCOPtrain2 set).

3.4. Fold space visualization

To visually inspect fold space structure, we used an unsupervised dimensionality reduction approach; this dimensionality reduction was performed by Laplacian-Eigenmaps (Belkin and Niyogi, 2003) over the CATCHOP set (see Methods). This method has been shown to give a better approximation of the vicinity of each point by approximating the Laplace Beltrami operator of the original space on the reduced space (Belkin and Niyogi, 2008). Projection into two dimensions is shown in Figure 8. In this representation, mainly- α and mainly- β domains are observed on opposite directions, while $\alpha\beta$ domains are found in between these groups.

3.5. Structural neighbour retrieval

ContactLib has been reported to achieve the best performance in finding protein structural neighbors when compared against the best alignment-free method FragBag (Xuefeng Cui et al., 2014); FragBag on the other hand was reported to improve on 6 state of the art structural aligners, including SGM, PRIDE, BLAST, STRUCTAL, CE and SSM. FragBag was tested in a subset of SCOP 1.75B domains, referred to as SCOP30; three SAS thresholds (2.0, 3.5 and 5 Å) identify three types of neighborhoods; SAS20 is the most sensitive and SAS50 the most relaxed. Agreement between SAS criterion and SCOP superfamily assignment is low (see supplementary Table S5), thus only the domains within each SAS neighborhood belonging to the same superfamily as defined in SCOP are considered true neighbours.

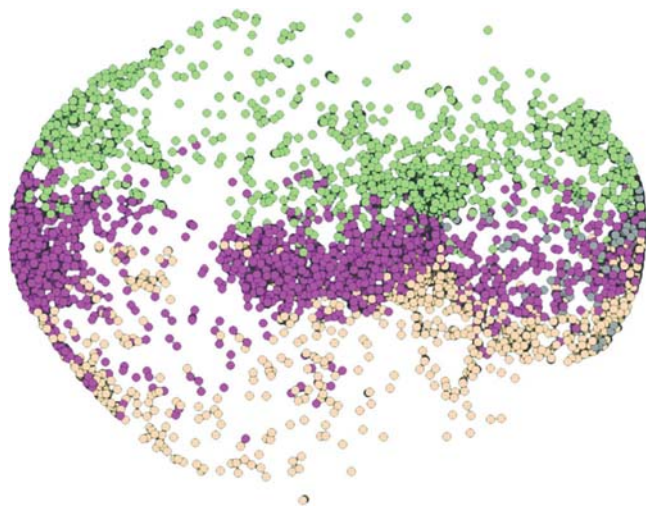


Fig. 8. Fold space visualization by dimensionality reduction. Domains in CATCHOP represented by Spectral-Embedding dimensionality reduction of the original 26-dimensional space. Different colours of the dots correspond to CATH class assignments: green dots for mainly- α , orange dots for mainly- β , purple dots for $\alpha\beta$ and gray dots for few secondary structure class.

Table 4
Mean AUROC

SAS level	Mean AUROC		Fraction above 75%
	RCC	ContactLib	
20	0.950	0.956	0.958
35	0.940	0.918	0.939
50	0.935	0.906	0.929

Summary of mean AUROC values for the task of neighbour retrieval. Direct comparison between mean AUROC obtained by our method (RCC) and ContactLib. Last column also shows the percentage of query structures for which AUROC was higher than 75%.

We calculated the Area Under the Receiver Operator Characteristic curve (AUROC) to test for the capacity of our fold space representation to identify true neighbors according to SAS and SCOP criteria; this approach was chosen to summarize the results obtained for all the SCOP domains analyzed. An AUROC score of 1.0 is a perfect score, and 0.5 the score expected from a random method. We calculated the AUROC for each structure in SCOP30 (3295 domains; see Methods) using both unsupervised and supervised approaches. In the unsupervised approach, SCOP30 structures were ordered from a given query structure using the euclidean distance; this ordered list was compared with the true neighbors provided by SAS and SCOP criterion to quantify the corresponding AUROC. Mean AUROCs for this experiment were 0.87, 0.85, 0.84 for SAS 20, 35 and 50 respectively. For comparison, note that the best score obtained in this experiment by FragBag is 0.747 for SAS20.

Table 4.

In the supervised analysis, AUROCs were obtained by training a random forest classifier with a subset of SCOP excluding structures in SCOP30 (see Methods). Then, for a query structure, its class is predicted and the ordering of the rest of domains is determined by the log-probabilities of belonging to this predicted class.

Our mean AUROCs for SAS 20, 35 and 50 are 0.950, 0.940 and 0.935, while ContactLib achieves 0.956, 0.918 and 0.906. ContactLib reported 75% of AUROC scores > 0.936, while our minimum AUROC score for 75% of queries is 0.958. For SAS35 and SAS50, 75% of AUROC scores are above 0.939 and 0.929 respectively. Our median AUROCs for any SAS level is > 0.99.

4. Discussion

Here we introduce a protein structure representation using a Sperner family defined by clusters of protein's residues that are in close proximity in the three dimensional space. In Figure 6 it is shown the linear relationship between protein size and total number of such clusters that may be associated to the near constant density of proteins. The small deviation in the number of clusters for small proteins is also consistent with the observation that small proteins tend to have higher average densities (Fischer et al., 2004).

We found that proteins usually harbour residue clusters with no more than six residues, independently of secondary structure composition. Thus, residue clusters of size 3 and up to 6 were identified according to the contiguity of residues in sequence (see Figure 2). This Residue Cluster Class System is used to construct a vector space model representation of protein fold space in 26 dimensions (see Figure 3). This representation of protein structure may be used to test for the separability of protein folds independently of protein similarity measures (e.g., root mean square deviation) or the use of any particular classifier.

It has been shown that the cut edge weight statistic is a proper test for data separability and consequently learnability of any given data set (Zighed et al., 2002). However, this requires computing an RNG that is time consuming ($O(n^3)$ operations, where n is the number of protein structures to be analyzed; see Methods). Here we show that a HSP can be computed faster than a RNG and

Supplemental Data for

Machine Learnable Fold Space Representation based on Residue Cluster Classes

R. Corral Corral¹, Edgar Chávez² and Gabriel del Rio¹

(1) Department of Biochemistry and Structural Biology, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México D. F., México

(2) Instituto de Matemáticas, Universidad Nacional Autónoma de México

This material includes:

3 supplemental tables: Table S1, S2A, S2B, S2C and S3.

1 supplemental figure: figure S1

Table S1. CATCHOP data set.

Index	PDB ID	Chain ID	Domain ID	Index	PDB ID	Chain ID	Domain ID
1	1sid	B	00	2	2r5k	B	00
3	2r5i	F	00	4	3bwq	D	00
5	1cn3	D	00	6	3nxg	B	00
7	1qau	A	00	8	2dlu	A	00
9	2wl7	A	00	10	2jin	A	00
11	2vz5	A	00	12	1n6e	I	04
13	1yq9	I	00	14	2frv	H	00
15	3cur	I	00	16	3cur	H	00
17	1wul	L	00	18	1ubl	L	00
19	1lm2	A	00	20	2cvc	A	03
21	2e84	A	02	22	1gws	A	04
23	2cyr	0	00	24	1ofw	A	01
25	1yew	B	00	26	1yew	F	00
27	3chx	J	00	28	1yew	J	00
29	3chx	F	00	30	3chx	B	00
31	1a41	A	01	32	1a36	A	03
33	1r49	A	03	34	1t8i	A	03
35	1rrj	A	03	36	1sc7	A	03
37	2hmo	A	01	38	1ulj	E	01
39	2bmo	A	01	40	1ndo	E	01
41	2hml	A	01	42	1o7p	A	01
43	2fin	A	00	44	2grk	A	00
45	1cq3	B	00	46	2grk	B	00
47	1cq3	A	00	48	2ffk	A	00
49	1nvp	D	02	50	1nh2	C	00
51	1nvp	C	00	52	1rml	B	02
53	1nh2	D	02	54	1rml	C	02
55	3pyt	G	01	56	3d5b	H	01
57	3ccl	E	01	58	3g6e	E	02
59	3i22	G	01	60	1kc8	G	01
61	1111	C	01	62	2pfl	A	00
63	3pfl	A	00	64	1h16	A	00
65	1qhm	A	00	66	1111	A	01
67	2quy	E	00	68	2iwm	D	00
69	3pva	C	00	70	2pva	D	00
71	3pva	G	00	72	3pva	B	00
73	1qo8	D	02	74	1m1p	F	00
75	1jry	B	02	76	1ksu	B	02
77	2b7r	A	01	78	1qjd	A	02
79	1q9m	D	00	80	2h42	B	00
81	1xot	A	00	82	3dy8	B	00
83	1xmu	B	00	84	3d3p	A	00
85	1al0	F	00	86	1rb8	F	00
87	2bpa	1	00	88	1gff	1	00
89	1cd3	F	00	90	1m06	F	00
91	11tl	A	01	92	11tl	C	01
93	11tl	E	01	94	11tl	F	01
95	11tl	B	01	96	11tl	D	01
97	2au7	A	00	98	3ld3	B	00
99	2ik0	A	00	100	2uxs	A	00
101	1jfd	B	00	102	2ik9	A	00
103	1pug	C	00	104	1pug	B	00
105	1j8b	A	00	106	3f42	B	00
107	1pug	D	00	108	1pug	A	00
109	5esc	B	02	110	1ixe	B	02
111	1ixe	A	02	112	1owb	B	03
113	1ixe	D	02	114	1iom	A	02
115	1pya	F	00	116	1ibt	D	00
117	1n13	L	00	118	1n13	H	00
119	2qqd	E	00	120	2qqd	C	00
121	1beg	A	00	122	2a8f	B	00

123	1ljp	A	00	124	1bxm	A	00
125	2aib	B	00	126	1lri	A	00
127	1wlz	C	01	128	3bum	B	02
129	2wcf	A	00	130	1bt6	A	00
131	1xfv	Q	02	132	1uhj	A	00
133	1qj2	B	01	134	1yjn	H	00
135	1nvj	A	00	136	1qj2	H	01
137	3oi3	Q	01	138	1n63	B	01
139	3ko3	D	00	140	3lmu	H	00
141	2okv	B	00	142	2dbo	A	00
143	3knf	E	00	144	3kob	B	00
145	1wr8	A	02	146	1kyt	A	02
147	1wr8	B	02	148	2b1q	A	02
149	1l6r	B	02	150	1tj4	A	02
151	1hr6	E	01	152	3h1l	N	01
153	3l7l	N	01	154	3h1h	O	02
155	3l72	N	01	156	1pp9	A	01
157	3kdt	A	00	158	2q60	A	00
159	2gbd	B	00	160	3jzb	A	00
161	2i4p	B	00	162	3ads	B	00
163	3knq	A	02	164	2g3p	B	02
165	3knq	B	02	166	3dgs	A	02
167	1g3p	A	02	168	2g3p	A	02
169	1mwt	A	02	170	1vqq	B	02
171	1mwu	A	02	172	1mwr	A	02
173	1qmf	A	01	174	1vqq	A	02
175	3il5	B	01	176	1ulq	F	01
177	1ou6	B	01	178	2vba	C	02
179	2vb7	C	02	180	2ib8	D	01
181	3g76	H	00	182	1i3o	F	00
183	1q4q	B	00	184	2uvl	B	00
185	3cm2	E	00	186	2opy	A	00
187	1c3o	A	04	188	1c30	A	04
189	1cs0	A	04	190	1m6v	C	04
191	1m6v	E	04	192	1ce8	E	04
193	1hfp	A	00	194	3cse	A	00
195	1rx5	A	00	196	2c2s	B	00
197	3ntz	A	00	198	1jol	A	00
199	3ij4	A	02	200	3hgc	A	02
201	2qts	E	02	202	2qts	C	02
203	2qts	A	02	204	2qts	B	02
205	1kmc	C	00	206	2bbv	D	00
207	1i4o	C	00	208	3cj8	B	01
209	1i5l	E	00	210	3cj8	C	01
211	2wtz	D	02	212	1rk2	A	00
213	1gqt	C	00	214	2y68	A	02
215	2r3e	A	00	216	3rt7	A	02
217	1j47	A	00	218	1hrz	A	00
219	1j5n	A	00	220	1l8y	A	00
221	1gt0	D	00	222	1v64	A	00
223	1zfi	A	00	224	1m3j	B	02
225	1m3j	A	02	226	2abz	E	00
227	1m3i	C	02	228	2abz	F	00
229	2qdy	A	00	230	3qyh	A	00
231	3qyh	E	00	232	1ugq	A	00
233	2zcf	A	00	234	3a8h	A	00
235	1l2a	C	02	236	1fae	A	02
237	1lly	D	02	238	1lly	B	02
239	2qno	A	02	240	1f9o	A	02
241	1ybj	A	00	242	3mol	B	00
243	3ddr	C	00	244	3csl	C	00
245	3csn	D	00	246	3ell	B	00
247	1kca	E	00	248	1jhh	B	00
249	1ay9	A	00	250	1umu	B	00

251	1kn9	A	01	252	1jhc	A	00
253	8gep	A	02	254	3geo	A	02
255	5gep	A	02	256	6gep	A	02
257	2aop	A	02	258	4aop	A	02
259	2iuo	H	02	260	2iv1	C	02
261	1dw9	E	02	262	2iv1	I	02
263	2iu7	B	02	264	1dwk	E	02
265	luaa	B	04	266	2pir	A	04
267	2is4	A	04	268	2is1	B	04
269	3k70	B	04	270	3k70	E	04
271	1t77	C	02	272	1t77	B	02
273	1t77	D	02	274	1mi1	A	02
275	1mi1	B	02	276	1t77	A	02
277	2eim	A	00	278	1m57	A	00
279	2zxw	N	00	280	1oco	N	00
281	3ag3	N	00	282	2dys	A	00
283	1om7	A	01	284	1jiw	P	01
285	1k7i	A	01	286	1k7g	A	01
287	1smp	A	01	288	1go8	P	01
289	1aes	A	01	290	1gw2	A	01
291	1apx	C	01	292	1sj2	B	01
293	1u2j	E	01	294	2vcf	X	01
295	1ng9	A	03	296	1nne	A	03
297	1e3m	B	04	298	1wb9	A	03
299	1oh6	A	03	300	1oh7	A	03
301	2j19	A	00	302	2j18	A	00
303	2j5m	A	00	304	2ciz	A	00
305	1cpo	A	00	306	2cj2	A	00
307	3cu8	B	00	308	1a38	A	00
309	1o9f	A	00	310	1w99	A	01
311	1ib1	B	00	312	2btp	B	00
313	1i6k	A	02	314	2cyb	B	02
315	3hzz	E	02	316	4ts1	B	02
317	2quj	B	02	318	3d6u	A	02
319	1c0a	A	03	320	2r93	K	00
321	1xxa	C	00	322	1iq0	A	02
323	3gtg	K	00	324	3gtm	C	01
325	3lnz	O	00	326	3g03	C	00
327	3lnj	C	00	328	2gv2	A	00
329	1ycq	A	00	330	1t4e	A	00
331	1o9x	A	05	332	1hk2	A	02
333	2bxg	A	03	334	2bxi	A	04
335	1e7f	A	01	336	1hk5	A	06
337	3i21	M	02	338	2i2p	M	02
339	1vs7	M	02	340	2v46	M	02
341	3ora	M	02	342	2wdm	M	02
343	3ek6	A	00	344	2jjx	C	00
345	3ek6	C	00	346	2we4	B	00
347	1e19	B	00	348	3ek5	B	00
349	1sal	C	00	350	1sae	C	00
351	1saf	A	00	352	1sag	A	00
353	1sah	D	00	354	1olh	B	00
355	1lcu	A	04	356	2d1k	A	04
357	1rgi	A	04	358	3buz	B	03
359	3kvg	B	04	360	3el2	A	03
361	1wp9	D	03	362	1wp9	B	03
363	1wp9	A	03	364	1wp9	E	03
365	1wp9	C	03	366	1wp9	F	03
367	1mw8	X	02	368	1cy2	A	02
369	1cy7	A	02	370	1ecl	A	02
371	1gl9	B	06	372	1gl9	C	06
373	1du3	I	01	374	1za3	R	01
375	1p0t	A	00	376	1p0t	E	00
377	1d0g	R	02	378	1p0t	W	00

379	3d31	D	00	380	3dhw	A	00
381	2onk	D	00	382	2r6g	G	01
383	2onk	C	00	384	2r6g	F	04
385	3ltw	A	01	386	1w5r	A	01
387	1gx3	A	01	388	1gx3	B	01
389	2bsz	A	01	390	2bsz	B	01
391	1sor	A	00	392	2o9f	B	00
393	2b6o	A	00	394	3ne2	E	00
395	2abm	G	00	396	2abm	B	00
397	3m3y	E	01	398	1twf	E	01
399	3m4o	E	01	400	3hoy	E	01
401	1i50	E	01	402	3gtk	E	01
403	2vao	B	02	404	1ahz	B	02
405	2pls	H	00	406	1zxi	C	02
407	1n61	F	03	408	2mbr	A	03
409	2pfo	A	03	410	2bbj	E	01
411	1ylq	A	00	412	1jqr	A	01
413	1mq2	A	03	414	1zqn	A	03
415	1mok	A	03	416	2oqy	D	01
417	3dfh	C	01	418	2ba6	D	01
419	1ec9	D	01	420	2wov	D	03
421	1d00	B	00	422	1ca4	F	00
423	1d01	F	00	424	1ca9	F	00
425	1ca4	D	00	426	2gkw	A	00
427	3l71	I	00	428	2fyu	I	00
429	1l0l	I	00	430	1sqb	I	00
431	3h1k	V	00	432	3l72	V	00
433	2czh	A	01	434	2fix	D	01
435	1yxi	A	01	436	2fhy	D	01
437	3ifa	D	01	438	2y5l	G	01
439	2dda	B	01	440	2vzn	B	00
441	1u53	A	00	442	1xx5	C	01
443	3mz8	A	01	444	1xta	B	01
445	1ogo	X	01	446	1wmr	B	01
447	2z8g	B	01	448	1wmr	A	01
449	1x0c	B	01	450	1ogm	X	01
451	1psq	A	00	452	3hxs	B	01
453	10gs	A	01	454	1tu7	A	01
455	1se1	A	02	456	1vgs	D	01
457	3g5c	B	02	458	3c05	A	00
459	1mpz	A	00	460	3g5c	A	02
461	1fvl	A	00	462	2w9u	A	00
463	1i3r	A	01	464	1zgl	E	01
465	1t5w	D	01	466	3c6l	D	01
467	2ian	P	01	468	3mbe	B	01
469	3hmg	E	01	470	2wrc	B	02
471	1hgg	E	01	472	2wrb	C	02
473	1mqm	G	01	474	1hgh	A	01
475	2hnw	A	00	476	1jk4	A	00
477	2hnw	B	00	478	2hnu	B	00
479	2hnu	D	00	480	2hnw	E	00
481	1c97	A	03	482	8acn	A	03
483	5acn	A	03	484	1b0k	A	03
485	1c97	A	01	486	1aco	A	03
487	3rjz	A	02	488	3rk1	B	02
489	3rk0	A	02	490	2d13	A	02
491	3h7e	A	02	492	1ru8	A	02
493	3eql	M	04	494	3eql	C	04
495	2o5j	C	04	496	2a6h	M	04
497	2o5i	C	04	498	1ddq	C	04
499	1xaf	B	00	500	1xaf	A	00
501	1z9t	A	00	502	1t8h	A	00
503	1xfj	A	00	504	1u05	A	00
505	1qag	A	01	506	2wa6	A	01

507	3fer	B	01	508	3f7p	A	01
509	1pxy	B	04	510	3hor	A	01
511	2k1d	A	00	512	3heq	B	00
513	1qm0	A	00	514	2w9e	A	00
515	1fkc	A	00	516	2k5o	A	00
517	1oi9	D	01	518	1oku	D	02
519	1jsu	B	01	520	3bhv	D	02
521	2x1n	B	02	522	1unl	E	00
523	1q86	N	00	524	1m1k	N	00
525	1k8a	N	00	526	3ccs	M	00
527	3i56	M	00	528	1jj2	L	00
529	1s28	D	00	530	1s28	B	00
531	1k6z	B	00	532	1k8k	D	02
533	1l2w	D	00	534	1l2w	H	00
535	4hir	A	00	536	1hrt	I	00
537	1e0f	I	01	538	6hir	A	00
539	5hir	A	00	540	4htc	I	00
541	2sli	A	03	542	2vw2	A	03
543	1sli	A	03	544	1sll	A	03
545	4sli	A	03	546	3sli	A	03
547	2czj	G	00	548	1j1h	A	00
549	2czj	E	00	550	1wjx	A	00
551	1p6v	C	00	552	1p6v	A	00
553	1toq	C	00	554	3lm1	O	00
555	1j4u	B	00	556	1zgs	A	01
557	1j4t	F	00	558	1kuj	C	00
559	1ifw	A	00	560	2o4t	A	00
561	2ghj	D	02	562	2hh6	A	00
563	1jgn	A	00	564	2ghj	B	02
565	2b5i	A	00	566	1exz	B	00
567	1pv7	B	01	568	3b5k	A	00
569	1fg9	A	00	570	1hwh	A	00
571	1otp	A	01	572	1uou	A	01
573	1zyk	D	01	574	1v8g	A	01
575	2j0f	B	01	576	1o17	D	01
577	1cc4	A	02	578	2phh	A	02
579	1kif	H	02	580	2gf3	B	02
581	1eli	A	02	582	2e82	B	02
583	3os4	A	00	584	2im5	A	00
585	1yir	D	00	586	1vlp	B	00
587	1vlp	A	00	588	2im5	D	00
589	1xqh	E	02	590	1xqh	A	02
591	3cbp	A	02	592	1o9s	B	02
593	1n3j	A	00	594	1ml9	A	00
595	2dxi	A	03	596	2dxi	B	03
597	1n77	A	02	598	1zjw	A	03
599	1euq	A	03	600	1j09	A	02
601	1ss8	C	02	602	1q2v	C	02
603	2nwc	B	02	604	1kp8	F	02
605	1we3	N	02	606	1q3q	A	02
607	3kak	B	03	608	3kaj	B	03
609	3kal	B	03	610	1m0w	B	03
611	1m0t	B	03	612	1m0t	A	03
613	1huu	C	00	614	1p78	A	00
615	1owg	B	00	616	1owf	A	00
617	3rhi	D	00	618	2o97	B	00
619	1nnn	A	02	620	2p2m	A	02
621	1o9t	B	02	622	1ega	B	02
623	1pg4	A	02	624	3df1	C	01
625	1k98	A	02	626	2o2k	A	01
627	3bul	A	03	628	1k7y	A	02
629	3iv9	A	03	630	2o2k	B	01
631	2qf7	B	04	632	3bg9	A	01
633	3bg3	C	01	634	3bg3	B	01

635	3bg9	D	01	636	3bg9	B	01
637	1hjaz	A	00	638	1hjaz	B	00
639	2j9a	A	01	640	2bfr	A	00
641	1lcp	B	01	642	3h8g	C	01
643	2v29	A	00	644	4fua	A	00
645	1e46	P	00	646	1pvt	A	00
647	1gt7	P	00	648	2v9l	A	00
649	2vpz	C	00	650	2vpw	C	00
651	2vpz	G	00	652	2vpx	C	00
653	2vpy	C	00	654	2vpx	G	00
655	1h5n	A	02	656	1e5v	C	02
657	1g8k	A	03	658	1ogy	I	03
659	1h0h	A	03	660	1e60	A	02
661	1l2g	A	00	662	2c3a	B	00
663	2c36	A	00	664	1l2g	B	00
665	1jma	A	00	666	1l2g	D	00
667	3mya	A	00	668	2k6m	S	00
669	2rjx	A	00	670	3myc	A	00
671	1ujs	A	00	672	2k6n	A	00
673	1ds6	B	00	674	1kmt	B	00
675	2jhv	F	00	676	2bem	B	00
677	1ft3	B	00	678	2jhv	C	00
679	8tln	E	01	680	2a7g	E	01
681	1ezm	A	01	682	1kro	A	01
683	7tli	A	01	684	1hyt	A	01
685	1rlz	A	00	686	1roz	B	00
687	1roz	A	00	688	1rqd	B	00
689	1rqd	A	00	690	1dhs	A	00
691	1o8l	A	00	692	1clw	A	00
693	1dbg	A	00	694	1idj	B	00
695	1dbo	A	00	696	1o8h	A	00
697	3fna	A	00	698	3ctu	A	00
699	3k2v	B	00	700	3ctu	B	00
701	3k6e	B	00	702	3k6e	A	00
703	2gmy	E	00	704	2oyo	B	02
705	2pfx	B	02	706	2pr	B	02
707	2pr	C	02	708	1gu9	D	00
709	3qi5	B	00	710	3qi5	A	00
711	1f4r	A	00	712	1bnk	A	00
713	1f6o	A	00	714	1ewn	A	00
715	3ogv	A	02	716	3ogs	A	02
717	3ogr	A	02	718	3og2	A	02
719	1xc6	A	02	720	1tg7	A	02
721	2h4j	A	02	722	1j8f	C	02
723	2h2f	A	02	724	2od2	A	02
725	1q17	B	02	726	2od9	A	02
727	1tue	G	02	728	1tue	Q	02
729	1qqh	A	00	730	1dto	A	02
731	1tue	J	02	732	1r6n	A	02
733	3hux	R	00	734	1gd8	B	00
735	2y17	R	00	736	3oh7	R	00
737	2y13	R	00	738	3pyr	N	00
739	1bol	A	00	740	1ucc	A	00
741	1ioo	A	00	742	1iyb	B	00
743	1vd1	A	00	744	1j1f	A	00
745	1j3k	D	00	746	3b9h	A	00
747	1axw	B	00	748	1bsp	A	00
749	1jut	A	00	750	1ev8	A	00
751	3g6e	B	01	752	1yj9	B	01
753	1yhq	B	01	754	1vq5	B	01
755	2qex	B	01	756	3g4s	B	01
757	1g8r	B	01	758	1r2k	B	00
759	1o8q	A	00	760	2nqq	D	01
761	1o8n	C	00	762	1wu2	B	01

763	1dd9	A	01	764	1eqn	E	01
765	1eqn	C	01	766	1eqn	A	01
767	1dde	A	01	768	3b39	B	01
769	1rkm	A	03	770	3m8u	A	03
771	1b9j	A	03	772	1jev	A	03
773	1vr5	B	03	774	1b3f	A	03
775	3hup	B	00	776	3c8j	D	00
777	1v4l	F	00	778	1ja3	A	00
779	1k9i	C	00	780	1sl6	A	00
781	1ypa	I	00	782	1egp	A	00
783	1dwm	A	00	784	1tec	I	00
785	1y34	I	00	786	1cis	A	00
787	1bse	C	00	788	1bsc	B	00
789	2bo9	B	01	790	1nww	B	00
791	1gy5	B	00	792	1jbs	A	00
793	2b76	A	04	794	1rmu	4	00
795	2pzs	A	05	796	3ae1	A	04
797	2wu5	A	04	798	1ev1	4	00
799	3aso	Z	00	800	1jo6	A	00
801	2ein	M	00	802	3abm	M	00
803	1ocz	Z	00	804	2eik	Z	00
805	2c6s	F	02	806	2c6s	N	02
807	2c6s	E	02	808	2c6s	O	02
809	2c6s	K	02	810	2c6s	G	02
811	3m1p	A	00	812	2bet	D	00
813	3k8c	B	00	814	1nn4	A	00
815	2bes	C	00	816	3qd5	B	00
817	2wdh	E	01	818	2uu9	E	01
819	1g2f	F	01	820	2qan	E	01
821	1hnz	E	01	822	1un6	B	03
823	1mw8	X	04	824	1cy4	A	04
825	1mw9	X	04	826	1d6m	A	04
827	1cy0	A	04	828	1cy2	A	04
829	2q17	B	00	830	1z70	X	00
831	2aft	X	00	832	2aii	X	00
833	2afy	X	00	834	2aik	X	00
835	2o01	4	01	836	1rwt	H	01
837	1rwt	A	01	838	1rwt	J	01
839	1rwt	I	01	840	1rwt	D	01
841	3nzw	Z	00	842	1fm2	B	01
843	1ofd	B	01	844	3bdm	Z	00
845	1yau	F	00	846	3nzj	2	00
847	1s3m	B	00	848	3h62	B	00
849	1ii7	A	01	850	1wao	4	02
851	1ush	A	01	852	1oid	B	01
853	1niu	B	02	854	3e6e	B	01
855	3b8v	C	01	856	2rjh	B	01
857	1qu4	D	01	858	2dy3	C	01
859	1ztp	B	01	860	1ejh	A	00
861	1ap8	A	00	862	1ejl	B	00
863	1ej4	A	00	864	1ejh	D	00
865	1pok	B	01	866	1kl1	C	01
867	1ejw	C	01	868	2i9u	A	01
869	3n2c	D	01	870	3feq	P	01
871	1pkl	B	03	872	1a5u	H	03
873	1lix	D	03	874	1aqf	H	03
875	2g50	H	03	876	2vgb	D	03
877	3ozv	B	02	878	2gre	G	02
879	3i1f	A	03	880	1sm4	B	01
881	1xfo	B	02	882	3oe7	T	01
883	1s3i	A	02	884	1fmt	A	02
885	1fmt	B	02	886	2bw0	A	02
887	2fmt	B	02	888	3r8x	A	02
889	1ohh	A	03	890	2jiz	J	03

891	3oe7	K	03	892	3oee	A	03
893	1h8e	B	03	894	1w0j	A	03
895	1fqk	D	01	896	5req	D	03
897	2d9j	A	01	898	1ezy	A	01
899	2af0	A	01	900	1fqi	A	01
901	1dbx	A	00	902	1dbu	A	00
903	2dxa	A	00	904	1wdv	A	00
905	1wdv	B	00	906	1vjf	A	00
907	3m95	B	00	908	1cqy	A	02
909	2w54	A	01	910	3fil	A	00
911	1lgr	F	01	912	3p4p	B	01
913	2bbk	L	00	914	2agx	D	00
915	2agy	H	00	916	2j57	N	00
917	2ok6	D	00	918	2h3x	B	00
919	1t8r	C	01	920	1t8s	C	01
921	1t8w	D	01	922	1t8w	A	01
923	1t8y	F	01	924	1t8r	D	01
925	3dob	B	00	926	2bpr	A	00
927	3dpq	E	01	928	3dpq	A	01
929	1dkx	A	01	930	1q5l	A	00
931	1kpo	X	01	932	1wf4	k	01
933	1svt	L	01	934	1mnf	F	01
935	1sx4	J	01	936	1j4z	H	01
937	1vkz	B	04	938	2xcl	A	04
939	1vkz	A	04	940	2yw2	B	04
941	2yrw	A	04	942	1gso	A	04
943	1b4n	D	02	944	1b4n	A	02
945	1aor	B	02	946	1b4n	B	02
947	1b25	B	02	948	1b25	C	02
949	1fjg	P	00	950	2vqe	P	00
951	2f4v	P	00	952	2ilc	q	00
953	1xmo	P	00	954	1hnw	P	00
955	1jvs	A	03	956	2egh	A	03
957	3anm	A	03	958	3ann	A	03
959	2y1c	B	03	960	2jcy	A	03
961	3lxv	A	00	962	3pch	A	00
963	2pcd	D	00	964	3pcn	F	00
965	3pck	A	00	966	2buw	A	00
967	2qgq	A	01	968	2qgq	C	01
969	2qgq	E	01	970	3can	A	00
971	2qgq	D	01	972	2qgq	G	01
973	1j2z	A	01	974	3fsy	B	03
975	3fsx	B	03	976	1mr9	B	00
977	1sst	B	02	978	2oi5	B	02
979	2w1b	A	05	980	2v50	C	05
981	2hrt	C	05	982	3noc	B	05
983	3nog	A	05	984	1oye	A	05
985	1nda	C	01	986	2bcg	G	01
987	2gr3	A	01	988	1qla	A	01
989	1gxf	B	01	990	1grt	A	02
991	1n7z	A	02	992	1n80	D	02
993	1n8b	C	02	994	1n8b	A	02
995	1n7z	C	02	996	1n8b	D	02
997	1vde	B	01	998	1at0	A	00
999	1ef0	B	01	1000	1vde	A	01
1001	1ef0	A	01	1002	1lws	A	01
1003	1vk1	A	01	1004	2hwj	F	01
1005	2hwj	D	01	1006	2hwj	A	01
1007	2hwj	B	01	1008	2hwj	E	01
1009	2a68	C	01	1010	1hqm	C	01
1011	1ynj	C	01	1012	2ppb	M	01
1013	2o5j	C	01	1014	1iw7	C	01
1015	1ly8	A	02	1016	2dv2	B	04
1017	1atj	E	02	1018	1yzz	A	02

1019	1x7u	A	04	1020	1b80	B	02
1021	1hlq	A	00	1022	1hrq	A	00
1023	1pij	A	00	1024	3hip	C	00
1025	3h31	A	00	1026	2fla	A	00
1027	3hk8	A	02	1028	3hk9	E	02
1029	3hka	B	02	1030	3hk7	I	02
1031	1j5s	A	02	1032	3hk7	E	02
1033	3e9i	B	02	1034	1qe0	B	01
1035	2rhs	D	05	1036	2aly	B	05
1037	1qe0	A	01	1038	3dsq	B	02
1039	1nj8	B	03	1040	2gtc	D	00
1041	1h4s	A	03	1042	1y2i	A	00
1043	2q3v	A	00	1044	1je3	A	01
1045	1z4q	A	02	1046	1i0a	A	03
1047	1tjw	A	03	1048	1k62	B	03
1049	3e04	D	03	1050	1u0l	C	03
1051	1spr	D	00	1052	1a81	K	02
1053	1tce	A	00	1054	2ptk	A	02
1055	1b47	A	03	1056	1rja	A	00
1057	1r6b	X	01	1058	3fes	D	00
1059	2k77	A	00	1060	1mbv	A	00
1061	1r6q	B	00	1062	1qvr	A	01
1063	1ahz	A	04	1064	1e8h	A	04
1065	1e8g	A	04	1066	1diq	A	04
1067	1e8f	B	04	1068	1qlt	B	04
1069	2v1k	A	01	1070	1m05	C	01
1071	2clr	A	01	1072	1bz9	A	01
1073	3h9s	A	01	1074	1kti	A	01
1075	3g74	B	00	1076	3g74	E	00
1077	3g74	C	00	1078	3g74	F	00
1079	3g74	H	00	1080	3g74	D	00
1081	1ahv	B	03	1082	1e8h	A	03
1083	2qpm	A	03	1084	1wvf	A	03
1085	2i0k	A	03	1086	1w1m	B	03
1087	1xty	B	00	1088	2zv3	C	00
1089	2zv3	E	00	1090	2gax	A	00
1091	1q7s	B	00	1092	2zv3	A	00
1093	3bjq	E	00	1094	3bjq	J	00
1095	3bjq	A	00	1096	3bjq	H	00
1097	3bjq	F	00	1098	3bjq	G	00
1099	1fm8	A	01	1100	1bg8	B	00
1101	1eyq	A	01	1102	1eyp	A	01
1103	1fm8	B	01	1104	1eyq	B	01
1105	1hxr	B	00	1106	111d	A	00
1107	2kzn	A	00	1108	2hr9	A	01
1109	1fwq	A	00	1110	1xm0	A	00
1111	1a8r	A	02	1112	1fb1	E	02
1113	1a9c	G	02	1114	1n3t	F	02
1115	1b9l	F	00	1116	1n3t	A	02
1117	3ckc	B	03	1118	3ck7	A	03
1119	3ck7	D	03	1120	3ckb	B	03
1121	3ck7	C	03	1122	3ckb	A	03
1123	2b4c	G	00	1124	2i5y	P	00
1125	1rzk	G	00	1126	1yyl	G	00
1127	2nxz	A	00	1128	1g9m	G	00
1129	1kg9	A	00	1130	2i1x	A	00
1131	3kj6	A	00	1132	2vt4	D	00
1133	1pxr	A	00	1134	2ycw	B	00
1135	1rz0	B	00	1136	3dnh	A	01
1137	3cp3	A	00	1138	1wli	B	00
1139	3k86	A	00	1140	2x1k	A	00
1141	2xk6	A	02	1142	3o17	B	02
1143	2fo0	A	04	1144	1nxk	B	02
1145	2bmc	B	02	1146	3e8e	A	01

1147	3ccv	2	00	1148	1kc8	3	00
1149	2qa4	2	00	1150	1vqo	2	00
1151	3o58	e	00	1152	2otl	2	00
1153	2uza	B	04	1154	2pda	A	04
1155	2c3m	B	04	1156	2pda	B	04
1157	2c3p	A	04	1158	2c42	B	04
1159	2ioj	B	00	1160	2fyw	C	01
1161	2fyw	C	02	1162	1nmo	A	01
1163	1nmo	F	01	1164	1knx	D	01
1165	2a5w	A	01	1166	3or1	C	01
1167	3or1	F	01	1168	1sau	A	01
1169	1ji8	A	01	1170	2a5w	B	01
1171	2wcd	N	00	1172	2wcd	P	00
1173	2wcd	D	00	1174	2wcd	W	00
1175	1qoy	A	00	1176	2wcd	B	00
1177	1s4d	A	01	1178	2dxw	A	01
1179	2e07	B	01	1180	2dsg	A	01
1181	2dsi	B	01	1182	2huv	A	01
1183	1hp0	B	00	1184	1kie	A	00
1185	3mkn	A	00	1186	3mkm	D	00
1187	1hoz	A	00	1188	1yoe	A	00
1189	1i0e	D	02	1190	1f52	I	01
1191	1f52	H	01	1192	2gls	F	01
1193	1m80	A	02	1194	2gls	H	01
1195	1xf6	A	00	1196	1qgw	B	00
1197	1xg0	B	00	1198	1xg0	A	00
1199	1xf6	B	00	1200	1qgw	A	00
1201	1vll	A	01	1202	1x7d	B	01
1203	1u7h	A	01	1204	1u7h	B	01
1205	1x7d	A	01	1206	1omo	B	01
1207	3ekf	B	00	1208	1og2	A	00
1209	3me6	D	00	1210	3m4v	A	00
1211	2h7r	A	00	1212	2ij4	A	00
1213	1eth	D	00	1214	1lpa	A	00
1215	1imt	A	00	1216	1pco	A	00
1217	1pcn	A	00	1218	1eth	B	00
1219	3do6	B	02	1220	1eg7	A	02
1221	1fpm	B	02	1222	1fp7	A	02
1223	1fp7	B	02	1224	1eg7	B	02
1225	1rqf	B	01	1226	1qf8	A	01
1227	1qf8	B	01	1228	1rqf	D	01
1229	1jwh	D	01	1230	1rqf	E	01
1231	1jch	A	03	1232	2b5u	A	03
1233	1e44	B	00	1234	1jch	C	03
1235	2xfz	Y	00	1236	2b5u	C	03
1237	3fr8	A	02	1238	2ew2	B	02
1239	2iyp	A	02	1240	1yj8	B	02
1241	2pla	A	02	1242	1qmg	C	02
1243	1fvh	A	03	1244	1dn1	A	02
1245	1mqs	A	03	1246	1fvf	A	03
1247	1epu	A	02	1248	1fvf	B	03
1249	3px2	D	02	1250	3px3	D	02
1251	3bxo	A	02	1252	3pfh	A	02
1253	3pfh	D	02	1254	3bxo	B	02
1255	2fu3	A	02	1256	1t3e	B	02
1257	2nqu	A	02	1258	1xi8	B	02
1259	1g8l	A	02	1260	1g8l	B	02
1261	2knx	A	00	1262	1k7b	A	00
1263	1cr8	A	00	1264	1ldr	A	00
1265	1ldl	A	00	1266	1n7d	A	01
1267	3l4j	A	03	1268	1ab4	A	01
1269	2rgr	A	03	1270	1bgw	A	05
1271	1bjt	A	05	1272	3l4k	A	03
1273	1o7k	A	00	1274	2csk	A	00

1275	1xtn	B	00	1276	1xte	A	00
1277	1ocs	A	00	1278	1xtn	A	00
1279	3ccu	C	00	1280	2otj	C	00
1281	1vq9	C	00	1282	1qvf	C	00
1283	3ccm	C	00	1284	3ccr	C	00
1285	2b26	B	02	1286	1eb0	A	01
1287	1gmu	C	02	1288	1gmu	A	02
1289	2jz8	A	00	1290	1gmv	B	02
1291	3kd1	E	01	1292	1clq	A	01
1293	1noy	B	01	1294	1ig9	A	01
1295	3ngi	A	01	1296	3cfr	A	01
1297	2oya	B	00	1298	2oy3	A	00
1299	2oya	A	00	1300	1z8g	A	01
1301	1o5e	L	00	1302	1o5f	L	00
1303	2cv2	A	02	1304	3akz	D	02
1305	1gln	A	03	1306	1n78	B	03
1307	1o0b	A	04	1308	3afh	A	02
1309	1a8h	A	02	1310	2x11	A	02
1311	1rqg	A	02	1312	1woy	A	02
1313	2x1m	A	02	1314	3kfl	A	02
1315	2otn	B	02	1316	2h9f	A	01
1317	3ekm	D	02	1318	3ekm	E	02
1319	1u1x	A	01	1320	1xub	A	02
1321	1fz2	E	01	1322	1mhy	G	02
1323	1xu5	F	01	1324	1xmg	F	01
1325	1xvg	F	01	1326	2rkh	A	02
1327	2b0w	B	03	1328	3mpz	C	00
1329	2iu0	B	02	1330	2fr6	C	00
1331	1vk9	A	00	1332	1ux0	B	00
1333	1vs6	I	01	1334	2e34	A	01
1335	2jl6	K	01	1336	2qam	I	01
1337	2wh2	K	01	1338	2qov	I	01
1339	1uf7	B	00	1340	2fku	A	00
1341	1j31	B	00	1342	2ggk	B	00
1343	1j31	D	00	1344	2ggk	D	00
1345	1nqg	A	02	1346	2hdi	A	02
1347	2grx	B	02	1348	1oh2	Q	00
1349	3m8d	A	02	1350	1nqf	A	02
1351	1eak	C	05	1352	1gxd	B	04
1353	1l6j	A	03	1354	1h8p	B	02
1355	1e6p	B	03	1356	1ogb	A	03
1357	2jb7	A	00	1358	1rlh	A	02
1359	1vgg	A	00	1360	1wvq	A	00
1361	1vgg	D	00	1362	1wvq	B	00
1363	1m3e	D	02	1364	1xr4	B	02
1365	2hj0	B	02	1366	3oxo	F	02
1367	1o9l	B	02	1368	3cdk	B	00
1369	1gn1	G	00	1370	1j4z	I	03
1371	1ss8	G	03	1372	1sx3	N	03
1373	1j4z	C	03	1374	1we3	H	03
1375	3gvd	C	01	1376	1sst	C	01
1377	1s80	A	01	1378	3gvd	B	01
1379	1ssq	D	01	1380	3gvd	L	01
1381	1h8x	B	00	1382	2nui	A	00
1383	1xpt	A	00	1384	2g8r	A	00
1385	2p4a	C	00	1386	2c05	A	00
1387	1fx3	A	00	1388	2o2a	D	00
1389	2o2a	B	00	1390	1ozb	B	00
1391	1ozb	H	00	1392	1qyn	D	00
1393	1cxt	A	03	1394	2dmr	A	03
1395	4dmr	A	03	1396	3dmr	A	03
1397	1tmo	A	03	1398	1dmr	A	03
1399	2oeq	A	00	1400	2oeq	A	00
1401	2iaz	C	00	1402	2oeq	D	00

1403	2iaz	A	00	1404	2iaz	B	00
1405	1tcz	D	00	1406	1c5e	B	00
1407	1td3	A	00	1408	1tcz	E	00
1409	1td0	D	00	1410	1td0	A	00
1411	1myk	B	00	1412	1myk	A	00
1413	1b01	A	00	1414	2hzv	G	01
1415	1ea4	G	00	1416	2hzv	A	01
1417	1n3r	I	01	1418	1n3t	O	01
1419	1n3s	F	01	1420	1fbx	M	01
1421	1fbx	G	01	1422	1is7	J	01
1423	1pfo	A	01	1424	1m3i	D	01
1425	1m3i	C	01	1426	1s3r	A	01
1427	3cqf	B	01	1428	1m3j	A	01
1429	3cct	B	02	1430	1hwi	A	02
1431	1hw8	C	02	1432	2r4f	D	02
1433	3bgl	D	02	1434	3cda	A	02
1435	1fhp	4	00	1436	1bbt	4	00
1437	2wzr	4	00	1438	1qqp	4	00
1439	1tme	4	00	1440	1zba	4	00
1441	2w8i	B	03	1442	2w8h	D	03
1443	2j58	H	03	1444	2w8h	A	03
1445	2w8i	A	03	1446	2w8h	C	03
1447	2gfv	A	00	1448	2gfv	C	00
1449	2gfv	D	00	1450	2gfv	F	00
1451	2gfv	B	00	1452	2gfv	E	00
1453	1vfp	B	01	1454	1xp5	A	01
1455	3ar8	A	01	1456	2ear	A	01
1457	3ar3	A	01	1458	3fps	A	01
1459	1t60	V	00	1460	1li1	C	00
1461	1li1	E	00	1462	1t61	C	00
1463	1m3d	G	00	1464	1m3d	D	00
1465	1cg0	A	03	1466	1soo	A	03
1467	1j4b	A	03	1468	1kkb	A	03
1469	1gim	A	03	1470	1lny	B	03
1471	2e51	E	02	1472	2qal	E	02
1473	2ier	B	01	1474	2fld	B	01
1475	2qb9	I	00	1476	2ilc	f	02
1477	1hcq	A	00	1478	1a6y	B	00
1479	2gat	A	00	1480	3g8x	A	00
1481	1ynw	A	00	1482	1hcq	E	00
1483	2amc	B	01	1484	1pys	B	01
1485	1eiy	B	01	1486	1mfq	B	00
1487	1b7y	B	01	1488	1kvn	A	00
1489	1zm2	D	00	1490	1mdt	A	01
1491	1xdt	T	01	1492	3ny6	A	00
1493	1xk9	B	00	1494	1xk9	A	00
1495	1ecr	A	02	1496	2i05	A	02
1497	1tje	A	03	1498	1qf6	A	03
1499	2i06	A	02	1500	1tkg	A	03
1501	1kh7	B	00	1502	3m7v	A	01
1503	3bdh	A	00	1504	1ejj	A	02
1505	3dpc	B	00	1506	1shq	A	00
1507	1jhs	A	00	1508	2i8g	A	00
1509	1tu1	B	00	1510	2vu4	A	00
1511	1eq6	A	00	1512	1v2b	A	00
1513	1bqa	B	02	1514	2vib	A	02
1515	3gbx	B	01	1516	1rv4	B	02
1517	1uu1	B	02	1518	1qj3	A	02
1519	1tkg	A	02	1520	1tke	A	02
1521	1tky	A	02	1522	1xvi	B	02
1523	1tje	A	02	1524	1xvi	A	02
1525	2ox6	A	00	1526	2ox6	C	00
1527	1s4k	B	00	1528	2ox6	D	00
1529	2ox6	B	00	1530	1s4k	A	00

1531	1oo0	A	00	1532	2x1g	D	00
1533	2j0q	C	00	1534	2j0q	F	00
1535	3ex7	E	00	1536	1p27	C	00
1537	3cdd	A	01	1538	3cdd	F	01
1539	3cdd	C	01	1540	3cdd	E	01
1541	3cdd	D	01	1542	3cdd	B	01
1543	1xd4	A	04	1544	1nvx	S	02
1545	1nvu	S	02	1546	1xd2	C	02
1547	1xd4	B	04	1548	1bkd	S	02
1549	1eak	C	01	1550	1eak	B	01
1551	1lbu	A	01	1552	1eak	D	01
1553	1l6j	A	01	1554	1eak	A	01
1555	2rcy	C	02	1556	2amf	D	02
1557	2rcy	E	02	1558	2amf	C	02
1559	2amf	B	02	1560	2amf	E	02
1561	3lri	A	00	1562	1zei	C	00
1563	1gZR	B	00	1564	1efe	A	00
1565	2dsq	I	00	1566	2dsp	I	00
1567	2nwl	B	00	1568	1xfh	B	00
1569	2nwl	A	00	1570	2nww	C	00
1571	2nwx	A	00	1572	1xfh	A	00
1573	1ig4	A	00	1574	3c2i	A	01
1575	1d9n	A	00	1576	1qk9	A	00
1577	1ub1	A	00	1578	2ky8	A	00
1579	1e8o	C	00	1580	2w9j	A	00
1581	1e8o	A	00	1582	1e8o	D	00
1583	1e8o	B	00	1584	1914	A	00
1585	2jx8	A	00	1586	2ysg	A	01
1587	3kah	A	01	1588	2dwv	B	01
1589	1o6w	A	02	1590	1o6w	A	01
1591	2zih	B	00	1592	2zih	A	00
1593	2zii	A	00	1594	2zii	D	00
1595	2zih	D	00	1596	2zii	B	00
1597	2c3p	A	03	1598	1kek	A	03
1599	2c3m	B	03	1600	2c3o	B	03
1601	2uza	B	03	1602	2c3p	B	03
1603	1hvr	A	00	1604	1ls5	B	02
1605	1mer	B	00	1606	2qci	A	00
1607	1flh	A	01	1608	3ixj	B	01
1609	1tx9	A	00	1610	1cd3	4	00
1611	1a10	4	00	1612	1cd3	1	00
1613	1a10	2	00	1614	1cd3	2	00
1615	1xhx	D	03	1616	1xhx	B	03
1617	2pyj	A	03	1618	1xhz	B	03
1619	2pyl	A	03	1620	2ex3	C	03
1621	1n8r	M	01	1622	1nji	M	01
1623	3cc4	L	01	1624	3ccv	L	01
1625	3ccm	L	01	1626	1s72	L	01
1627	6ame	A	00	1628	1c89	A	02
1629	7ame	A	00	1630	3msi	A	00
1631	1hg7	A	00	1632	3ame	A	00
1633	3gce	A	00	1634	2d2c	D	02
1635	1zrt	E	02	1636	3l75	R	02
1637	1bcc	E	02	1638	3h1h	E	02
1639	3ene	A	05	1640	3csf	A	05
1641	1e90	A	05	1642	3dpd	A	05
1643	2chz	A	05	1644	1he8	A	05
1645	1avy	C	00	1646	1byy	A	00
1647	3ae8	C	01	1648	2iz0	A	03
1649	3a0h	h	01	1650	2bas	A	02
1651	1vgr	A	02	1652	1xk7	A	02
1653	2vjq	A	02	1654	1p5r	A	02
1655	1p5r	B	02	1656	2vjk	A	02
1657	2gk9	C	01	1658	2gk9	A	01

1659	2gk9	B	01	1660	2gk9	D	01
1661	1bo1	B	01	1662	1bo1	A	01
1663	1d8i	C	00	1664	1d8h	A	00
1665	1d8i	B	00	1666	1d8h	B	00
1667	1d8i	A	00	1668	1d8h	C	00
1669	3ilx	A	01	1670	3ilw	B	01
1671	3mg1	B	01	1672	3ilv	A	01
1673	1m98	B	01	1674	3mg2	A	01
1675	1kdg	B	02	1676	1cbo	A	02
1677	1kdg	A	02	1678	1naa	B	02
1679	2gew	A	02	1680	1n4v	A	02
1681	1cns	B	02	1682	1dxj	A	02
1683	2z39	A	02	1684	2baa	A	02
1685	3hbe	X	02	1686	2z39	B	02
1687	2isg	A	00	1688	1zn3	B	00
1689	1ile	A	01	1690	1yvg	A	00
1691	3fie	B	00	1692	1xtf	B	00
1693	3spu	D	00	1694	3m8t	A	00
1695	2y87	A	00	1696	1jje	A	00
1697	3s0z	A	00	1698	1sml	A	00
1699	3bf0	C	02	1700	3bf0	B	02
1701	3bez	A	02	1702	3bf0	A	02
1703	3bez	C	02	1704	3bf0	D	02
1705	1b4n	C	01	1706	1b25	A	01
1707	1b25	B	01	1708	1aor	B	01
1709	1b25	C	01	1710	1aor	A	01
1711	2v0s	A	00	1712	1nzh	A	00
1713	1y21	A	00	1714	1e9n	B	00
1715	2a40	E	00	1716	1dew	A	00
1717	1x6a	A	01	1718	1j2o	A	00
1719	1b8t	A	02	1720	1a7i	A	00
1721	2dfy	X	02	1722	1rut	X	02
1723	1t3e	B	03	1724	2nqq	C	03
1725	2nqr	B	03	1726	2nqk	A	03
1727	2nqm	B	03	1728	2nqs	A	03
1729	1smn	A	00	1730	1g8t	B	00
1731	1smn	B	00	1732	1qae	B	00
1733	1ql0	B	00	1734	1g8t	A	00
1735	2rg9	A	01	1736	2jdl	A	01
1737	1tcs	A	01	1738	3nfm	A	01
1739	1hwn	A	01	1740	1qcj	B	01
1741	3axy	B	00	1742	2qyq	A	00
1743	1wkp	D	00	1744	1fjj	A	00
1745	1b7a	A	00	1746	1beh	B	00
1747	1w27	B	02	1748	1dcn	C	02
1749	3no9	B	02	1750	2o7d	E	02
1751	1tju	D	02	1752	1dof	D	01
1753	3mli	B	00	1754	2ouf	A	00
1755	2bo3	A	01	1756	3mli	A	01
1757	3mli	D	01	1758	3mli	C	01
1759	2wqt	B	00	1760	1qco	B	02
1761	2wqt	E	00	1762	1nkq	B	00
1763	1i7o	C	01	1764	2wqt	J	00
1765	2vr0	C	00	1766	2j7a	C	00
1767	2j7a	O	00	1768	2j7a	I	00
1769	2j7a	F	00	1770	2j7a	L	00
1771	3slp	C	00	1772	1hbo	C	00
1773	1hbm	F	00	1774	3m30	F	00
1775	3sm4	A	00	1776	1w36	B	05
1777	1y5i	B	03	1778	1y4z	B	03
1779	3ir6	B	03	1780	1siw	B	03
1781	1y5l	B	03	1782	1q16	B	03
1783	1ief	A	00	1784	1nju	A	00
1785	1ied	A	00	1786	1nkk	D	00

1787	2pbk	A	00	1788	1o6e	A	00
1789	3om6	C	01	1790	1gyh	A	00
1791	3cpn	A	00	1792	1yi7	D	01
1793	2exi	D	01	1794	3om6	D	01
1795	3bom	B	00	1796	3d4x	D	00
1797	6hbi	B	00	1798	1yvq	D	00
1799	1ye0	D	00	1800	1a0w	B	00
1801	1am4	C	00	1802	3byi	C	00
1803	low3	A	00	1804	1pbw	B	00
1805	1xa6	A	03	1806	3byi	B	00
1807	1ini	A	00	1808	1h7t	A	00
1809	1g1l	H	00	1810	2v0l	A	01
1811	1g0r	E	00	1812	1hv9	B	01
1813	1fui	F	03	1814	1fui	E	03
1815	1fui	D	03	1816	3a9t	C	03
1817	3a9r	A	03	1818	3a9t	A	03
1819	1ueh	A	00	1820	1f75	B	00
1821	1x07	A	00	1822	2e9c	A	00
1823	1v7u	A	00	1824	1jp3	B	00
1825	1f5a	A	03	1826	2o1p	A	01
1827	1px5	A	02	1828	2o1p	B	01
1829	1fa0	B	03	1830	1px5	B	02
1831	1rkx	D	02	1832	1kep	A	02
1833	2vuu	B	02	1834	1g1a	A	02
1835	1bsv	A	02	1836	2vuu	D	02
1837	1wz2	B	03	1838	1a8h	A	03
1839	1bs2	A	02	1840	1pg2	A	02
1841	1obc	A	04	1842	3h9c	A	03
1843	1zzj	A	00	1844	1ug8	A	00
1845	3i8h	K	01	1846	2axy	D	00
1847	1vig	A	00	1848	2y12	H	01
1849	1bvp	6	01	1850	1bvp	1	01
1851	1bvp	4	01	1852	2btv	Q	01
1853	2btv	S	01	1854	2btv	D	01
1855	3e3i	A	00	1856	3e2a	B	00
1857	2a8c	E	00	1858	1t75	A	00
1859	3e2a	A	00	1860	3e3i	F	00
1861	3m5h	F	00	1862	2wr7	A	01
1863	1rvx	H	00	1864	2wr5	C	01
1865	3al4	D	00	1866	3gbm	D	00
1867	1l3e	B	00	1868	1r8u	B	00
1869	2kje	A	00	1870	1f81	A	00
1871	2ka4	A	00	1872	2k8f	A	00
1873	2ata	A	00	1874	3aqd	N	00
1875	1j88	E	02	1876	3c6l	H	02
1877	3miq	H	00	1878	3i3b	D	04
1879	1qf1	A	02	1880	3chr	A	03
1881	3fh7	A	03	1882	1trl	A	00
1883	1kto	A	02	1884	7tli	A	02
1885	3ar5	A	04	1886	1xp5	A	04
1887	2eau	A	04	1888	3fgo	A	04
1889	3ar7	A	04	1890	3ar6	A	04
1891	1b65	F	00	1892	3n5i	B	00
1893	3n33	C	00	1894	1vz8	A	01
1895	1b65	A	00	1896	3n33	B	00
1897	1nvb	A	02	1898	1rrm	A	02
1899	3bfj	M	02	1900	1nve	A	02
1901	1oj7	C	02	1902	1nvf	C	02
1903	1re0	B	02	1904	3l8n	A	02
1905	1xsz	B	02	1906	1pbv	A	02
1907	1xt0	B	02	1908	1r8s	E	02
1909	1kuo	A	00	1910	1aq3	B	00
1911	1gav	M	00	1912	1bms	C	00
1913	6msf	C	00	1914	1zdi	A	00

1915	1n7p	A	03	1916	1loh	A	03
1917	1fls	A	03	1918	1xli	A	03
1919	1ojm	A	03	1920	1hmw	A	03
1921	1nqm	A	00	1922	2q2m	A	00
1923	1u18	B	00	1924	3bx8	E	00
1925	1vyo	A	00	1926	1fel	A	00
1927	1ihn	A	00	1928	2cyj	A	00
1929	2fi9	A	00	1930	2k2e	A	01
1931	1ihn	B	00	1932	2fvt	A	01
1933	1ok8	A	03	1934	2hg0	A	03
1935	3p54	A	03	1936	1urz	B	03
1937	1urz	C	03	1938	3c5x	A	03
1939	1n0v	D	02	1940	1zm3	E	02
1941	1zm9	A	02	1942	1zm4	A	02
1943	1zm3	C	02	1944	1u2r	A	03
1945	1n8b	C	01	1946	1n8b	D	01
1947	1n8b	B	01	1948	1n7z	C	01
1949	1n7z	A	01	1950	1n7z	D	01
1951	2i00	D	03	1952	2i00	A	03
1953	3bn8	A	00	1954	2hv2	C	02
1955	3cnu	A	00	1956	2i00	F	03
1957	2abj	D	01	1958	2hgx	A	01
1959	1cko	A	01	1960	1ggo	A	06
1961	1ce8	G	02	1962	1kee	C	06
1963	1nph	A	02	1964	2fh3	C	03
1965	1h1v	G	01	1966	2fh1	B	01
1967	1esv	S	00	1968	1p8x	C	01
1969	1aip	D	01	1970	3lx3	A	00
1971	1aip	G	01	1972	3d7j	F	00
1973	1b66	A	00	1974	3i2b	D	00
1975	1lti	A	00	1976	2a5f	B	00
1977	1ltt	A	00	1978	1htl	A	00
1979	1s5b	A	00	1980	1lta	A	00
1981	1n8w	B	04	1982	2gq3	B	04
1983	1n8w	A	04	1984	1p7t	B	04
1985	1n8i	A	04	1986	2gq3	A	04
1987	2ffd	F	02	1988	1lt9	B	01
1989	1n86	B	01	1990	2hod	E	02
1991	1fzf	F	01	1992	2j0g	F	01
1993	1mur	A	02	1994	1muh	A	02
1995	1b7e	A	01	1996	1mus	A	02
1997	3ecp	A	02	1998	1f3i	A	01
1999	1i7w	C	00	2000	2c1m	A	00
2001	2j dq	A	00	2002	3bct	A	00
2003	2c1t	A	00	2004	1upl	B	00
2005	1yki	C	00	2006	2i7h	D	00
2007	2ifa	A	00	2008	3bem	A	00
2009	3hzn	C	00	2010	3n2s	A	00
2011	1cl4	A	00	2012	1q3y	A	00
2013	1aaf	A	00	2014	1f6u	A	00
2015	1a1t	A	00	2016	2exf	A	00
2017	3m2u	D	03	2018	3pot	B	01
2019	3m30	E	01	2020	1mro	A	03
2021	3m30	D	03	2022	3m2r	E	01
2023	3k38	M	00	2024	1n7d	A	09
2025	1ofz	B	00	2026	3b7e	B	00
2027	2hqs	F	02	2028	3k39	J	00
2029	1hex	A	00	2030	1hex	B	00
2031	2bml	A	00	2032	1gvm	A	00
2033	2j8g	A	02	2034	1h8g	B	00
2035	2x62	A	00	2036	1ro7	D	00
2037	2wqq	A	00	2038	1ro7	C	00
2039	2x61	A	00	2040	1ro8	B	00
2041	3d3c	C	00	2042	3d3c	B	00

2043	3imq	A	00	2044	1tzv	A	00
2045	1tzx	B	00	2046	1tzu	A	00
2047	2pzx	C	01	2048	2pzx	D	01
2049	1fjr	A	01	2050	2pzx	A	01
2051	2pzx	B	01	2052	1fjr	B	01
2053	1qxo	B	00	2054	2g85	A	00
2055	1q11	D	00	2056	1qxo	D	00
2057	1qxo	A	00	2058	2o12	A	00
2059	1puc	A	00	2060	1dkt	B	00
2061	1buh	B	00	2062	1sce	D	00
2063	1dks	B	00	2064	1qb3	C	00
2065	2a99	A	01	2066	1xdy	C	00
2067	3hbp	A	01	2068	1ogp	C	01
2069	1xdy	F	00	2070	1xdy	D	00
2071	2g2w	B	01	2072	3c7u	B	01
2073	2b5r	C	02	2074	3c7v	D	02
2075	2b5r	C	01	2076	2b5r	D	01
2077	2prc	C	02	2078	7prc	C	01
2079	1dxr	C	01	2080	1prc	C	02
2081	2wjm	C	01	2082	3g7f	C	02
2083	3ktu	A	02	2084	3cwa	C	03
2085	1wei	A	01	2086	2nof	A	02
2087	3evs	A	03	2088	3fsq	A	01
2089	1w27	B	03	2090	3eze	A	02
2091	3ezd	A	02	2092	3ezc	A	02
2093	1zym	A	02	2094	1ezb	A	02
2095	2nrf	A	00	2096	2xow	A	00
2097	2xtu	A	00	2098	2nrf	B	00
2099	3b45	A	00	2100	2irv	A	00
2101	1xi8	B	03	2102	1fc5	B	04
2103	1g8r	A	04	2104	2nqk	B	04
2105	2fu3	B	04	2106	2nqr	B	04
2107	2q9y	B	00	2108	2f8z	F	00
2109	1ps1	A	00	2110	3m9u	A	00
2111	1hm7	A	00	2112	1dgp	B	00
2113	2pmb	A	01	2114	3bq9	A	01
2115	3bq9	B	01	2116	3gh1	C	01
2117	3gh1	B	01	2118	3gh1	D	01
2119	3l00	A	01	2120	1yfh	C	01
2121	1eh6	A	01	2122	3kzz	A	01
2123	1t39	B	01	2124	1t39	A	01
2125	1vns	A	02	2126	1vni	A	02
2127	1qhb	E	00	2128	1vnh	A	02
2129	1vnc	A	02	2130	1qhb	D	00
2131	3bkn	H	00	2132	2xjo	G	00
2133	1vlg	D	00	2134	2cf7	E	00
2135	2xkq	C	00	2136	2iu2	A	00
2137	1jly	A	02	2138	1hwm	B	02
2139	2amz	B	01	2140	3jut	C	00
2141	3a21	B	04	2142	3ah1	B	02
2143	2j0g	E	02	2144	2j0y	B	02
2145	1fzd	C	02	2146	1n73	B	02
2147	1jfe	E	02	2148	1n86	C	02
2149	3n99	S	03	2150	3n99	L	03
2151	3dcl	D	03	2152	3n99	J	03
2153	3n99	E	03	2154	3dcl	B	03
2155	1knv	A	00	2156	2pfj	A	00
2157	2bam	B	00	2158	1m0d	C	00
2159	1m0d	B	00	2160	3bam	B	00
2161	1g6q	4	02	2162	1f3l	A	02
2163	1g6q	2	02	2164	1g6q	6	02
2165	1g6q	5	02	2166	2fyt	A	02
2167	2rbh	B	00	2168	2qik	A	02
2169	2rbh	A	00	2170	3juc	A	00

2171	1v30	A	00	2172	2i5t	A	00
2173	1b3s	E	00	2174	1bv0	E	00
2175	1x1y	D	00	2176	3da7	D	00
2177	1b2s	F	00	2178	1x1x	F	00
2179	1eil	A	01	2180	1sp9	A	01
2181	2rk0	B	01	2182	3hnq	A	01
2183	3hq0	C	02	2184	1cix	A	01
2185	2ag2	C	00	2186	1g13	B	00
2187	1tjj	C	00	2188	2ag9	A	00
2189	1pu5	C	00	2190	1tjj	B	00
2191	3ifg	L	01	2192	2qe0	C	01
2193	2vle	A	01	2194	1o05	B	01
2195	3ifg	M	01	2196	1wnd	D	01
2197	3eq5	J	00	2198	3eq5	L	00
2199	118r	B	00	2200	113g	A	00
2201	118r	A	00	2202	3eq5	G	00
2203	1az3	B	00	2204	2rve	A	00
2205	1tx3	A	00	2206	1kc6	B	00
2207	1bsu	A	00	2208	1eo3	A	00
2209	1w2m	D	01	2210	3biq	A	01
2211	1a16	A	01	2212	3cb5	A	01
2213	2bwv	A	01	2214	1m35	E	01
2215	1w4k	A	00	2216	1w4e	A	00
2217	1w88	J	00	2218	1w88	I	00
2219	2eq9	C	00	2220	1bbl	A	00
2221	1pp6	C	00	2222	2rci	A	00
2223	1pp0	D	00	2224	1pp6	D	00
2225	1vgf	B	00	2226	1pp0	A	00
2227	2pvi	B	00	2228	1eyu	A	00
2229	1pvu	B	00	2230	1h56	A	00
2231	3pvi	A	00	2232	3pvi	B	00
2233	3lup	A	02	2234	3pnq	D	02
2235	1vpv	B	02	2236	1pzx	A	03
2237	3fdj	A	03	2238	1oi2	A	02
2239	1ldj	A	03	2240	3rtr	E	01
2241	1ldk	A	01	2242	2wzk	A	02
2243	1ldk	A	03	2244	1u6g	A	03
2245	2drd	B	01	2246	2hqd	A	01
2247	2hrt	A	01	2248	2v50	A	01
2249	3noc	B	01	2250	1oy6	A	01
2251	3etg	A	02	2252	1gtm	B	01
2253	3mw9	F	02	2254	2hk8	B	01
2255	1dib	B	02	2256	1nr7	K	01
2257	1smc	C	00	2258	3ca9	B	00
2259	3hhq	I	00	2260	2oke	C	00
2261	1xs4	E	00	2262	3ecy	B	00
2263	2wgm	B	01	2264	1wu0	A	00
2265	2xqt	D	01	2266	2wgm	e	01
2267	1yce	J	01	2268	1yce	M	01
2269	1eyq	A	02	2270	1jep	A	02
2271	1jx1	B	02	2272	1fm7	B	02
2273	1fm8	B	02	2274	1eyp	B	02
2275	2qvz	X	03	2276	2qvy	X	03
2277	1amu	B	03	2278	3a9u	A	03
2279	2qw0	X	03	2280	1mdf	A	03
2281	2qvj	B	01	2282	3pto	D	01
2283	2gic	A	01	2284	3hhz	L	01
2285	3pu1	E	01	2286	3hhw	M	01
2287	3cdd	D	03	2288	3cdd	A	03
2289	3cdd	F	03	2290	1wru	A	03
2291	3cdd	B	03	2292	3cdd	C	03
2293	3sr2	D	01	2294	2qm4	B	01
2295	3ii6	A	01	2296	2qm4	A	01
2297	3ii6	C	01	2298	3sr2	G	01

2299	lit7	A	02	2300	liq8	B	02
2301	liq8	A	02	2302	lit7	B	02
2303	lit8	A	02	2304	lj2b	B	02
2305	3bum	B	01	2306	3buo	B	01
2307	3ob2	B	01	2308	1de4	F	03
2309	1cx8	G	03	2310	1fbv	A	04
2311	3dkg	A	01	2312	1g3n	A	01
2313	2wqo	A	01	2314	3dkc	A	01
2315	ljbp	E	02	2316	3ad4	A	01
2317	lnkl	A	00	2318	1sn6	A	01
2319	lof9	A	00	2320	1n69	B	00
2321	2qyp	A	00	2322	2z9a	B	00
2323	li6b	A	02	2324	1quj	A	02
2325	loqh	A	02	2326	1dk4	B	02
2327	1bka	A	03	2328	1ftk	A	01
2329	2jdw	A	00	2330	1yni	D	00
2331	2jer	D	00	2332	2ewo	H	00
2333	1yni	C	00	2334	2ewo	B	00
2335	ljxo	B	03	2336	1ex6	B	01
2337	3ney	C	02	2338	1jxm	A	03
2339	1s96	A	02	2340	2zzz	B	02
2341	3rbb	A	00	2342	1avv	A	00
2343	3ioz	A	00	2344	1avz	A	00
2345	3reb	A	00	2346	3rea	A	00
2347	1hev	A	00	2348	2uwz	A	04
2349	2sgf	I	00	2350	1ulk	B	01
2351	1ct0	I	00	2352	1wgc	B	01
2353	2x2r	C	00	2354	2zfl	A	00
2355	2fme	A	00	2356	2owm	A	00
2357	3hqd	B	00	2358	3h4s	A	01
2359	lvzy	A	01	2360	li7f	A	01
2361	lvzy	B	01	2362	lvq0	A	01
2363	1hw7	A	01	2364	lvq0	B	01
2365	3d5k	A	02	2366	1wp1	A	02
2367	1yc9	A	02	2368	3pik	A	02
2369	3d5k	B	02	2370	3d5k	C	02
2371	1xlr	A	00	2372	2ooi	B	01
2373	2fa1	B	00	2374	2ahc	C	00
2375	3eet	B	02	2376	1glb	B	00
2377	1fwc	B	00	2378	1fwa	B	00
2379	1e9z	A	01	2380	1kra	B	00
2381	lubp	B	00	2382	3kau	B	00
2383	1mbr	0	01	2384	3i99	A	03
2385	2mbr	A	01	2386	luxy	A	01
2387	1mbt	A	01	2388	2q85	A	03
2389	1q27	A	00	2390	1pvf	B	00
2391	1mut	A	00	2392	1k2e	B	00
2393	2gt2	B	00	2394	1v8r	A	00
2395	1rkm	A	02	2396	1b40	A	02
2397	1b3g	A	02	2398	1b3f	A	02
2399	1b6h	A	02	2400	1dpp	E	01
2401	1a73	B	00	2402	1cz0	A	00
2403	2o6m	A	00	2404	1evw	B	00
2405	1evw	A	00	2406	1cyq	A	00
2407	1x6m	C	00	2408	1x6m	B	00
2409	1x6m	A	00	2410	1xa8	D	00
2411	1xa8	B	00	2412	1xa8	C	00
2413	2op3	A	00	2414	1yt7	A	00
2415	1u9x	A	00	2416	3bwk	B	00
2417	3mpf	A	00	2418	3ovx	A	00
2419	1cfa	A	00	2420	3md2	C	01
2421	1aa7	B	01	2422	1kjs	A	00
2423	1aa7	A	01	2424	3md2	A	01
2425	2h4c	D	00	2426	2azy	A	00

2427	2qu9	A	00	2428	1jq9	A	00
2429	2not	B	00	2430	1poa	A	00
2431	2jio	A	04	2432	1i50	A	03
2433	3eng	A	00	2434	1fdo	A	04
2435	1g8k	E	04	2436	3cf2	D	01
2437	1n6d	C	03	2438	1n6f	A	03
2439	1hf2	D	01	2440	1thn	D	00
2441	1h4y	B	00	2442	1v0f	C	01
2443	1t84	A	00	2444	1ees	B	00
2445	1cee	B	00	2446	1ej5	A	00
2447	1f3m	A	00	2448	1e0a	B	00
2449	2z0t	C	00	2450	2z0t	D	00
2451	1xne	A	00	2452	2z0t	A	00
2453	1s04	A	00	2454	2z0t	B	00
2455	1s0v	A	01	2456	1s0v	D	01
2457	2pi5	A	01	2458	1h38	C	01
2459	1h38	B	01	2460	2pi4	A	01
2461	2p0n	A	00	2462	3de8	B	00
2463	1rke	B	00	2464	3ipm	O	01
2465	1c02	A	00	2466	1fnt	m	01
2467	1nrz	D	00	2468	2jzn	C	00
2469	1nrz	C	00	2470	2jzh	A	00
2471	1vsq	C	00	2472	1ble	A	00
2473	1u5w	D	00	2474	1u5w	H	00
2475	2mjp	B	00	2476	1ex2	A	00
2477	2pyu	A	00	2478	1v7r	A	00
2479	1l8q	A	03	2480	3pvp	A	00
2481	2hcb	D	03	2482	2hcb	C	03
2483	3pvv	A	00	2484	3pvp	B	00
2485	1ltz	A	00	2486	2xsn	B	00
2487	1dmw	A	00	2488	2xsn	D	00
2489	1j8u	A	00	2490	1lrm	A	00
2491	2ier	B	02	2492	1p42	A	02
2493	2ies	B	02	2494	1yhc	A	02
2495	1yhc	B	02	2496	1yh8	A	02
2497	2a6h	C	03	2498	2a69	C	03
2499	3eql	C	03	2500	1smy	M	03
2501	1twf	B	04	2502	1y1v	B	04
2503	2chs	A	00	2504	1com	G	00
2505	1com	K	00	2506	2uzh	A	00
2507	1h48	B	00	2508	2lbw	A	00
2509	1utg	A	00	2510	2utg	B	00
2511	1ccd	A	00	2512	2utg	A	00
2513	1utr	A	00	2514	1utr	B	00
2515	3be1	A	03	2516	1mox	A	01
2517	2hr7	A	03	2518	1s78	B	01
2519	2hr7	B	03	2520	2hr7	A	01
2521	3crj	B	00	2522	2np3	A	00
2523	1jum	B	02	2524	3bjb	E	00
2525	2zoy	B	00	2526	1rkw	E	02
2527	2ota	B	00	2528	2juz	B	00
2529	2jr2	A	00	2530	2jrx	A	00
2531	2juz	A	00	2532	2ota	A	00
2533	1cgp	B	01	2534	1l7q	A	02
2535	1p8j	D	02	2536	1yi7	B	02
2537	2w68	C	00	2538	1axz	A	00
2539	1d6u	A	01	2540	2w0q	A	01
2541	1qaf	A	01	2542	1spu	A	01
2543	1qal	B	01	2544	1d6y	A	01
2545	2p1x	A	00	2546	3s4m	A	00
2547	1xaq	A	01	2548	1dlx	A	00
2549	1ly7	A	00	2550	1whz	A	00
2551	1byt	A	02	2552	2iuj	A	02
2553	1n8q	A	02	2554	1fgo	A	02

2555	2sbl	B	02	2556	1lnh	A	02
2557	2ipr	A	00	2558	3qfq	A	00
2559	2ntc	A	00	2560	3qfq	B	00
2561	1tbd	A	00	2562	1ksy	A	00
2563	2sbl	B	03	2564	1fgm	A	03
2565	1rrl	A	03	2566	2sbl	A	03
2567	2iuk	B	03	2568	1hu9	A	03
2569	3knk	Z	01	2570	3i8f	V	01
2571	1nxz	B	01	2572	2xg2	Z	01
2573	2re8	A	04	2574	2qbi	V	00
2575	1sj8	A	02	2576	3b55	A	03
2577	2rad	B	03	2578	2qzg	A	01
2579	2qzg	C	01	2580	1jqo	A	03
2581	1vq9	L	02	2582	3ccl	L	02
2583	1yj9	O	00	2584	2qex	L	02
2585	1k8a	M	02	2586	1yhq	L	02
2587	2ffs	A	00	2588	1xn5	A	00
2589	2pcs	A	00	2590	2kte	A	00
2591	3ie5	A	00	2592	1h2o	A	00
2593	2p2e	A	00	2594	1nwb	A	00
2595	1r94	A	00	2596	1s98	B	00
2597	1r95	B	00	2598	2qgo	A	00
2599	1k5q	A	01	2600	3ml0	A	01
2601	1e3a	A	01	2602	2adv	A	00
2603	1pnm	A	01	2604	1gm7	A	01
2605	1tbc	A	00	2606	1tiv	A	00
2607	1tac	A	00	2608	1jfw	A	00
2609	1k5k	A	00	2610	1fku	A	00
2611	1yox	B	00	2612	1yox	D	00
2613	1yox	A	00	2614	1yox	F	00
2615	1yox	E	00	2616	1yox	C	00
2617	2ixu	A	03	2618	1oba	A	03
2619	1h09	A	03	2620	2j8g	A	03
2621	2ixv	A	03	2622	2j8f	A	03
2623	1ja1	B	03	2624	1j9z	B	03
2625	1f20	A	02	2626	1amo	B	03
2627	1tll	A	02	2628	1j9z	A	03
2629	1ai6	B	02	2630	1fxv	B	02
2631	1gk1	B	03	2632	1k5s	B	02
2633	3ml0	B	02	2634	1jvz	B	03
2635	1a36	A	04	2636	1ej9	A	04
2637	1lpq	A	04	2638	1ge9	A	01
2639	1ise	A	01	2640	3r8n	Y	01
2641	2pvg	A	00	2642	2pu9	A	00
2643	2puk	E	00	2644	2pvd	A	00
2645	2puo	A	00	2646	2puk	A	00
2647	1wpl	S	00	2648	1wpl	Q	00
2649	1jg5	B	00	2650	1wpl	R	00
2651	1is7	N	00	2652	1is7	O	00
2653	2qbd	C	02	2654	3ilz	C	02
2655	2v48	C	02	2656	2wdg	C	02
2657	2ow8	d	02	2658	3oi2	C	02
2659	2crk	A	01	2660	1qk1	E	01
2661	2gl6	C	01	2662	3ju5	C	01
2663	3l2g	N	01	2664	1crk	B	01
2665	2ols	A	03	2666	1vbg	A	05
2667	1kee	D	01	2668	1c30	F	01
2669	1ce8	D	01	2670	1m6v	H	01
2671	3l88	J	00	2672	3qnd	A	00
2673	1nob	B	00	2674	3l89	D	00
2675	1uxe	A	00	2676	1nob	C	00
2677	1a5u	C	01	2678	1aqf	D	01
2679	2vgi	D	01	2680	2jj1	N	02
2681	1f3x	A	01	2682	2vgg	C	01

2683	2pda	A	07	2684	2c3m	B	07
2685	2c3p	A	07	2686	1kek	B	07
2687	2c3u	B	07	2688	2c3y	A	07
2689	3kkj	B	02	2690	1h83	B	02
2691	2xfq	B	02	2692	2c75	A	02
2693	3bi2	A	02	2694	1rsg	A	02
2695	1oan	B	01	2696	1rer	A	01
2697	1urz	C	01	2698	1urz	E	01
2699	2ala	A	01	2700	1rer	C	01
2701	1wc9	A	01	2702	1r44	E	00
2703	2pwn	A	00	2704	1u10	E	00
2705	1vhh	A	00	2706	1lbu	A	02
2707	2qg7	B	02	2708	3c5i	B	02
2709	2b0q	A	02	2710	3c5i	C	02
2711	2qg7	D	02	2712	1l8t	A	02
2713	1ar3	0	02	2714	1b0m	A	02
2715	1b0j	A	02	2716	1as9	0	02
2717	1l5j	B	03	2718	8acn	A	02
2719	3k40	A	01	2720	1js6	B	01
2721	1js6	A	01	2722	3k40	B	01
2723	1js3	B	01	2724	1js3	A	01
2725	2wzs	E	02	2726	2wzs	F	02
2727	2wvz	B	02	2728	2ww0	A	02
2729	2wzs	A	02	2730	2wzs	D	02
2731	1pth	A	02	2732	1prh	B	02
2733	1eqh	B	02	2734	1dnw	D	00
2735	3n8x	B	02	2736	1diy	A	02
2737	2vbl	B	00	2738	2i3p	A	00
2739	2vbn	B	00	2740	1dfa	A	03
2741	1dfa	A	02	2742	3mxb	S	00
2743	3b8n	E	01	2744	3b8m	A	01
2745	3b8n	B	01	2746	3b8n	F	01
2747	3b8o	D	01	2748	3b8o	A	01
2749	1rvz	G	01	2750	3fku	C	01
2751	3al4	C	01	2752	3m6s	K	01
2753	3htt	A	01	2754	3lyj	I	01
2755	3qkj	D	02	2756	1e7d	B	02
2757	1khc	A	02	2758	1en7	B	02
2759	1h9f	A	00	2760	1jeq	A	05
2761	3mkp	B	01	2762	1nk1	B	01
2763	1i8n	A	00	2764	2qj2	A	01
2765	1gmo	C	01	2766	3hmt	B	00
2767	1e5g	A	01	2768	1qo6	A	01
2769	2rkz	E	01	2770	1g44	C	01
2771	3l89	T	02	2772	2jra	B	01
2773	1y2x	D	00	2774	1tzq	A	00
2775	3lim	E	00	2776	1kd6	A	00
2777	1y2t	A	00	2778	1o7l	B	00
2779	1r4q	A	02	2780	1oql	A	02
2781	1obt	A	02	2782	1mrh	A	02
2783	1qci	A	02	2784	3njs	A	02
2785	1c4z	A	02	2786	1d5f	A	02
2787	1c4z	C	02	2788	1d5f	C	02
2789	1d5f	B	02	2790	1c4z	B	02
2791	1eqn	C	02	2792	2i5r	B	00
2793	1eqn	A	02	2794	2zbk	G	02
2795	2fcj	A	00	2796	2i5r	A	00
2797	3mof	A	03	2798	3p0p	A	00
2799	2paw	A	02	2800	3dt2	A	03
2801	2fah	A	03	2802	2pc9	A	03
2803	2qsd	F	02	2804	2qsd	C	02
2805	2qsd	E	02	2806	2qsd	A	02
2807	2qsd	H	02	2808	2qsd	B	02
2809	2uzq	A	00	2810	3p3a	B	02

2811	1tq1	A	00	2812	1uar	A	02
2813	2uzq	C	00	2814	3aax	B	02
2815	1olm	E	01	2816	3b7q	A	02
2817	3b7n	A	02	2818	3b74	A	02
2819	3b7q	B	02	2820	1olm	A	01
2821	3asn	K	00	2822	3abk	X	00
2823	3abm	X	00	2824	1v54	K	00
2825	3ag4	X	00	2826	2eik	X	00
2827	2ih4	A	02	2828	2np7	A	02
2829	1adm	A	02	2830	2ibt	D	02
2831	1aqi	A	02	2832	1g38	A	02
2833	2yb9	A	02	2834	1y8j	A	02
2835	1rlj	A	02	2836	1dmt	A	02
2837	1rli	A	02	2838	1r1h	A	02
2839	1q51	L	01	2840	3qxz	C	01
2841	2fzs	N	00	2842	2zl2	E	00
2843	3ktk	B	00	2844	2a7s	C	01
2845	2wy6	A	01	2846	2ffz	A	00
2847	2wy6	B	01	2848	1qmd	B	01
2849	1qmd	A	01	2850	1p5x	A	00
2851	1m61	A	02	2852	2ozo	A	02
2853	1a81	A	02	2854	1a81	I	02
2855	2oq1	A	02	2856	1a81	C	02
2857	1nf2	B	02	2858	1nf2	C	02
2859	2hf2	B	02	2860	1nf2	A	02
2861	1rlo	C	02	2862	1rkq	A	02
2863	1nh3	A	02	2864	1seu	A	01
2865	1a35	A	02	2866	1ois	A	02
2867	1k4s	A	01	2868	1k4t	A	01
2869	3myr	A	02	2870	1wuk	S	02
2871	2frv	A	02	2872	1ubt	S	02
2873	1ubh	S	02	2874	1yrq	A	02
2875	1w3a	A	02	2876	1z52	A	03
2877	1w3g	A	02	2878	3c0m	A	03
2879	3c0m	B	03	2880	3c0o	A	03
2881	1lkt	E	00	2882	1lkt	B	00
2883	1lkt	C	00	2884	1lkt	D	00
2885	1lkt	A	00	2886	1lkt	F	00
2887	1k0e	B	00	2888	1i7s	A	00
2889	1qdl	A	00	2890	1k0e	A	00
2891	1k0g	A	00	2892	1i7q	A	00
2893	2r4l	B	00	2894	8prn	A	00
2895	3bz1	O	01	2896	2r89	A	00
2897	3prq	O	01	2898	1opf	E	00
2899	3skk	B	00	2900	3sfh	A	00
2901	2i4y	C	00	2902	1tal	A	00
2903	2aeb	A	00	2904	1woi	A	00
2905	1j5q	B	02	2906	1dhx	A	04
2907	1hqn	B	01	2908	1m3y	B	01
2909	1j5q	A	02	2910	1m3y	A	01
2911	1x3x	B	00	2912	1b5a	A	00
2913	1lco	A	01	2914	1cyo	A	00
2915	1icc	B	00	2916	3ks0	A	00
2917	3kz4	H	01	2918	3n09	I	01
2919	3kz4	N	01	2920	3gzu	F	01
2921	3gzu	C	01	2922	3gzu	K	01
2923	1c8u	B	01	2924	1vi8	G	00
2925	1pn4	B	01	2926	2qwz	B	01
2927	1s9c	G	02	2928	3b7j	D	00
2929	2jiz	K	03	2930	3oe7	M	03
2931	3oeh	D	03	2932	2w6j	F	03
2933	2jdi	D	03	2934	1e1r	F	03
2935	2ikb	C	00	2936	2nr7	A	00
2937	1chk	B	01	2938	2is5	C	00

2939	2ikb	B	00	2940	2is5	A	00
2941	1wro	A	00	2942	1wrq	B	00
2943	1wro	C	00	2944	1wpt	A	00
2945	2zh0	H	00	2946	2gzt	B	00
2947	1mjg	A	00	2948	1mj1	B	00
2949	1mjp	B	00	2950	1cmb	A	00
2951	1mjg	D	00	2952	1mjg	J	00
2953	5cro	B	00	2954	3c6f	C	01
2955	6cro	A	00	2956	1d11	A	00
2957	2orc	A	00	2958	2a63	A	00
2959	3glh	B	02	2960	1ht2	H	03
2961	1bvs	D	03	2962	1g3i	U	03
2963	2wrk	M	01	2964	2qow	M	01
2965	2bed	B	00	2966	3ksl	B	00
2967	1o6h	A	02	2968	1h35	C	01
2969	3d5s	B	00	2970	1qaz	A	00
2971	1uj1	B	03	2972	2alv	A	03
2973	2amp	B	03	2974	1uk4	A	03
2975	2gz9	A	03	2976	2gz7	A	03
2977	1i21	Y	00	2978	2q4y	A	00
2979	1nsl	D	00	2980	2pdo	F	01
2981	1gyt	G	02	2982	2glj	G	01
2983	1xu8	A	01	2984	1dvm	B	01
2985	2ach	A	02	2986	1d5s	A	02
2987	1lj5	A	01	2988	7api	A	02
2989	3brm	B	02	2990	1mki	B	02
2991	2pby	D	02	2992	2pby	B	02
2993	2pby	A	02	2994	2osu	A	02
2995	2glj	P	02	2996	2glj	W	02
2997	2ijz	K	02	2998	2glj	S	02
2999	2glj	J	02	3000	2glf	C	02
3001	1rj2	A	01	3002	2rgn	B	01
3003	1rj2	J	01	3004	1x86	G	01
3005	1xd4	B	01	3006	2nz8	B	01
3007	1tkv	A	00	3008	1jr5	A	00
3009	1ka3	B	00	3010	1tl6	A	00
3011	1jr5	B	00	3012	1tlh	A	00
3013	2qb9	S	00	3014	2zm6	S	00
3015	3ora	S	00	3016	3huw	S	00
3017	2vho	S	00	3018	2wdg	S	00
3019	1fs1	B	00	3020	2p1q	A	00
3021	1fqv	J	00	3022	3e4u	C	00
3023	2r9r	H	01	3024	3ogk	E	00
3025	1ijj	A	02	3026	1t44	A	02
3027	1j6z	A	02	3028	1atn	A	02
3029	1sqk	A	02	3030	1mdu	B	02
3031	2c62	B	00	3032	2it9	D	00
3033	1l3a	A	00	3034	1pcf	D	00
3035	2phe	B	00	3036	1pcf	F	00
3037	3er0	A	01	3038	1xni	A	02
3039	1h64	I	00	3040	3ku8	C	00
3041	2fg0	B	01	3042	3bw1	B	00
3043	1we3	P	00	3044	1aon	U	00
3045	1pcq	R	00	3046	1svt	P	00
3047	1sx4	S	00	3048	1wf4	t	00
3049	1gho	N	03	3050	3s6o	B	00
3051	1djg	A	02	3052	1pdy	A	02
3053	3aiw	A	00	3054	1twz	A	00
3055	1xx7	D	00	3056	2ogi	A	00
3057	2q14	F	01	3058	1xx7	C	00
3059	3bg2	A	01	3060	3ccg	A	00
3061	1lq8	A	01	3062	1yxa	B	02
3063	3f5n	B	01	3064	3cvm	B	02
3065	3dlw	A	02	3066	2b4x	I	02

3067	Intc	A	00	3068	1a5j	A	02
3069	1bl0	A	01	3070	3bdd	C	00
3071	2dtz	D	01	3072	1jt0	B	01
3073	1wpb	D	02	3074	1wpb	E	02
3075	1wpb	A	02	3076	1wpb	N	02
3077	1wpb	M	02	3078	1wpb	F	02
3079	2few	A	00	3080	1xiz	A	00
3081	3lf6	A	00	3082	1a3a	A	00
3083	1a3a	C	00	3084	1a6j	B	00
3085	2v19	F	00	3086	2v18	G	00
3087	2ux9	A	00	3088	2v21	B	00
3089	2v18	H	00	3090	2cif	A	00
3091	1qte	A	02	3092	1bmt	B	01
3093	3bul	A	01	3094	1sv1	B	00
3095	2j70	A	00	3096	1q6a	B	00
3097	2zd2	B	02	3098	2r31	A	02
3099	2p4x	B	02	3100	2p4x	A	02
3101	2zd2	A	02	3102	2r6i	B	02
3103	1x6r	A	00	3104	2hi2	A	00
3105	1qve	B	00	3106	1hpw	A	00
3107	1qve	A	00	3108	1wsc	A	01
3109	1n8w	B	02	3110	1y8b	A	03
3111	1d8c	A	02	3112	1p7t	B	02
3113	1p7t	A	02	3114	1n8w	A	02
3115	2c0r	A	02	3116	2tpl	A	01
3117	2ay8	A	01	3118	1i48	K	02
3119	1kaf	F	00	3120	3kgw	B	01
3121	2hbw	A	02	3122	2fg0	A	02
3123	2fg0	B	02	3124	2jyx	A	00
3125	2k1g	A	00	3126	2evr	A	02
3127	1q48	A	00	3128	1r9p	A	00
3129	3g0m	A	00	3130	1ni7	A	00
3131	1mzg	A	00	3132	1mzg	B	00
3133	1zvg	A	00	3134	1bcg	A	00
3135	2bup	A	02	3136	1wt7	A	00
3137	1c56	A	00	3138	1fh3	A	00
3139	2fyv	A	01	3140	3bvu	A	01
3141	2f7r	A	01	3142	1o7d	A	01
3143	3dx0	A	01	3144	2alw	A	01
3145	1koj	A	03	3146	2o2c	C	03
3147	3hjb	C	03	3148	2o2c	A	03
3149	2cxr	A	03	3150	1q50	A	03
3151	2wh2	I	02	3152	3i8i	K	02
3153	3hux	I	02	3154	3i8f	K	02
3155	3d5b	I	02	3156	3ohk	I	02
3157	1n8w	B	01	3158	1p7t	B	01
3159	1n8i	A	01	3160	1d8c	A	01
3161	1p7t	A	01	3162	2gq3	B	01
3163	2iex	A	02	3164	2dub	A	02
3165	1rjn	B	02	3166	1mj3	D	02
3167	2dub	B	02	3168	2zqr	C	02
3169	3g5p	B	00	3170	1ws0	A	00
3171	1qly	A	00	3172	1rn5	A	00
3173	1g27	A	00	3174	1g27	C	00
3175	1zvi	A	03	3176	1ed4	B	03
3177	1dmj	A	03	3178	1ed6	A	03
3179	1om4	A	03	3180	1p6k	B	03
3181	1vq4	W	02	3182	1nji	X	02
3183	3g4s	W	02	3184	3cxc	V	02
3185	3cej	W	02	3186	2otl	W	02
3187	2r1a	E	00	3188	2r1a	B	00
3189	2r1a	H	00	3190	2r1a	G	00
3191	2r1a	C	00	3192	2r1a	D	00
3193	1xs0	A	00	3194	1hke	A	00

3195	1xs0	C	00	3196	1hke	B	00
3197	1gpq	A	00	3198	1uuz	A	00
3199	3kqn	A	03	3200	2zjo	A	03
3201	3kqu	E	03	3202	3o8c	A	05
3203	1hei	B	03	3204	2f55	B	03
3205	2pzx	B	02	3206	1fjr	A	02
3207	2pzx	C	02	3208	2pzx	D	02
3209	1fjr	B	02	3210	2pzx	A	02
3211	1oqj	A	00	3212	1h5p	A	00
3213	1oqj	B	00	3214	1mr1	D	00
3215	1mr1	C	00	3216	1ufn	A	00
3217	3glf	E	03	3218	3gli	I	03
3219	3glh	C	03	3220	3glg	H	03
3221	2qw6	D	00	3222	2qw6	C	00
3223	1zs7	A	02	3224	1j2b	B	04
3225	3d79	A	02	3226	1ze2	A	02
3227	1it7	A	04	3228	1sqw	A	02
3229	3jtr	B	03	3230	1kec	B	03
3231	1jvz	B	02	3232	1h2g	B	03
3233	1fm2	B	02	3234	2hzz	A	04
3235	3qsv	A	00	3236	3kmp	A	00
3237	1mhd	B	00	3238	3qsv	C	00
3239	1ozj	B	00	3240	3qsv	B	00
3241	2pys	A	00	3242	2ezn	A	00
3243	3gxz	A	00	3244	1m5j	A	00
3245	1m5m	A	00	3246	115e	A	00
3247	1uzi	A	00	3248	1gze	C	00
3249	2wn6	A	01	3250	1g24	A	00
3251	2a78	B	00	3252	1qs1	D	01
3253	1adt	A	02	3254	2wb0	X	02
3255	1adu	A	02	3256	2waz	X	02
3257	1adu	B	02	3258	1adv	B	02
3259	1u2v	E	00	3260	2p9n	E	00
3261	2p9p	E	00	3262	2p9l	E	00
3263	1tyq	E	00	3264	3rse	E	00
3265	2oc7	C	02	3266	1vcw	A	01
3267	2a4r	C	01	3268	3gct	G	00
3269	3lh3	D	02	3270	1yyy	I	02
3271	3gjn	C	00	3272	1v13	A	00
3273	1m08	A	00	3274	2wpt	B	00
3275	2gzg	B	00	3276	2gzj	F	00
3277	1ff3	B	00	3278	1ff3	C	00
3279	1fva	B	00	3280	2iem	A	00
3281	1nwa	A	00	3282	1fva	A	00
3283	2d29	A	01	3284	1sir	A	01
3285	2a1t	C	01	3286	3swo	C	01
3287	1ukw	B	01	3288	2rfq	D	01
3289	1xnr	N	00	3290	2y10	N	00
3291	3knh	N	00	3292	2wrn	N	00
3293	3kis	n	00	3294	3huw	N	00
3295	1q5f	A	01	3296	3hrv	A	00
3297	3hrv	B	00	3298	1oqv	C	00
3299	1oqv	A	00	3300	1oqv	B	00
3301	1qzx	A	03	3302	1zvz	B	01
3303	3bd1	A	00	3304	2ivq	G	01
3305	2ivg	F	01	3306	2iv1	C	01
3307	1sph	B	00	3308	1opd	A	00
3309	1zer	A	00	3310	2hpr	A	00
3311	1k1c	A	00	3312	1fu0	B	00
3313	1x2g	A	01	3314	2qhv	A	01
3315	1x2h	C	01	3316	2qhs	A	01
3317	1x2g	B	01	3318	2qht	A	01
3319	1ho5	A	02	3320	1oie	A	02
3321	2ush	B	02	3322	1hpu	B	02

3323	lho5	B	02	3324	lhpu	A	02
3325	lthw	A	00	3326	llox	A	00
3327	2pe7	A	00	3328	2i0w	A	00
3329	lpev	B	00	3330	2vu7	A	00
3331	2psp	A	02	3332	1e9t	A	00
3333	lhi7	B	00	3334	lpsp	B	02
3335	2psp	B	02	3336	lpsp	B	01
3337	lp0y	A	02	3338	lmlv	B	02
3339	lozv	A	02	3340	2h21	C	02
3341	2h2j	C	02	3342	2h21	B	02
3343	lnyo	A	00	3344	1w7d	A	00
3345	1x3b	A	00	3346	1w7e	A	00
3347	1o70	A	02	3348	2vxp	B	00
3349	2cx9	D	02	3350	1u8v	B	02
3351	2dvl	B	02	3352	1t9g	C	02
3353	3nf4	B	02	3354	2cx9	B	02
3355	2c1d	A	02	3356	1fj0	B	00
3357	2c1d	C	01	3358	1mg2	D	00
3359	3hq8	A	01	3360	1b7c	A	00
3361	ly2g	A	00	3362	1f46	B	00
3363	1s1s	B	00	3364	1f7x	A	00
3365	1s1j	A	00	3366	1f7w	A	00
3367	lrke	A	01	3368	1u6h	A	01
3369	1u89	A	00	3370	2ibf	A	01
3371	2gww	A	01	3372	1zw2	A	01
3373	lnzb	B	01	3374	lpvp	B	01
3375	1xns	B	02	3376	2crx	B	01
3377	1f44	A	01	3378	1p7d	A	02
3379	lend	0	00	3380	1enk	A	00
3381	2fcc	B	00	3382	1eni	A	00
3383	1enj	A	00	3384	2end	A	00
3385	2wdo	A	00	3386	3g4s	R	00
3387	1f7t	C	00	3388	2v47	W	00
3389	1fte	B	00	3390	1f80	C	00
3391	3qvr	A	03	3392	3qvp	A	03
3393	lgal	A	03	3394	1cf3	A	03
3395	lgpe	A	03	3396	lgpe	B	03
3397	3hek	A	00	3398	1thn	A	00
3399	2bt0	A	00	3400	2vw5	D	00
3401	luy8	A	00	3402	2vw5	B	00
3403	lijy	A	00	3404	lijx	F	00
3405	lijx	E	00	3406	lijy	B	00
3407	lijx	D	00	3408	lijx	A	00
3409	1c0w	B	02	3410	1u8r	C	02
3411	1fx7	A	02	3412	1fwz	A	02
3413	2f5f	A	02	3414	1c0w	D	02
3415	3bgt	D	01	3416	3cmb	D	00
3417	3bh3	B	01	3418	3bh3	C	01
3419	3bh2	B	01	3420	3bh2	A	01
3421	1xd6	A	00	3422	1bwu	A	00
3423	ldlp	F	02	3424	3mez	C	00
3425	3mez	A	00	3426	3a0c	C	00
3427	2ii0	A	01	3428	1nvx	S	01
3429	1xd4	A	03	3430	1nvu	S	01
3431	1xd4	B	03	3432	1xd2	C	01
3433	1mfl	A	01	3434	1kkb	A	01
3435	lson	A	01	3436	1mez	A	01
3437	lade	B	01	3438	1kkf	A	01
3439	lrle	A	02	3440	119m	B	02
3441	lnug	B	02	3442	1nud	A	02
3443	levu	A	02	3444	lggy	A	02
3445	1m1k	V	00	3446	3g4s	U	00
3447	2qez	F	01	3448	3cme	U	00
3449	2qov	Z	00	3450	2qbe	Z	00

3451	1khf	A	01	3452	2zci	A	01
3453	3dtb	A	01	3454	1ii2	B	01
3455	2pe9	C	01	3456	3mof	A	01
3457	3nos	B	01	3458	1r35	A	01
3459	1dm7	A	01	3460	1nsi	B	01
3461	1foo	B	01	3462	1m8e	B	01
3463	1xkh	B	01	3464	1nqh	A	01
3465	2grx	B	01	3466	1po0	A	01
3467	1po3	B	01	3468	1nqe	A	01
3469	3cc4	D	00	3470	2ogk	D	00
3471	2j03	G	00	3472	1k9m	F	00
3473	2pzz	D	00	3474	1qvf	D	00
3475	1ovn	A	02	3476	1g7d	A	00
3477	2c1y	A	02	3478	2c0g	A	02
3479	2c0f	A	02	3480	2c0g	B	02
3481	1gc5	A	02	3482	1dh2	A	02
3483	2i6a	A	02	3484	1dh1	A	02
3485	2abs	A	02	3486	3drw	B	02
3487	1omw	A	02	3488	1fqk	B	02
3489	2bt2	C	02	3490	3cik	A	02
3491	2crp	A	02	3492	1agr	E	02
3493	1tdt	A	01	3494	1kgt	A	01
3495	2tdt	A	01	3496	3eg4	A	01
3497	3gos	A	01	3498	1kgq	A	01
3499	1gqi	A	03	3500	1k9d	A	03
3501	1mqr	A	03	3502	1gqj	B	03
3503	1h4l	A	03	3504	1gql	A	03
3505	3d4y	A	02	3506	3i04	D	01
3507	3i0l	D	01	3508	1gnl	A	02
3509	1u00	A	02	3510	1e2u	A	02
3511	3ir7	C	01	3512	1y4z	C	01
3513	1y5n	C	01	3514	2bs2	F	00
3515	1qla	C	00	3516	1qlb	C	00
3517	1fbl	A	02	3518	1rtg	A	00
3519	1su3	A	02	3520	2clt	B	02
3521	3c7x	A	00	3522	1itv	B	00
3523	1q0a	A	00	3524	1pm6	A	00
3525	2ief	A	00	3526	3q5s	A	01
3527	1exi	A	02	3528	1q09	A	00
3529	1r2a	A	00	3530	1l6e	B	00
3531	2hwn	D	00	3532	2h9r	B	00
3533	2h9r	A	00	3534	2hwn	C	00
3535	2fvh	B	00	3536	1a5n	A	00
3537	1krc	A	00	3538	2fvh	C	00
3539	1ejw	A	00	3540	1fwj	A	00
3541	1xil	A	02	3542	1xhx	B	02
3543	1xhz	D	02	3544	1xhx	C	02
3545	1xhx	D	02	3546	2pzs	D	02
3547	2bs2	E	02	3548	3ae1	B	02
3549	1gte	C	02	3550	2wp9	F	02
3551	2ws3	F	02	3552	3cir	B	02
3553	2ux4	H	01	3554	3g7f	H	01
3555	1ry5	H	01	3556	3dta	H	01
3557	1dxr	H	01	3558	1f6n	H	01
3559	1xko	B	00	3560	1squ	B	00
3561	2f9z	B	00	3562	1xkr	A	00
3563	2f9z	A	00	3564	1xko	A	00
3565	3prq	O	02	3566	3bz2	O	02
3567	3kzi	O	02	3568	3bz1	O	02
3569	1s5l	o	02	3570	3a0b	o	02
3571	1nog	A	00	3572	2idx	A	00
3573	3cil	A	01	3574	1wyl	C	00
3575	1wvt	B	00	3576	2zhz	C	01
3577	1g9b	A	02	3578	1s0b	A	02

3579	1s0c	A	02	3580	1s0f	A	02
3581	1g9d	A	02	3582	1g9c	A	02
3583	6mht	A	02	3584	2z6q	A	02
3585	8mht	A	02	3586	2c7o	A	02
3587	1svu	A	02	3588	2c7p	A	02
3589	2iuk	A	05	3590	3bne	A	05
3591	2iuj	A	05	3592	1rov	A	05
3593	2p0m	A	02	3594	1lnh	A	05
3595	1jro	H	04	3596	1jrp	B	02
3597	1jrp	F	04	3598	1n62	E	02
3599	1sb3	A	03	3600	1zcs	A	05
3601	3k89	A	01	3602	3r97	A	01
3603	2cuy	A	01	3604	3ezo	A	01
3605	1nm2	A	02	3606	3im8	A	01
3607	3qic	A	01	3608	1sz2	A	02
3609	3h1v	X	01	3610	1qha	B	02
3611	2nzt	B	03	3612	2nzt	A	03
3613	1i32	C	02	3614	3cmc	R	02
3615	1hdg	Q	02	3616	1ofg	D	02
3617	2xf8	G	02	3618	3ids	A	02
3619	1o17	D	02	3620	1zyk	A	02
3621	1gxb	B	02	3622	1o17	C	02
3623	2wk5	A	02	3624	1otp	A	02
3625	3hi8	F	00	3626	3p87	D	00
3627	3lx2	B	00	3628	2zvw	G	00
3629	1rwz	A	00	3630	1ud9	B	00
3631	2y2b	A	00	3632	2y2e	A	00
3633	2y2d	C	00	3634	1iya	A	00
3635	2y2b	B	00	3636	2y2d	A	00
3637	2yze	G	00	3638	1r51	A	00
3639	3bk8	A	00	3640	1xy3	H	00
3641	2fub	A	00	3642	1xxj	D	00
3643	2k7w	A	00	3644	1g5o	A	00
3645	1ysg	A	00	3646	1ysi	A	00
3647	1q59	A	00	3648	1r2d	A	00
3649	1qtu	A	00	3650	1a1x	A	00
3651	1qtt	A	00	3652	1jnp	B	00
3653	1jsg	A	00	3654	1jnp	A	00
3655	1ik3	A	01	3656	1y4k	A	01
3657	1fgq	A	01	3658	1gpl	A	02
3659	3bnc	A	01	3660	2qz7	A	01
3661	3p4q	D	00	3662	3ae5	C	02
3663	2wdr	D	00	3664	2h88	C	02
3665	2wdq	K	00	3666	1yq3	C	02
3667	1m8p	C	02	3668	1g8g	B	02
3669	1t62	A	00	3670	1m8p	A	02
3671	1jhd	A	02	3672	1j70	C	02
3673	1ggo	A	02	3674	1jde	A	02
3675	1vbg	A	04	3676	1kc7	A	02
3677	1buk	0	02	3678	2r82	A	04
3679	2vpw	E	04	3680	2vpw	A	04
3681	2vpz	A	04	3682	2vpy	A	04
3683	2vpz	E	04	3684	2vpy	E	04
3685	1f3z	A	00	3686	1o2f	A	00
3687	1f3g	A	00	3688	3our	D	00
3689	3our	F	00	3690	1glc	F	00
3691	2xur	B	02	3692	3d1f	B	02
3693	1ok7	B	02	3694	1mmi	B	02
3695	3f1v	A	01	3696	1vpk	A	03
3697	3cdd	A	02	3698	3cdd	C	02
3699	3cdd	F	02	3700	3cdd	B	02
3701	3cdd	E	02	3702	3cdd	D	02
3703	115o	A	01	3704	114n	A	01
3705	114h	A	01	3706	1j33	A	01

3707	1jhr	A	01	3708	1jho	A	01
3709	2qa4	P	03	3710	2gze	A	00
3711	2l4b	A	00	3712	1qvf	O	02
3713	3h4e	L	02	3714	3cce	P	03
3715	1k1f	A	00	3716	2ql2	D	00
3717	1k1f	H	00	3718	1am9	A	00
3719	1k1f	E	00	3720	1am9	D	00
3721	1b79	D	00	3722	1t3w	B	00
3723	1jwe	A	00	3724	1b79	B	00
3725	1b79	C	00	3726	1t3w	A	00
3727	1tx6	J	02	3728	2r33	B	00
3729	1mvz	A	00	3730	2iln	I	00
3731	1d6r	I	00	3732	1bbi	A	00
3733	3bsu	C	00	3734	3bl4	A	02
3735	1mjg	N	04	3736	3i04	P	04
3737	3i01	N	04	3738	3git	B	02
3739	1pto	H	01	3740	3g4o	B	01
3741	1pre	B	01	3742	1prt	C	01
3743	3g4n	B	01	3744	3g4n	A	01
3745	1zck	C	00	3746	2zn7	A	00
3747	1c83	A	00	3748	2qdm	A	00
3749	2b4s	A	00	3750	2qbq	A	00
3751	1zqf	A	04	3752	3mby	A	04
3753	1jms	A	04	3754	2bpg	B	04
3755	1zqy	A	03	3756	7icl	A	04
3757	2h21	B	01	3758	2h2j	A	01
3759	2h21	A	01	3760	1p0y	B	01
3761	2h21	C	01	3762	2h2e	B	01
3763	2dr6	A	02	3764	1jlf	B	02
3765	2qov	T	00	3766	3i9c	T	00
3767	1u8s	B	02	3768	2p9e	B	03
3769	2yzo	A	00	3770	1vr0	A	00
3771	1vr0	B	00	3772	2yzo	B	00
3773	2yyv	A	00	3774	1vr0	C	00
3775	3cw0	D	00	3776	3cw0	C	00
3777	3cw0	A	00	3778	2idg	A	00
3779	3efp	A	00	3780	1s9u	A	00
3781	2a1h	B	02	3782	1a0g	B	02
3783	2hg8	B	02	3784	2ej2	B	02
3785	2hgx	B	02	3786	2dab	A	02
3787	1nzu	A	02	3788	3ita	B	02
3789	3a3j	A	02	3790	3ita	C	02
3791	3beb	A	02	3792	3it9	B	02
3793	2ad6	D	00	3794	1h4i	D	00
3795	4aah	B	00	3796	1h4i	B	00
3797	2d0v	B	00	3798	1b2n	B	00
3799	1oh8	B	01	3800	1oh6	A	01
3801	1ng9	A	01	3802	1wbd	A	01
3803	1r0v	A	02	3804	1fw6	A	01
3805	2isn	A	00	3806	3fxo	A	01
3807	2v06	A	00	3808	1txo	B	00
3809	3nmn	B	00	3810	2pom	A	00
3811	1blw	B	00	3812	1f95	B	00
3813	1yo3	A	00	3814	1bkq	O	00
3815	1f3c	B	00	3816	3e2b	A	00
3817	1nis	A	04	3818	1c97	A	04
3819	1l5j	A	02	3820	1b0m	A	04
3821	6acn	A	04	3822	1b0k	A	04
3823	1hbu	D	01	3824	1hbm	D	01
3825	1hbn	A	01	3826	1hbo	A	01
3827	3m2r	D	01	3828	3m2v	A	01
3829	1pfs	A	00	3830	3gv6	A	00
3831	3mxg	A	00	3832	2r6g	B	02
3833	2chb	H	00	3834	1b9m	B	02

3835	1sfo	E	02	3836	2e2h	F	00
3837	3gtl	E	02	3838	1twg	F	00
3839	2nvy	E	02	3840	2ja6	E	02
3841	3l4j	A	02	3842	3i1o	H	02
3843	1s9f	C	04	3844	1b6s	C	03
3845	3gv7	B	04	3846	3fki	G	01
3847	1m1n	D	04	3848	1g21	B	04
3849	1m34	D	04	3850	1qh1	D	04
3851	1qgu	B	04	3852	115h	B	04
3853	1r0d	G	00	3854	1r0d	I	00
3855	1r0d	F	00	3856	1r0d	A	00
3857	1r0d	D	00	3858	1r0d	B	00
3859	3dtr	M	02	3860	3dsy	L	02
3861	1jgw	M	02	3862	2wjn	M	01
3863	1rzz	R	02	3864	2hh1	L	01
3865	3i9j	A	01	3866	3i9k	B	01
3867	1r16	A	01	3868	2i67	A	01
3869	3dzg	A	01	3870	1isf	B	01
3871	1hbk	A	00	3872	1y19	J	01
3873	2d2q	B	02	3874	2wh5	A	00
3875	1mk9	B	01	3876	2dik	A	05
3877	3git	E	01	3878	3git	B	01
3879	3git	D	01	3880	1oao	D	03
3881	1ru3	A	03	3882	3i01	P	03
3883	2hyw	A	03	3884	2ie6	A	01
3885	1hak	A	03	3886	1aei	F	02
3887	1pbv	A	01	3888	1ku1	B	01
3889	2a7x	A	02	3890	1n2j	A	02
3891	3ivg	B	02	3892	2x3f	A	02
3893	1v8f	B	02	3894	2gr7	A	00
3895	1nou	A	01	3896	1gql	A	01
3897	1mqp	A	01	3898	1qba	A	02
3899	2gk1	G	01	3900	1k9d	A	01
3901	1jfz	D	00	3902	1rc5	B	00
3903	1rc5	D	00	3904	1yyk	B	01
3905	1yyw	A	01	3906	1yz9	A	01
3907	3cw7	B	02	3908	3d4v	D	02
3909	1diz	A	02	3910	1rrq	A	02
3911	1lww	A	03	3912	3g0q	A	02
3913	2eij	S	00	3914	1occ	F	00
3915	2occ	F	00	3916	3asn	S	00
3917	1ocr	F	00	3918	3abl	S	00
3919	6gep	A	03	3920	5aop	A	03
3921	1aop	A	01	3922	4gep	A	01
3923	3aop	A	03	3924	2aop	A	01
3925	2e5y	A	01	3926	1bsn	A	01
3927	1h8e	H	00	3928	2rq6	A	01
3929	1e79	H	01	3930	1aqt	A	01
3931	3ku8	A	00	3932	3ku8	B	00
3933	1x75	B	00	3934	3l4j	A	05
3935	1ab4	A	03	3936	1x75	A	00
3937	1fd4	M	00	3938	1fd3	A	00
3939	1fd4	H	00	3940	1fqq	A	00
3941	1fd4	D	00	3942	1fd3	C	00
3943	2btv	D	03	3944	2btv	G	03
3945	2btv	F	03	3946	1bvp	2	03
3947	2btv	J	03	3948	2btv	R	03
3949	3ep3	A	00	3950	1jl0	A	00
3951	1msv	B	00	3952	3epb	A	00
3953	3dz4	A	00	3954	1i7b	A	00
3955	1jl5	A	00	3956	1koo	C	02
3957	2ast	B	00	3958	1gwb	B	00
3959	1o6t	A	01	3960	1oto	A	00
3961	2an6	C	01	3962	2an6	D	01

3963	2an6	B	01	3964	1k2f	A	02
3965	1k2f	B	02	3966	2an6	A	01
3967	1tbg	F	00	3968	1tbg	H	00
3969	1ioj	A	00	3970	1b9x	B	00
3971	3ah8	G	00	3972	1b9y	B	00
3973	2e10	A	02	3974	2pb6	B	02
3975	2dv5	A	02	3976	2e17	B	02
3977	2p6k	A	02	3978	2e16	A	02
3979	1m22	B	00	3980	3qk5	A	00
3981	2wj1	B	00	3982	1ocm	B	00
3983	1o9n	A	00	3984	1ocl	A	00
3985	1lxa	A	02	3986	2jf2	A	02
3987	3i3x	C	02	3988	2qia	A	02
3989	2aq9	A	02	3990	3i3a	B	02
3991	1kor	B	02	3992	1kp3	A	02
3993	1kh3	B	02	3994	1kp2	A	02
3995	1jlz	D	02	3996	1kh3	A	02
3997	1h6d	F	01	3998	1y1s	A	00
3999	2v54	B	00	4000	3d8u	A	01
4001	1mmf	E	00	4002	3oi7	C	01
4003	2a69	C	02	4004	2be5	M	02
4005	2be5	C	02	4006	1smy	C	02
4007	2ppb	C	02	4008	1zyr	C	02
4009	2z8y	O	05	4010	1mjg	P	05
4011	1mjg	O	05	4012	1mjg	M	05
4013	3git	E	03	4014	3git	F	03
4015	1y9i	B	00	4016	1rfz	B	00
4017	1y9i	C	00	4018	1tlq	A	00
4019	1y9i	A	00	4020	1y9i	D	00
4021	1p3f	C	00	4022	1p3i	B	00
4023	1p3o	D	00	4024	3afa	H	00
4025	1kx5	F	00	4026	1h3o	B	00
4027	3me1	A	00	4028	2nuu	A	00
4029	2b2h	A	00	4030	2nmr	A	00
4031	3me1	B	00	4032	2nuu	B	00
4033	1s20	B	01	4034	1nxu	A	01
4035	2g8y	B	01	4036	1nxu	B	01
4037	1s20	E	01	4038	1s20	D	01
4039	2oa7	B	02	4040	2vcr	G	02
4041	3ik9	A	02	4042	2dsa	C	02
4043	3fr3	B	02	4044	3cwg	A	01
4045	2xq9	D	01	4046	2zjv	B	00
4047	1uw6	S	00	4048	2xqa	A	01
4049	2yks	A	01	4050	2xq3	C	01
4051	1o83	D	00	4052	1o82	D	00
4053	1o82	A	00	4054	1o84	A	00
4055	1o83	B	00	4056	1o83	C	00
4057	3q5s	A	03	4058	1d5y	B	03
4059	1bow	A	00	4060	1d5y	A	03
4061	2bow	A	00	4062	3gk6	A	00
4063	1nwt	A	02	4064	2anf	A	02
4065	1p5q	C	01	4066	1e6z	A	02
4067	3fxy	B	02	4068	1ll4	A	02
4069	3b9r	B	02	4070	3ar7	A	02
4071	3nal	A	02	4072	3ar8	A	02
4073	1wpg	D	02	4074	2agv	B	02
4075	1a2v	A	03	4076	1hn1	C	05
4077	1k1y	B	02	4078	1f4h	B	05
4079	2oqe	F	03	4080	3bup	A	04
4081	1jpw	D	00	4082	1jpw	F	00
4083	1i7w	D	00	4084	1jpw	E	00
4085	1g3j	D	00	4086	1g3j	B	00
4087	3ezq	D	00	4088	3mop	G	00
4089	3ygs	P	00	4090	1alz	A	00

4091	3ezq	F	00	4092	1ucp	A	00
4093	1l17	A	00	4094	1nhe	A	00
4095	1gf3	A	00	4096	1b7s	A	00
4097	3c7z	A	00	4098	3rw8	A	00
4099	2a4z	A	04	4100	1e7u	A	04
4101	3oaw	A	04	4102	3cst	A	04
4103	1e7v	A	04	4104	1e8y	A	04
4105	2wh4	I	01	4106	1vt2	H	01
4107	3kiw	I	01	4108	3i8i	K	01
4109	3df4	H	01	4110	1vs9	G	01
4111	1n8r	X	01	4112	3gw1	B	00
4113	1yjn	W	01	4114	3i1t	Z	00
4115	1mws	B	03	4116	2awb	Y	00
4117	3i8g	U	00	4118	3oto	R	00
4119	2y16	R	00	4120	3ofx	R	00
4121	2qbb	R	00	4122	2xqd	R	00
4123	3og2	A	03	4124	1tg7	A	03
4125	1xc6	A	03	4126	3ogs	A	03
4127	3ogr	A	03	4128	3ogv	A	03
4129	2hqd	A	04	4130	3nog	A	08
4131	2hrt	C	08	4132	2dhh	C	08
4133	2hrt	C	04	4134	1t9w	A	08
4135	1o14	B	02	4136	3bf5	B	02
4137	2ajr	A	02	4138	3bf5	A	02
4139	2qcv	A	02	4140	1o14	A	02
4141	3nid	C	00	4142	2ce8	B	00
4143	1vyh	C	01	4144	2bbk	J	00
4145	1gp2	B	00	4146	3qi0	B	00
4147	3aso	D	00	4148	1occ	Q	00
4149	3aso	Q	00	4150	3abm	D	00
4151	2eij	D	00	4152	2eil	D	00
4153	3era	A	00	4154	1es7	B	00
4155	1kbs	A	00	4156	2ux2	B	00
4157	1rew	D	00	4158	1xt3	A	00
4159	1yfl	F	02	4160	1q0t	A	02
4161	1yfl	D	02	4162	1yfl	E	02
4163	1yfl	A	02	4164	1yfl	A	02
4165	1q23	H	00	4166	2bgh	B	02
4167	1pd5	G	00	4168	1dpd	A	00
4169	1cla	A	00	4170	2i9d	C	00
4171	3o9v	A	01	4172	2gbi	B	02
4173	2rgu	A	01	4174	1hj3	A	02
4175	3o95	D	01	4176	2g63	D	02
4177	1z52	A	02	4178	3c0m	B	02
4179	3c0o	A	02	4180	3c0o	B	02
4181	1pre	B	02	4182	3g4o	B	02
4183	2jqj	A	00	4184	1u7v	C	00
4185	3a77	B	00	4186	1j2f	B	00
4187	3i6w	C	01	4188	1k3q	A	00
4189	2phj	B	00	4190	115x	B	00
4191	2e6g	A	00	4192	2e6b	D	00
4193	2e6g	E	00	4194	2e6h	A	00
4195	2cyh	A	00	4196	1c5f	M	00
4197	1qtl	E	00	4198	1qtl	G	00
4199	1qtl	K	00	4200	1c5f	A	00
4201	1nl0	G	00	4202	1j34	C	00
4203	1cfh	A	00	4204	1whf	A	01
4205	1whe	A	01	4206	1pfx	L	01
4207	1a79	B	01	4208	2fco	B	00
4209	1ob8	B	00	4210	1ob8	A	00
4211	1rzn	B	00	4212	2wew	C	00
4213	1cmx	A	00	4214	1xd3	C	00
4215	3irt	A	00	4216	3irt	B	00
4217	1xd3	A	00	4218	2etl	B	00

4219	3cf2	A	02	4220	3hu2	B	02
4221	1cr5	A	02	4222	1ypw	C	02
4223	3hu2	A	02	4224	1r7r	A	02
4225	2roo	A	00	4226	2dcw	A	00
4227	1hy9	A	00	4228	1oaw	A	00
4229	1oma	A	00	4230	1boe	A	00
4231	2qzy	A	02	4232	2rke	A	02
4233	2fah	D	02	4234	2faf	B	02
4235	3dt7	B	02	4236	1k3d	A	02
4237	1nql	A	02	4238	2ahx	A	02
4239	1m6b	B	02	4240	3i2t	A	02
4241	1n8z	C	02	4242	1m6b	A	04
4243	1n5l	A	02	4244	1wbq	C	02
4245	1xgo	A	01	4246	1wbq	B	02
4247	1w2m	F	02	4248	1a16	A	02
4249	1wqa	A	01	4250	1c47	A	02
4251	1c47	B	01	4252	1tuo	A	01
4253	1k2y	X	01	4254	1kfq	A	03
4255	1v55	L	00	4256	2eik	Y	00
4257	2dyr	L	00	4258	3ag2	Y	00
4259	1v54	L	00	4260	2eij	L	00
4261	2cel	A	00	4262	2rfy	A	00
4263	1egn	A	00	4264	2rfy	B	00
4265	2rfz	A	00	4266	4cel	A	00
4267	1oih	A	00	4268	1oih	B	00
4269	3r1j	B	00	4270	1oii	B	00
4271	1gqw	B	00	4272	1nx4	A	00
4273	2wh4	O	00	4274	1yjw	K	00
4275	3i9e	N	00	4276	3ccv	K	00
4277	2qp1	K	00	4278	2wh2	O	00
4279	1bgv	A	03	4280	1hrd	B	03
4281	1hrd	C	03	4282	1k89	A	03
4283	1hrd	A	03	4284	1aup	A	03
4285	2emn	A	00	4286	1gfl	B	00
4287	2psl	C	00	4288	2b3q	B	00
4289	2pxs	A	00	4290	2a47	A	00
4291	2alz	A	03	4292	1sv0	B	00
4293	1xsp	A	01	4294	2hoi	H	01
4295	3crx	A	02	4296	2f1j	A	01
4297	1io4	D	00	4298	1e50	B	00
4299	1h9d	D	00	4300	1cl3	A	00
4301	1h9d	B	00	4302	1e50	D	00
4303	1tad	A	02	4304	3c16	C	02
4305	1tad	C	02	4306	1as2	A	02
4307	2om2	C	02	4308	1y3a	B	02
4309	3l8u	A	00	4310	1k3r	A	01
4311	1k3r	B	01	4312	1gz0	D	02
4313	1to0	G	00	4314	3aia	B	00
4315	1ojd	E	03	4316	1s2q	B	03
4317	3cpi	G	03	4318	1s3b	B	03
4319	2v60	A	03	4320	1oj9	A	03
4321	1y8r	B	03	4322	2uub	D	02
4323	3pyn	D	02	4324	2qbb	D	02
4325	3ofa	D	02	4326	1ibm	D	02
4327	1tp7	C	03	4328	2b43	A	04
4329	1wt4	B	00	4330	1om2	A	00
4331	2e9t	D	03	4332	1tp7	D	03
4333	1e5l	B	03	4334	2axq	A	03
4335	1e5l	A	03	4336	1e5q	A	03
4337	1e5q	C	03	4338	1e5q	H	03
4339	1f3i	A	02	4340	1l1a	A	03
4341	1mm8	A	03	4342	1mus	A	03
4343	1b7e	A	02	4344	1mur	A	03
4345	1vq7	P	01	4346	3cc4	P	01

4347	1qvg	O	01	4348	1mlk	Q	01
4349	1vql	P	01	4350	1kqs	O	01
4351	1si3	A	00	4352	1r6z	P	03
4353	1si2	A	00	4354	1r6z	Z	03
4355	1t2s	A	00	4356	1r6z	A	03
4357	1tk1	A	00	4358	3e8j	A	00
4359	3dws	A	00	4360	1tlb	Q	00
4361	3e8j	B	00	4362	1tlb	U	00
4363	2a06	S	00	4364	1ntm	F	00
4365	3h1i	S	00	4366	3174	S	00
4367	1bgy	R	00	4368	1kyo	R	00
4369	2csc	A	01	4370	2h12	F	02
4371	1iom	A	01	4372	5csc	A	01
4373	3o8j	E	01	4374	1nxe	A	02
4375	3hun	A	02	4376	1jq3	A	02
4377	3hun	B	02	4378	2qfm	C	02
4379	2zsu	F	01	4380	1jq3	B	02
4381	1u99	A	02	4382	1mo6	A	02
4383	1ubf	A	02	4384	1g18	A	02
4385	1mo3	A	02	4386	1xp8	A	02
4387	3bzs	A	00	4388	3bzo	B	00
4389	3bzt	A	00	4390	3c00	B	00
4391	3bzp	A	00	4392	3c03	A	00
4393	2r5x	A	00	4394	3cws	B	01
4395	2nwv	A	00	4396	2z8u	P	01
4397	1m5q	N	02	4398	1tba	B	01
4399	1mpx	C	02	4400	1ju4	A	03
4401	2b4k	C	02	4402	1ryy	F	02
4403	117q	A	03	4404	2b9v	P	02
4405	2hnc	A	00	4406	3ieo	A	00
4407	1t9n	A	00	4408	2nxt	A	00
4409	1cva	A	00	4410	2qoa	A	00
4411	1ooy	B	01	4412	3oxo	D	01
4413	2nrb	B	01	4414	1k6d	B	00
4415	3oxo	C	01	4416	1xr4	A	01
4417	1wow	B	01	4418	2gm7	B	00
4419	1rcw	B	00	4420	1ix3	A	00
4421	1we1	A	00	4422	2q4x	B	00
4423	1qrh	A	00	4424	1eri	A	00
4425	1qc9	C	00	4426	1d02	B	00
4427	1qps	A	00	4428	1ckq	A	00
4429	1q32	B	01	4430	1qzq	A	02
4431	1v0t	A	01	4432	1xdp	A	03
4433	1xdo	B	04	4434	2o8r	B	03
4435	1ovd	A	02	4436	3c61	C	02
4437	2b4g	C	02	4438	2e6a	B	02
4439	1jrb	B	02	4440	3oix	D	02
4441	3knq	B	01	4442	1g3p	A	01
4443	3knq	A	01	4444	1fgp	A	00
4445	3dgs	A	01	4446	3dgs	B	01
4447	1cyy	A	02	4448	1cy9	B	02
4449	1cy7	A	03	4450	1cy0	A	03
4451	1cy9	A	02	4452	1cy2	A	03
4453	1pww	B	03	4454	1pww	B	03
4455	1zxv	A	03	4456	1pww	A	03
4457	1yqy	A	02	4458	1pwp	A	03
4459	1x1g	A	00	4460	1miz	B	02
4461	1k5g	E	00	4462	1unp	A	00
4463	2ysl	A	00	4464	1oqn	A	00
4465	1twj	C	00	4466	1t4a	B	00
4467	1t4a	A	00	4468	3d54	J	00
4469	1vq3	C	00	4470	3d54	G	00
4471	1jg9	A	03	4472	2pwh	B	02
4473	2pwf	D	02	4474	2pwf	B	02

4475	1s46	A	03	4476	1jgi	A	03
4477	2z48	B	03	4478	2z48	A	03
4479	2z49	A	03	4480	2z49	B	03
4481	1vcl	A	03	4482	1vcl	B	03
4483	3ptx	A	02	4484	3hhz	N	02
4485	2qvj	E	02	4486	3pu1	C	02
4487	3pu4	E	02	4488	2gic	B	02
4489	3dcl	D	01	4490	3n99	E	01
4491	3dcl	C	01	4492	3n99	J	01
4493	3n99	G	01	4494	3dcl	A	01
4495	1n8v	A	00	4496	1kx8	A	00
4497	1kx9	A	00	4498	2gvs	A	00
4499	1n8u	A	00	4500	1k19	A	00
4501	3ij4	A	03	4502	2qts	D	03
4503	2qts	E	03	4504	2qts	C	03
4505	2qts	B	03	4506	3hgc	A	03
4507	2oz9	R	00	4508	1uam	A	02
4509	1zt9	D	00	4510	1tro	E	00
4511	1jhg	A	00	4512	1oy5	C	02
4513	3p4r	M	02	4514	1chu	A	02
4515	1kf6	M	02	4516	2fje	A	02
4517	2b76	M	02	4518	2fja	A	02
4519	1dj3	A	02	4520	1hop	B	02
4521	1hoo	B	02	4522	1gin	A	02
4523	1kkf	A	02	4524	1cgl	A	02
4525	2pl6	H	00	4526	1r2m	A	00
4527	2pl6	D	00	4528	2pl7	B	00
4529	2gvm	B	00	4530	2gvm	A	00
4531	1yaw	A	02	4532	1vqv	B	02
4533	3c9u	B	02	4534	3c9r	B	02
4535	1cli	D	02	4536	2btu	B	02
4537	1key	B	02	4538	1jnr	A	03
4539	1zoy	A	03	4540	3ehz	E	02
4541	1cun	B	01	4542	1u4q	A	02
4543	1gp0	A	00	4544	3k43	A	00
4545	2rl7	A	00	4546	1sz0	A	01
4547	3k41	A	00	4548	2cnj	D	00
4549	1zdn	B	00	4550	1c4z	D	00
4551	2bep	A	00	4552	3a4s	B	00
4553	1kpp	A	00	4554	2kly	A	00
4555	1kqs	W	00	4556	3ccl	X	00
4557	1m90	Y	00	4558	3cd6	X	00
4559	3g71	X	00	4560	3cc4	X	00
4561	1dd9	A	03	4562	1eqn	E	03
4563	1erc	A	00	4564	3b39	A	03
4565	1eqn	A	03	4566	1eqn	D	03
4567	1tly	B	00	4568	2wjq	A	00
4569	1qd6	C	00	4570	2wjr	A	00
4571	1ild	A	00	4572	1im0	A	00
4573	2zxw	V	00	4574	3aso	I	00
4575	3asn	I	00	4576	2eik	I	00
4577	3abm	I	00	4578	2dys	V	00
4579	3ag2	J	00	4580	3aso	W	00
4581	1v54	W	00	4582	3ag1	J	00
4583	3abk	W	00	4584	1ocz	J	00
4585	1dgs	A	01	4586	3chx	I	02
4587	2ebo	B	00	4588	3m8a	H	00
4589	1en5	A	01	4590	2a9h	D	00
4591	1rgn	H	02	4592	1fnp	H	02
4593	1pss	H	02	4594	1e6d	H	02
4595	1txw	H	02	4596	1aij	T	02
4597	2eij	T	00	4598	2eim	T	00
4599	3asn	G	00	4600	2eij	G	00
4601	1v55	T	00	4602	3asn	T	00

4603	2jmu	A	01	4604	3bhd	A	00
4605	1yem	A	00	4606	2fbl	A	00
4607	2dc4	B	00	4608	3bhd	B	00
4609	117v	B	00	4610	2qi9	A	00
4611	117v	A	00	4612	2nq2	B	00
4613	2qi9	B	00	4614	2nq2	A	00
4615	1k98	A	04	4616	3bul	A	04
4617	1msk	A	02	4618	3iv9	A	04
4619	1k7y	A	04	4620	3iva	A	04
4621	2wdj	N	00	4622	1vs6	J	00
4623	3msl	J	00	4624	3ccu	J	00
4625	2j03	N	00	4626	2qoz	J	00
4627	1j5o	B	01	4628	3hvt	A	01
4629	1hmv	E	01	4630	1rti	B	01
4631	3kk1	A	01	4632	2r2t	A	02
4633	3mgz	A	00	4634	2ris	A	00
4635	1hru	A	00	4636	1tku	A	00
4637	1pvw	A	00	4638	1g58	B	00
4639	1ua8	A	00	4640	3buu	A	00
4641	2zpd	A	00	4642	1iwm	B	00
4643	1iwl	A	00	4644	1iwn	A	00
4645	1ccw	B	02	4646	1i9c	B	02
4647	1i9c	D	02	4648	1cb7	B	02
4649	1ccw	D	02	4650	1cb7	D	02
4651	3cd5	C	01	4652	1hw8	C	01
4653	3cd7	A	01	4654	1hw8	D	01
4655	2q6c	A	01	4656	3cd5	B	01
4657	1ajy	A	01	4658	1f5e	P	00
4659	1d66	B	01	4660	1ajy	B	01
4661	1zme	C	01	4662	125d	0	00
4663	2h5n	D	00	4664	2ou3	B	01
4665	2h5n	C	00	4666	2ou3	A	01
4667	2h5n	B	01	4668	2h5n	A	01
4669	2akc	A	00	4670	1vnc	A	01
4671	1idu	A	01	4672	2a96	A	00
4673	2ipb	A	00	4674	2ipb	D	00
4675	1blb	D	02	4676	2a5m	A	02
4677	2jdf	A	01	4678	2g98	A	01
4679	3hzb	B	00	4680	3qk3	A	01
4681	1gs0	A	01	4682	1uk1	B	01
4683	3c49	A	01	4684	3l3m	A	01
4685	2rcw	A	01	4686	1uk0	B	01
4687	1yww	A	00	4688	1xkt	A	02
4689	1xkt	B	02	4690	1jyg	A	00
4691	1ryk	A	00	4692	2px6	B	02
4693	3r7k	D	03	4694	2q8g	A	01
4695	3r7k	B	03	4696	2cj8	A	00
4697	3aff	A	03	4698	1x90	B	00
4699	2pvx	B	00	4700	1pfw	A	01
4701	2pvx	C	00	4702	3h3v	M	00
4703	3qvd	A	02	4704	1i50	L	00
4705	1jnb	G	03	4706	1ijg	H	03
4707	1cxr	A	00	4708	1ed0	A	00
4709	2plh	A	00	4710	1ijg	D	03
4711	2i9a	B	02	4712	1gp9	B	02
4713	1tpk	C	00	4714	2i9a	D	02
4715	2i9b	A	02	4716	3e6p	L	01
4717	1rqf	A	02	4718	2otl	Z	00
4719	1yij	Z	00	4720	3gtm	I	02
4721	1qvg	Y	00	4722	2hfl	A	01
4723	2j09	A	03	4724	1own	A	03
4725	1qnf	A	03	4726	1iqu	A	03
4727	1owo	A	03	4728	1tez	A	03
4729	2o6y	A	01	4730	2o7d	E	01

4731	1t6p	C	01	4732	1u15	C	01
4733	1q5n	A	01	4734	2o78	C	01
4735	1ph9	B	00	4736	1ph1	B	00
4737	2i0q	B	00	4738	1pa6	B	00
4739	1ph3	B	00	4740	1ph5	B	00
4741	1zue	A	00	4742	1z99	A	00
4743	1ahl	A	00	4744	1el6	B	02
4745	1h5o	A	00	4746	1atx	A	00
4747	2haw	B	01	4748	1wpm	A	01
4749	1wpm	A	00	4750	1wpp	A	01
4751	1i74	A	01	4752	1wpp	B	01
4753	3bp9	D	00	4754	3h4e	B	01
4755	2x82	D	00	4756	3bp9	X	00
4757	3h4e	K	01	4758	1afv	B	00
4759	1olz	B	02	4760	2uzy	B	02
4761	2uzy	D	02	4762	1shy	B	02
4763	3ol2	A	02	4764	1olz	A	02
4765	3hfz	B	03	4766	1b70	B	03
4767	1pys	B	03	4768	2rhs	B	03
4769	2akw	B	03	4770	1eiy	B	03
4771	1u56	B	00	4772	1xbn	A	00
4773	3m0b	A	00	4774	3nvu	A	00
4775	3eee	D	00	4776	3lah	B	00
4777	1t2p	C	00	4778	1ng5	B	00
4779	1rz2	A	00	4780	1t2w	A	00
4781	1t2p	A	00	4782	1ng5	A	00
4783	1b25	A	03	4784	1b4n	C	03
4785	1b25	D	03	4786	1b4n	A	03
4787	1b25	B	03	4788	1b4n	D	03
4789	2igs	H	00	4790	2igs	E	00
4791	2igs	D	00	4792	2igs	C	00
4793	2igs	F	00	4794	2igs	A	00
4795	2dpv	0	00	4796	4dpv	Z	00
4797	3kic	I	00	4798	3kic	L	00
4799	3kie	O	00	4800	3kie	G	00
4801	2xgl	G	00	4802	3d5a	G	00
4803	3ohd	G	00	4804	3huy	G	00
4805	3df1	G	00	4806	1ibk	G	00
4807	3esm	A	00	4808	1esm	A	00
4809	1esm	B	00	4810	5esm	A	00
4811	2esm	A	00	4812	4esm	A	00
4813	1duj	A	00	4814	1s2h	A	00
4815	1go4	B	00	4816	1go4	A	00
4817	1go4	D	00	4818	1klq	A	00
4819	1v9y	A	00	4820	2v8c	A	00
4821	1drq	A	00	4822	2vv7	A	00
4823	2pt7	A	01	4824	3bwl	C	00
4825	3fk1	A	01	4826	2z2c	B	01
4827	2gg6	A	01	4828	1ejd	B	01
4829	3lth	A	02	4830	3iss	G	01
4831	1m1g	B	02	4832	1m1g	C	02
4833	1npp	C	02	4834	1npr	A	02
4835	1m1g	A	02	4836	1m1g	D	02
4837	1vg9	C	02	4838	3cpi	H	02
4839	3plw	A	02	4840	1lv0	A	02
4841	3cph	G	02	4842	2bcg	G	02
4843	2a68	K	02	4844	2a68	A	02
4845	2be5	B	02	4846	2wro	Z	02
4847	3knk	Z	02	4848	2vum	C	02
4849	1iu3	F	00	4850	1j3e	A	00
4851	1iu3	C	00	4852	1lrr	A	00
4853	3fmt	E	02	4854	1lrr	D	00
4855	2zhx	M	00	4856	1mwj	A	00
4857	4eug	A	00	4858	1emj	A	00

4859	llau	E	00	4860	2oxm	A	00
4861	loam	B	02	4862	3c5x	A	02
4863	loke	B	02	4864	loan	A	02
4865	lurz	A	02	4866	lsvb	A	02
4867	ljkv	D	02	4868	lo9i	C	02
4869	lo9i	A	02	4870	ljkv	A	02
4871	ljkv	C	02	4872	ljkv	E	02
4873	lnk8	A	02	4874	lxc9	A	02
4875	2xy5	A	02	4876	lt8e	A	02
4877	lqss	A	02	4878	2kzm	A	02
4879	lgxd	C	01	4880	luea	D	01
4881	lgxd	D	01	4882	lbuv	T	01
4883	luea	B	01	4884	lbr9	A	01
4885	lpd7	A	00	4886	le9l	A	00
4887	ls5q	B	00	4888	ls5r	B	00
4889	2l9s	B	00	4890	2f05	A	00
4891	lfop	B	02	4892	ldlx	A	02
4893	lp6h	B	02	4894	lk2r	A	02
4895	lrs8	A	02	4896	ldww	A	02
4897	lw2b	B	02	4898	3g6e	B	03
4899	lyjw	B	03	4900	lyhq	B	03
4901	3cce	B	03	4902	3ccj	B	03
4903	2i9z	B	00	4904	3p37	B	02
4905	2ia9	E	00	4906	le2s	P	02
4907	ln2l	A	02	4908	2ia9	F	00
4909	lm90	C	03	4910	ls72	A	03
4911	3cd6	A	03	4912	2qbe	C	03
4913	2qa4	A	03	4914	2qba	C	03
4915	la92	B	00	4916	lijd	E	00
4917	2fkw	R	00	4918	2fkw	O	00
4919	2fkw	M	00	4920	lnkz	C	00
4921	ltb6	L	00	4922	lavg	L	00
4923	ljmo	L	00	4924	2b5t	C	00
4925	2b5t	A	00	4926	2pgq	A	00
4927	lc8z	A	00	4928	li7e	A	00
4929	ls3l	A	00	4930	2fim	B	00
4931	2fim	A	00	4932	3c5n	B	00
4933	ln72	A	00	4934	lzs5	A	00
4935	3i3j	J	00	4936	2r0v	C	01
4937	3plc	A	00	4938	3ple	A	00
4939	lqty	R	00	4940	3evs	B	00
4941	2qr0	I	00	4942	2xac	A	00
4943	2vpf	F	00	4944	3eol	C	00
4945	2wsf	L	00	4946	3lw5	L	00
4947	2wse	L	00	4948	2o0l	L	01
4949	2wsc	L	00	4950	ljb0	L	00
4951	lmmo	E	00	4952	lxu5	D	00
4953	lxu3	C	00	4954	lxvf	C	00
4955	lfz2	D	00	4956	lfz2	C	00
4957	2waz	X	01	4958	ladt	A	01
4959	lanv	A	01	4960	2wb0	X	01
4961	ladu	B	01	4962	ladu	A	01
4963	3nls	B	00	4964	2q4h	B	02
4965	3nls	A	00	4966	lhxp	B	01
4967	2q4l	A	02	4968	3llj	A	00
4969	3ker	C	00	4970	lbjp	C	00
4971	2ooh	C	00	4972	lca7	B	00
4973	3mb2	G	00	4974	lgif	A	00
4975	li10	D	02	4976	3gvi	A	02
4977	lup4	C	02	4978	3gvh	B	02
4979	2cvq	B	02	4980	lbdm	B	02
4981	2y14	D	01	4982	2uxd	D	01
4983	2uu9	D	01	4984	ln32	D	01
4985	2xqd	D	01	4986	2qb9	D	01

4987	1fkp	A	05	4988	3klh	A	05
4989	1qqn	A	03	4990	1noz	A	02
4991	3i8h	N	00	4992	1rth	A	05
4993	1i3p	A	00	4994	1p35	C	00
4995	1i3s	B	00	4996	2fun	A	00
4997	1i3s	C	00	4998	1i3s	A	00
4999	2ra3	I	00	5000	1dtx	A	00
5001	6pti	A	00	5002	2fi3	I	00
5003	3fp8	I	00	5004	1co7	I	00
5005	1rhx	A	00	5006	2hyb	H	00
5007	2hyb	Q	00	5008	2hy5	B	00
5009	2qs7	A	00	5010	2hyb	B	00
5011	1ibu	A	00	5012	1ibw	C	00
5013	1ibv	A	00	5014	1ibt	E	00
5015	1hq6	A	00	5016	1ibw	A	00
5017	4rub	U	00	5018	1rsc	N	00
5019	1rcx	F	00	5020	8ruc	I	00
5021	1uw9	J	00	5022	1zho	G	01
5023	1fui	C	02	5024	3a9s	B	02
5025	1fui	A	02	5026	3a9r	B	02
5027	1fui	B	02	5028	3a9r	C	02
5029	2v2g	C	02	5030	2v41	C	02
5031	2v32	C	02	5032	1xcc	A	02
5033	1prx	B	02	5034	2v41	F	02
5035	3qvp	A	02	5036	1gal	A	02
5037	1gpe	A	02	5038	1cf3	A	02
5039	3qvr	A	02	5040	1gpe	B	02
5041	1pp9	C	00	5042	1l0n	C	00
5043	2qjk	J	00	5044	1kyo	N	00
5045	1vf5	N	00	5046	3l70	P	00
5047	1bx7	A	00	5048	1eja	B	00
5049	1hia	I	00	5050	1c9t	H	00
5051	1c9t	G	00	5052	1c9p	B	00
5053	2fsq	A	00	5054	1vdx	A	00
5055	1jh6	B	00	5056	1iuh	A	00
5057	1jh7	A	00	5058	1vgj	A	00
5059	4cox	B	01	5060	1eai	C	00
5061	6cox	A	01	5062	1apo	A	00
5063	1xkb	B	02	5064	1pxx	C	01
5065	2qk7	A	00	5066	7ahl	F	01
5067	1t5r	B	00	5068	1t5r	H	00
5069	3lkf	A	00	5070	1t5r	D	00
5071	2p9s	G	00	5072	1hu3	A	00
5073	1tnb	K	00	5074	3g2w	B	00
5075	1n95	A	00	5076	1hxi	A	00
5077	1vcb	I	01	5078	1lm8	V	01
5079	1vcb	C	01	5080	1lqb	C	01
5081	1vcb	L	01	5082	1vcb	F	01
5083	1hss	C	00	5084	1bip	A	00
5085	1hss	B	00	5086	1s6d	A	01
5087	1w2q	A	01	5088	1hss	D	00
5089	1fp7	B	03	5090	1fp7	A	03
5091	1eg7	A	03	5092	1fpm	A	03
5093	1fpm	B	03	5094	1eg7	B	03
5095	1nyc	A	00	5096	1y4h	C	00
5097	1pxv	D	00	5098	1oh1	A	00
5099	1y4h	D	00	5100	1pxv	C	00
5101	3h90	D	01	5102	3h90	B	01
5103	3h90	C	01	5104	2qfi	A	01
5105	3h90	A	01	5106	2qfi	B	01
5107	1q3a	B	00	5108	1ck7	A	01
5109	2j0t	C	00	5110	1qic	B	00
5111	1pwq	B	04	5112	3edz	B	00
5113	1lly	B	03	5114	1lly	C	03

5115	1f9d	A	03	5116	1l2a	A	03
5117	1l2a	D	03	5118	1lly	A	03
5119	3gvk	A	04	5120	1v0f	C	04
5121	3gvk	C	04	5122	1v0e	A	04
5123	1v0f	E	04	5124	1v0e	D	04
5125	1kpk	F	00	5126	2fee	B	00
5127	lots	A	00	5128	2h2p	A	00
5129	lotu	B	00	5130	2ht3	B	00
5131	1kn9	A	02	5132	1t7d	A	02
5133	1kn9	D	02	5134	1b12	C	02
5135	1kn9	C	02	5136	1t7d	B	02
5137	3gsh	A	00	5138	1siy	A	00
5139	1fk6	A	00	5140	3cee	A	00
5141	1mzl	A	00	5142	1pnb	B	00
5143	3o0a	A	00	5144	1udz	B	00
5145	2ajg	A	00	5146	1gax	B	03
5147	1wk8	B	00	5148	2ajg	B	00
5149	1kia	C	01	5150	1nbh	C	01
5151	2idk	D	01	5152	1xva	A	01
5153	1d2h	A	01	5154	1r8y	H	01
5155	3m01	A	02	5156	1n24	B	01
5157	1n21	A	01	5158	3m02	A	02
5159	1n22	B	01	5160	3g4d	B	02
5161	1bor	A	00	5162	1rmd	A	01
5163	1weu	A	00	5164	1fp0	A	00
5165	2ida	A	01	5166	2jrj	A	00
5167	1vao	B	01	5168	1q1t	B	01
5169	2w55	G	03	5170	1i19	A	01
5171	1ahz	B	01	5172	1ahu	A	01
5173	1ksa	B	00	5174	3eni	C	00
5175	3oeg	A	00	5176	1ksa	A	00
5177	3eoj	A	00	5178	3eni	A	00
5179	1zsv	D	01	5180	3fpc	D	01
5181	2oui	D	01	5182	2fze	B	01
5183	1h0k	B	01	5184	1deh	A	01
5185	3iiu	M	00	5186	1ppr	O	01
5187	2x20	M	00	5188	1ppr	M	02
5189	2x21	M	00	5190	1ppr	O	02
5191	3q81	B	00	5192	1zkj	A	00
5193	1k38	A	00	5194	1kdw	A	00
5195	1li0	A	00	5196	1mwx	B	04
5197	2w8q	A	02	5198	3rjl	C	02
5199	1o01	G	02	5200	3rhj	A	02
5201	3rhq	C	02	5202	2onm	F	02
5203	1wal	A	00	5204	1tyo	B	00
5205	1t09	A	00	5206	1cnz	B	00
5207	3flk	A	00	5208	3icd	A	00
5209	1hc1	D	02	5210	1bt2	B	00
5211	1bt1	B	00	5212	1hcy	B	02
5213	1bug	A	00	5214	1hcy	D	02
5215	2w8i	H	01	5216	2w8i	C	01
5217	2j58	C	02	5218	2w8i	C	02
5219	2w8h	E	01	5220	2w8i	H	02

Table S2A. Domain classification conflicts between CATH and automatic classification of protein fold space using Class criterium.

Index	PDB ID	Chain ID	Domain ID	Predicted class	Annotated class	Index	PDB ID	Chain ID	Domain ID	Predicted class	Annotated class
1	3die	A	01	Mainly Alpha	Alpha Beta	2	3kod	D	00	Mainly Beta	Alpha Beta
3	1opc	A	00	Alpha Beta	Mainly Alpha	4	1fxi	A	00	Mainly Beta	Alpha Beta
5	1j6n	C	02	Mainly Alpha	Alpha Beta	6	3sc6	E	02	Mainly Alpha	Alpha Beta
7	2aj2	A	02	Mainly Beta	Alpha Beta	8	2zwk	B	00	Mainly Alpha	Few Secondary Structures
9	3en6	A	01	Mainly Beta	Alpha Beta	10	2j3f	C	01	Mainly Beta	Alpha Beta
11	1yld	B	00	Mainly Beta	Few Secondary Structures	12	1rf1	E	02	Mainly Beta	Alpha Beta
13	3qu6	C	01	Alpha Beta	Mainly Alpha	14	2nn2	B	00	Mainly Alpha	Alpha Beta
15	1wg6	A	01	Alpha Beta	Mainly Beta	16	1na6	B	01	Alpha Beta	Mainly Beta
17	1cfh	A	00	Mainly Alpha	Few Secondary Structures	18	1kew	A	02	Mainly Alpha	Alpha Beta
19	2z7r	A	01	Mainly Beta	Alpha Beta	20	1wlv	C	00	Mainly Beta	Alpha Beta
21	1s98	B	00	Alpha Beta	Mainly Beta	22	1gaq	B	00	Mainly Beta	Alpha Beta
23	1jsb	A	00	Mainly Beta	Alpha Beta	24	1htm	D	00	Mainly Alpha	Alpha Beta
25	1eh6	A	01	Alpha Beta	Mainly Beta	26	1fzd	A	01	Mainly Beta	Alpha Beta
27	3g03	A	00	Alpha Beta	Mainly Alpha	28	2he0	B	00	Alpha Beta	Mainly Alpha
29	1k75	B	02	Mainly Alpha	Alpha Beta	30	1qu1	E	00	Mainly Alpha	Alpha Beta
31	2jxo	A	00	Alpha Beta	Mainly Beta	32	1go3	M	02	Mainly Beta	Alpha Beta
33	1bef	A	01	Alpha Beta	Mainly Beta	34	1j1q	A	02	Alpha Beta	Few Secondary Structures
35	2pji	A	00	Mainly Beta	Few Secondary Structures	36	2dwz	C	00	Alpha Beta	Mainly Alpha
37	1g7o	A	01	Mainly Alpha	Alpha Beta	38	1iaw	B	02	Alpha Beta	Mainly Alpha
39	2q81	D	00	Mainly Alpha	Alpha Beta	40	2f8p	A	00	Alpha Beta	Mainly Alpha
41	1s7d	A	00	Alpha Beta	Mainly Beta	42	1ii2	A	02	Alpha Beta	Mainly Beta
43	3n52	D	00	Alpha Beta	Mainly Beta	44	1h6w	A	02	Few Secondary Structures	Mainly Beta
45	1ok8	A	03	Mainly Beta	Alpha Beta	46	1xja	C	00	Alpha Beta	Mainly Beta
47	1y9e	F	01	Mainly Beta	Alpha Beta	48	1b2s	E	00	Mainly Alpha	Alpha Beta
49	1zh0	A	01	Mainly Alpha	Alpha Beta	50	1htm	F	00	Mainly Alpha	Alpha Beta
51	1n2m	F	00	Mainly Beta	Alpha Beta	52	1nag	A	00	Alpha Beta	Few Secondary Structures
53	3g7k	C	00	Mainly Beta	Alpha Beta	54	2p4p	A	00	Mainly Beta	Alpha Beta
55	1mry	A	02	Mainly Beta	Alpha Beta	56	1qhh	A	00	Mainly Alpha	Alpha Beta
57	1fyk	A	00	Alpha Beta	Mainly Alpha	58	1ezl	A	00	Alpha Beta	Mainly Beta

59	3a60	B	01	Mainly Beta	Alpha Beta	60	1pfn	D	00	Alpha Beta	Mainly Beta
61	1zki	A	00	Mainly Beta	Alpha Beta	62	1yox	C	00	Mainly Beta	Alpha Beta
63	1nkq	B	00	Mainly Beta	Alpha Beta	64	1oc2	A	02	Mainly Alpha	Alpha Beta
65	1ouz	A	00	Alpha Beta	Few Secondary Structures	66	1phj	A	01	Alpha Beta	Mainly Beta
67	1bor	A	00	Mainly Beta	Alpha Beta	68	1n0q	A	00	Alpha Beta	Mainly Alpha
69	2czc	A	01	Mainly Alpha	Alpha Beta	70	1fjf	J	00	Mainly Beta	Alpha Beta
71	2z6k	B	00	Alpha Beta	Mainly Beta	72	1qzh	D	00	Alpha Beta	Mainly Beta
73	1dm0	A	02	Alpha Beta	Few Secondary Structures	74	1le9	B	01	Alpha Beta	Mainly Beta
75	1nyo	A	00	Alpha Beta	Mainly Beta	76	1dkm	A	02	Mainly Alpha	Alpha Beta
77	1y7k	A	00	Mainly Beta	Few Secondary Structures	78	1dd4	B	02	Mainly Alpha	Alpha Beta
79	1v17	A	00	Alpha Beta	Mainly Beta	80	1sl1	A	02	Alpha Beta	Mainly Alpha
81	1doy	A	00	Mainly Beta	Alpha Beta	82	2js3	B	01	Alpha Beta	Mainly Alpha
83	1ckn	B	03	Mainly Alpha	Few Secondary Structures	84	3htu	E	00	Alpha Beta	Mainly Alpha
85	1hfe	S	00	Mainly Alpha	Few Secondary Structures	86	1b6e	A	00	Mainly Beta	Alpha Beta
87	2hgh	A	03	Mainly Alpha	Alpha Beta	88	3ku0	B	02	Alpha Beta	Few Secondary Structures
89	1osp	O	02	Mainly Beta	Alpha Beta	90	1gvj	A	00	Alpha Beta	Mainly Alpha
91	3myr	E	02	Alpha Beta	Few Secondary Structures	92	1mk7	B	02	Alpha Beta	Mainly Beta
93	1qu1	F	00	Mainly Alpha	Alpha Beta	94	2ffd	B	02	Mainly Beta	Alpha Beta
95	1df9	B	01	Alpha Beta	Mainly Beta	96	1gg2	G	00	Mainly Alpha	Few Secondary Structures
97	1v3u	A	01	Mainly Beta	Alpha Beta	98	1ewf	A	02	Mainly Beta	Alpha Beta
99	3alc	A	00	Alpha Beta	Few Secondary Structures	100	2jyy	A	00	Mainly Beta	Alpha Beta
101	3e0r	D	01	Mainly Beta	Alpha Beta	102	1u2r	A	03	Mainly Alpha	Alpha Beta
103	2qsd	E	01	Mainly Beta	Alpha Beta	104	1yst	H	01	Alpha Beta	Few Secondary Structures
105	2x5u	H	02	Mainly Beta	Alpha Beta	106	3mip	B	00	Mainly Alpha	Alpha Beta
107	1iy0	A	01	Mainly Alpha	Alpha Beta	108	1si3	A	00	Alpha Beta	Mainly Beta
109	1flo	C	01	Mainly Alpha	Alpha Beta	110	3gku	B	03	Mainly Alpha	Alpha Beta
111	2pah	A	00	Alpha Beta	Mainly Alpha	112	1i6x	A	02	Alpha Beta	Mainly Alpha
113	2hzt	B	00	Alpha Beta	Mainly Alpha	114	1btv	A	00	Mainly Beta	Alpha Beta
115	1ba3	A	05	Alpha Beta	Few Secondary Structures	116	1seg	A	00	Mainly Beta	Alpha Beta
117	3kaj	A	02	Mainly	Alpha Beta	118	1e50	D	00	Alpha Beta	Mainly

119	1yf8	A	02	Beta Alpha Beta	Few Secondary Structures	120	2h2r	B	01	Mainly Beta	Beta Alpha Beta
121	1l6o	C	00	Alpha Beta	Mainly Beta	122	2wem	D	01	Mainly Alpha	Alpha Beta
123	2pn0	D	02	Mainly Beta	Alpha Beta	124	2aw4	V	00	Alpha Beta	Mainly Beta
125	1o54	A	01	Mainly Beta	Alpha Beta	126	1sap	A	00	Alpha Beta	Mainly Beta
127	2vhm	V	00	Alpha Beta	Mainly Beta	128	1i9g	A	01	Mainly Beta	Alpha Beta
129	1t8c	A	01	Mainly Beta	Alpha Beta	130	1fxi	B	00	Mainly Beta	Alpha Beta
131	2dzo	A	00	Alpha Beta	Mainly Alpha	132	3lw5	N	00	Mainly Alpha	Few Secondary Structures
133	3ckg	A	02	Mainly Beta	Alpha Beta	134	3kal	B	02	Mainly Beta	Alpha Beta
135	1vg0	A	02	Mainly Beta	Alpha Beta	136	1k8h	A	00	Alpha Beta	Mainly Beta
137	1scm	A	00	Mainly Alpha	Few Secondary Structures	138	1nt4	B	02	Mainly Alpha	Alpha Beta
139	3s52	D	00	Mainly Beta	Alpha Beta	140	2wl9	D	02	Mainly Beta	Alpha Beta
141	2i1p	A	00	Alpha Beta	Few Secondary Structures	142	1qu1	C	00	Mainly Alpha	Alpha Beta
143	1ixl	A	00	Mainly Beta	Alpha Beta	144	1cac	0	00	Mainly Beta	Alpha Beta
145	1omw	G	00	Mainly Alpha	Few Secondary Structures	146	1o67	C	00	Alpha Beta	Mainly Beta
147	1llm	D	02	Mainly Alpha	Alpha Beta	148	2onl	C	03	Mainly Alpha	Few Secondary Structures
149	1ldt	L	00	Mainly Beta	Alpha Beta	150	1bxk	A	02	Mainly Alpha	Alpha Beta
151	1jlr	A	02	Alpha Beta	Few Secondary Structures	152	3obg	A	02	Alpha Beta	Mainly Alpha
153	3o17	B	02	Alpha Beta	Mainly Alpha	154	2fum	C	02	Alpha Beta	Mainly Alpha
155	2dn5	A	00	Mainly Alpha	Alpha Beta	156	3b9e	A	03	Mainly Beta	Alpha Beta
157	1jnd	A	02	Mainly Beta	Alpha Beta	158	3ars	A	03	Mainly Beta	Alpha Beta
159	2ijr	A	01	Mainly Beta	Alpha Beta	160	1rmj	A	01	Alpha Beta	Few Secondary Structures
161	1ukl	C	00	Mainly Alpha	Few Secondary Structures	162	2ql2	C	00	Mainly Alpha	Few Secondary Structures
163	1mdy	A	00	Mainly Alpha	Few Secondary Structures	164	1ukl	F	00	Mainly Alpha	Few Secondary Structures
165	1am9	C	00	Mainly Alpha	Few Secondary Structures	166	1nkp	D	00	Mainly Alpha	Few Secondary Structures
167	1ng5	A	00	Alpha Beta	Mainly Beta	168	1fjj	A	00	Mainly Beta	Alpha Beta
169	1vi3	A	00	Mainly Beta	Alpha Beta	170	1wkp	A	00	Mainly Beta	Alpha Beta
171	2ig7	B	02	Mainly Alpha	Alpha Beta	172	3qfu	A	04	Mainly Alpha	Alpha Beta
173	3ldp	B	04	Mainly Alpha	Alpha Beta	174	2p9u	A	04	Mainly Alpha	Alpha Beta
175	3d2f	D	04	Mainly Alpha	Alpha Beta	176	3qfp	A	04	Mainly Alpha	Alpha Beta

177	3qml	A	04	Mainly Alpha	Alpha Beta	178	3cjb	A	04	Mainly Alpha	Alpha Beta
179	3iuc	A	04	Mainly Alpha	Alpha Beta	180	1q18	A	02	Mainly Alpha	Alpha Beta
181	1a3u	A	00	Alpha Beta	Mainly Beta	182	2eym	A	00	Alpha Beta	Mainly Beta
183	1joq	A	00	Alpha Beta	Mainly Beta	184	2fim	B	00	Mainly Beta	Alpha Beta
185	3ck2	A	00	Mainly Beta	Alpha Beta	186	1jl5	A	00	Mainly Beta	Alpha Beta
187	1der	H	02	Mainly Alpha	Alpha Beta	188	2qts	B	04	Alpha Beta	Mainly Alpha
189	1zxi	E	01	Mainly Beta	Alpha Beta	190	1vq7	T	00	Alpha Beta	Mainly Beta
191	1cc1	L	00	Alpha Beta	Mainly Alpha	192	2prg	C	00	Mainly Alpha	Few Secondary Structures
193	3cf5	E	02	Mainly Beta	Alpha Beta	194	1lnh	A	02	Alpha Beta	Few Secondary Structures
195	1jmq	A	02	Alpha Beta	Few Secondary Structures	196	1chp	F	00	Alpha Beta	Mainly Beta
197	3hip	C	00	Mainly Beta	Few Secondary Structures	198	1isu	B	00	Alpha Beta	Few Secondary Structures
199	1t0y	A	00	Mainly Beta	Alpha Beta	200	1iva	A	00	Mainly Beta	Few Secondary Structures
201	1omb	A	00	Mainly Beta	Few Secondary Structures	202	1oma	A	00	Mainly Beta	Few Secondary Structures
203	1fkr	A	00	Mainly Beta	Alpha Beta	204	1kt1	A	02	Mainly Beta	Alpha Beta
205	1fks	A	00	Mainly Beta	Alpha Beta	206	1kt1	A	01	Mainly Beta	Alpha Beta
207	1rou	A	00	Mainly Beta	Alpha Beta	208	1hxv	A	00	Mainly Beta	Alpha Beta
209	1w26	A	03	Mainly Beta	Alpha Beta	210	1w26	B	03	Mainly Beta	Alpha Beta
211	1puq	A	00	Mainly Beta	Alpha Beta	212	1rdr	A	01	Mainly Alpha	Few Secondary Structures
213	1qcs	A	02	Mainly Beta	Alpha Beta	214	3r8o	R	00	Mainly Alpha	Few Secondary Structures
215	3ora	R	00	Mainly Alpha	Few Secondary Structures	216	3otf	A	02	Alpha Beta	Mainly Beta
217	1cx4	A	02	Alpha Beta	Mainly Beta	218	1pmi	A	01	Alpha Beta	Mainly Beta
219	2psp	A	02	Alpha Beta	Few Secondary Structures	220	1h31	F	00	Alpha Beta	Mainly Alpha
221	1zko	A	00	Alpha Beta	Mainly Beta						

Table S2B. Domain classification conflicts between CATH and automatic classification of protein fold space using Class and Architecture criteria.

Index	PDB ID	Chain ID	Domain ID	Predicted class	Annotated class	Index	PDB ID	Chain ID	Domain ID	Predicted class	Annotated class
1	1u0l	C	01	Roll	Beta Barrel	2	2z9i	B	01	2-Layer Sandwich	Beta Barrel
3	1gos	A	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	4	1nvo	A	00	Orthogonal Bundle	Up-down Bundle
5	3coj	E	01	2-Layer Sandwich	3-Layer(aba) Sandwich	6	1xmq	Q	00	Roll	Beta Barrel
7	1e3m	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich	8	2f4z	A	00	2-Layer Sandwich	Roll
9	1mkf	B	02	Beta Barrel	Sandwich	10	3h5v	C	01	Orthogonal Bundle	3-Layer(aba) Sandwich
11	3srx	B	00	3-Layer(aba) Sandwich	4-Layer Sandwich	12	2hma	A	01	Orthogonal Bundle	3-Layer(aba) Sandwich
13	1iqp	C	03	Orthogonal Bundle	Up-down Bundle	14	1o64	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
15	3his	A	02	Orthogonal Bundle	Irregular	16	3eyc	C	00	Sandwich	Beta Barrel
17	2ahe	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich	18	2iee	B	01	2-Layer Sandwich	3-Layer(aba) Sandwich
19	1nti	A	00	Orthogonal Bundle	Up-down Bundle	20	2gdr	C	01	2-Layer Sandwich	3-Layer(aba) Sandwich
21	1gk3	A	01	3-Layer(aba) Sandwich	Orthogonal Bundle	22	2vpf	G	00	Beta Barrel	Ribbon
23	1d8y	A	02	Orthogonal Bundle	Up-down Bundle	24	1sir	A	02	Sandwich	Beta Barrel
25	3k6l	C	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	26	1vma	A	01	Orthogonal Bundle	Up-down Bundle
27	2ip4	A	04	2-Layer Sandwich	Alpha-Beta Complex	28	3i8h	Q	00	2-Layer Sandwich	Irregular
29	2hrv	A	01	Ribbon	Beta Barrel	30	1xpa	A	00	2-Layer Sandwich	Alpha-Beta Complex
31	1tha	B	00	Beta Barrel	Sandwich	32	1qf6	A	04	3-Layer(aba) Sandwich	2-Layer Sandwich
33	1xxa	D	00	Roll	2-Layer Sandwich	34	1cpo	A	00	3-Layer(aba) Sandwich	Orthogonal Bundle
35	2ojl	A	00	2-Layer Sandwich	3-Layer(aba) Sandwich	36	1hxi	A	00	Up-down Bundle	Alpha Horseshoe
37	1qyd	C	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	38	1syo	B	03	Sandwich	Distorted Sandwich
39	119y	A	00	Alpha-Beta Complex	4-Layer Sandwich	40	3pr8	C	01	2-Layer Sandwich	3-Layer(aba) Sandwich
41	1lxa	A	02	Orthogonal Bundle	Up-down Bundle	42	1q1w	B	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
43	1bnd	A	00	Sandwich	Ribbon	44	3pgj	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich
45	3ncw	B	02	2-Layer Sandwich	Roll	46	2wda	A	01	Orthogonal Bundle	Alpha/alpha barrel
47	1qfj	C	01	2-Layer Sandwich	Beta Barrel	48	3cjp	B	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel
49	1bgw	A	04	Roll	2-Layer Sandwich	50	1q1k	A	02	Roll	3-Layer(aba) Sandwich
51	2j3n	C	02	2-Layer Sandwich	3-Layer(bba) Sandwich	52	1bbg	A	00	Ribbon	2-Layer Sandwich
53	1df9	B	02	Sandwich	Beta Barrel	54	1wj0	A	00	Single Sheet	Irregular
55	3gbm	B	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	56	2w9p	N	00	2-Layer Sandwich	Roll
57	1n7g	C	01	Orthogonal Bundle	Alpha-Beta Complex	58	3cgv	A	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
59	1lnz	A	01	Sandwich	Distorted Sandwich	60	1ox8	A	00	Roll	Roll
61	1i10	A	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	62	1q15	C	01	2-Layer Sandwich	4-Layer Sandwich
63	1z1d	B	00	2-Layer Sandwich	3-Layer(aba) Sandwich	64	2pma	A	01	Sandwich	Beta Barrel
65	3ktz	A	02	2-Layer Sandwich	Irregular	66	2cai	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich
67	1ng3	A	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	68	2pgg	A	01	3-Layer(aba) Sandwich	Alpha-Beta Complex
69	3bwg	B	02	Roll	3-Layer(aba) Sandwich	70	2qsx	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
71	2r8c	B	01	Sandwich	Roll	72	3bim	G	00	Orthogonal Bundle	2-Layer Sandwich

73	1uuu	A	00	2-Layer Sandwich	Irregular	74	1gn1	A	00	2-Layer Sandwich	3-Layer(bba) Sandwich
75	1bye	C	01	Orthogonal Bundle	3-Layer(aba) Sandwich	76	1yov	D	03	Orthogonal Bundle	Roll
77	3bnm	B	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	78	3khj	E	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel
79	1h83	C	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	80	1mk4	A	00	Roll	3-Layer(aba) Sandwich
81	2b5c	0	00	Orthogonal Bundle	Roll	82	1ep1	B	01	Sandwich	Beta Barrel
83	1mft	B	00	Orthogonal Bundle	Up-down Bundle	84	1ryi	B	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
85	2r7j	A	01	Roll	Alpha-Beta Complex	86	1oao	A	03	Orthogonal Bundle	Up-down Bundle
87	1o63	B	01	2-Layer Sandwich	3-Layer(aba) Sandwich	88	2p50	B	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel
89	1h84	A	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	90	1iz1	P	02	2-Layer Sandwich	3-Layer(aba) Sandwich
91	1nr4	F	00	2-Layer Sandwich	Beta Barrel	92	1d0d	A	00	2-Layer Sandwich	Irregular
93	2ogj	B	01	Sandwich	Roll	94	3dex	E	00	2-Layer Sandwich	3-Layer(aba) Sandwich
95	1v76	A	00	Sandwich	Roll	96	1mq5	A	04	Orthogonal Bundle	Alpha Horseshoe
97	1d0n	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich	98	3ej1	D	00	3-Layer(aba) Sandwich	2-Layer Sandwich
99	1pwu	B	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	100	1q25	A	01	Beta Barrel	Distorted Sandwich
101	3ckd	A	03	Orthogonal Bundle	Up-down Bundle	102	1n7g	D	01	Orthogonal Bundle	Alpha-Beta Complex
103	3h32	E	03	2-Layer Sandwich	Irregular	104	1p4r	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
105	1t2f	B	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	106	1xmq	T	00	Orthogonal Bundle	Up-down Bundle
107	1pdo	A	00	Orthogonal Bundle	3-Layer(aba) Sandwich	108	3fdi	B	00	Orthogonal Bundle	3-Layer(aba) Sandwich
109	1twy	C	01	2-Layer Sandwich	3-Layer(aba) Sandwich	110	1ylm	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
111	1l3p	A	00	Orthogonal Bundle	Up-down Bundle	112	1hwo	A	02	2-Layer Sandwich	Irregular
113	1sr8	A	03	2-Layer Sandwich	3-Layer(aba) Sandwich	114	1xs6	B	00	Sandwich	Distorted Sandwich
115	1zud	4	00	2-Layer Sandwich	Roll	116	2dpn	A	01	3-Layer(aba) Sandwich	2-Layer Sandwich
117	1u25	B	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	118	2zhx	N	00	2-Layer Sandwich	Roll
119	1uxk	A	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	120	2pq5	B	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
121	3oto	N	00	2-Layer Sandwich	Irregular	122	2b49	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
123	1i74	A	01	3-Layer(aba) Sandwich	Alpha-Beta Complex	124	1o6t	A	02	Roll	Sandwich
125	2eq9	J	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	126	3fi8	A	01	2-Layer Sandwich	Roll
127	3gku	A	03	Orthogonal Bundle	2-Layer Sandwich	128	3rhy	B	00	3-Layer(aba) Sandwich	5-stranded Propeller
129	1f3v	A	00	3-Layer(aba) Sandwich	2-Layer Sandwich	130	2p1q	A	00	Orthogonal Bundle	2-Layer Sandwich
131	2alc	A	00	Orthogonal Bundle	Irregular	132	2ozg	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich
133	1uk0	B	01	Orthogonal Bundle	Up-down Bundle	134	2fe7	B	00	2-Layer Sandwich	3-Layer(aba) Sandwich
135	3huw	T	00	Orthogonal Bundle	Up-down Bundle	136	3fy5	A	00	2-Layer Sandwich	Roll
137	1bu6	Z	02	3-Layer(aba) Sandwich	2-Layer Sandwich	138	1gqq	A	03	3-Layer(aba) Sandwich	Alpha-Beta Complex
139	2dzo	C	00	3-Layer(aba) Sandwich	Alpha Horseshoe	140	3ov7	A	02	3-Layer(aba) Sandwich	2-Layer Sandwich
141	2qmx	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich	142	2qez	B	03	3-Layer(aba) Sandwich	Alpha-Beta Barrel
143	1vs0	B	02	Roll	2-Layer Sandwich	144	1zov	A	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
145	3i99	A	03	2-Layer Sandwich	Alpha-Beta Complex	146	2fb2	A	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel
147	2enx	B	01	3-Layer(aba) Sandwich	Alpha-Beta Complex	148	2ggg	A	01	2-Layer Sandwich	Roll
149	3dy0	A	01	3-Layer(aba) Sandwich	2-Layer	150	2q85	A	03	3-Layer(aba) Sandwich	Alpha-Beta

151	3aad	A	01	Sandwich Orthogonal Bundle	Sandwich Up-down Bundle	152	3i8h	W	00	Sandwich Orthogonal Bundle	Complex Up-down Bundle
153	2pq5	C	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	154	3p0w	C	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel
155	1aoc	A	00	Alpha-Beta Complex	Ribbon	156	1jw2	A	00	Orthogonal Bundle	Up-down Bundle
157	1e61	A	03	2-Layer Sandwich	Alpha-Beta Complex	158	1fc3	C	00	Up-down Bundle	Orthogonal Bundle
159	2i3u	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	160	1f56	A	00	Beta Barrel	Sandwich
161	2hzg	B	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel	162	1fpz	E	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
163	2haw	A	01	3-Layer(aba) Sandwich	Alpha-Beta Complex	164	2yvk	B	01	Orthogonal Bundle	Up-down Bundle
165	1ycf	D	01	3-Layer(aba) Sandwich	4-Layer Sandwich	166	1lvm	B	01	Sandwich	Beta Barrel
167	2i06	A	01	2-Layer Sandwich	3-Layer(bba) Sandwich	168	2vnt	B	01	Sandwich	Beta Barrel
169	1e91	A	00	Orthogonal Bundle	Up-down Bundle	170	2pw9	B	02	2-Layer Sandwich	Roll
171	1k0n	B	01	2-Layer Sandwich	3-Layer(aba) Sandwich	172	1e68	A	00	Orthogonal Bundle	Up-down Bundle
173	3cme	R	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	174	1sd7	A	01	2-Layer Sandwich	Orthogonal Bundle
175	1c96	A	03	3-Layer(aba) Sandwich	2-Layer Sandwich	176	3tmk	B	00	Orthogonal Bundle	3-Layer(aba) Sandwich
177	1k9a	E	01	Beta Barrel	Roll	178	2fuh	A	00	Orthogonal Bundle	Roll
179	1wnr	D	00	Beta Barrel	Roll	180	2zyz	D	02	2-Layer Sandwich	3-Layer(aba) Sandwich
181	1niq	C	00	2-Layer Sandwich	Roll	182	1dee	H	00	Orthogonal Bundle	Up-down Bundle
183	1u3a	E	00	Orthogonal Bundle	3-Layer(aba) Sandwich	184	1qlp	A	02	3-Layer(aba) Sandwich	2-Layer Sandwich
185	2dfy	X	02	Roll	Ribbon	186	3mg4	T	00	3-Layer(aba) Sandwich	4-Layer Sandwich
187	1s16	B	02	3-Layer(aba) Sandwich	2-Layer Sandwich	188	1gml	A	00	2-Layer Sandwich	3-Layer(bba) Sandwich
189	2eo0	B	00	2-Layer Sandwich	3-Layer(aba) Sandwich	190	1ka1	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
191	1oc2	A	02	Orthogonal Bundle	Alpha-Beta Complex	192	1slt	B	04	Orthogonal Bundle	2-Layer Sandwich
193	2bbv	D	00	Up-down Bundle	Irregular	194	2i0o	A	00	3-Layer(aba) Sandwich	4-Layer Sandwich
195	1omx	B	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	196	2wri	Y	02	Sandwich	Beta Barrel
197	2zh6	A	02	3-Layer(aba) Sandwich	2-Layer Sandwich	198	1tu3	G	00	Orthogonal Bundle	Up-down Bundle
199	2yvw	A	02	3-Layer(aba) Sandwich	Alpha-beta prism	200	1sw2	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
201	2kse	A	00	Orthogonal Bundle	Up-down Bundle	202	1ui8	A	01	2-Layer Sandwich	Roll
203	2gzw	C	02	2-Layer Sandwich	Orthogonal Bundle	204	3d8x	A	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
205	3i04	M	04	2-Layer Sandwich	3-Layer(aba) Sandwich	206	1n0l	B	00	Beta Barrel	Sandwich
207	2vdy	B	01	3-Layer(aba) Sandwich	2-Layer Sandwich	208	1dd9	A	01	2-Layer Sandwich	Alpha-Beta Complex
209	2gax	B	00	2-Layer Sandwich	3-Layer(aba) Sandwich	210	1okj	D	02	Roll	2-Layer Sandwich
211	1sd4	A	01	2-Layer Sandwich	Orthogonal Bundle	212	1lox	A	01	Orthogonal Bundle	Up-down Bundle
213	1yq4	D	00	Orthogonal Bundle	Up-down Bundle	214	1s4b	P	01	Orthogonal Bundle	Up-down Bundle
215	3rhi	B	00	2-Layer Sandwich	Irregular	216	2b9b	C	04	Up-down Bundle	2-Layer Sandwich
217	1vcn	A	01	Alpha-Beta Barrel	3-Layer(aba) Sandwich	218	1wtu	A	00	2-Layer Sandwich	Irregular
219	3lxz	C	01	2-Layer Sandwich	3-Layer(aba) Sandwich	220	2eq7	A	02	2-Layer Sandwich	3-Layer(bba) Sandwich
221	1f5m	A	00	Orthogonal Bundle	2-Layer Sandwich	222	1n9y	D	00	Sandwich	Beta Barrel
223	1k35	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich	224	1der	J	03	2-Layer Sandwich	3-Layer(bba) Sandwich
225	3cny	B	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel	226	1umb	C	00	Alpha-Beta Barrel	3-Layer(aba) Sandwich

227	liqp	F	03	Orthogonal Bundle	Up-down Bundle	228	2r0b	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
229	likz	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	230	2ob2	C	00	Alpha-Beta Barrel	3-Layer(aba) Sandwich
231	3omg	B	02	2-Layer Sandwich	Roll	232	1r8e	A	01	Sandwich	Alpha-Beta Barrel
233	3bpo	A	00	Orthogonal Bundle	Up-down Bundle	234	3fd3	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
235	1kfd	A	02	Orthogonal Bundle	Up-down Bundle	236	1avx	B	00	Sandwich	Trefoil
237	1ryi	C	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	238	1fav	A	00	Up-down Bundle	Orthogonal Bundle
239	3knm	N	00	2-Layer Sandwich	Alpha-Beta Complex	240	1h0t	B	00	Orthogonal Bundle	Up-down Bundle
241	3gzx	A	01	2-Layer Sandwich	Alpha-Beta Complex	242	1ysa	D	00	Up-down Bundle	2-Layer Sandwich
243	2d13	D	02	2-Layer Sandwich	Alpha-Beta Complex	244	1gv4	B	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
245	3nva	A	01	Alpha-Beta Barrel	3-Layer(aba) Sandwich	246	2kyg	B	00	Orthogonal Bundle	Up-down Bundle
247	1vs6	S	00	2-Layer Sandwich	Alpha-Beta Complex	248	3lok	A	01	Beta Barrel	2-Layer Sandwich
249	3ku0	B	02	2-Layer Sandwich	Irregular	250	2wh4	Z	01	2-Layer Sandwich	Beta Barrel
251	1vz8	B	01	3-Layer(aba) Sandwich	4-Layer Sandwich	252	3oar	T	00	Orthogonal Bundle	Up-down Bundle
253	1uk4	B	02	2-Layer Sandwich	Beta Barrel	254	1xpv	A	00	2-Layer Sandwich	3-Layer(aba) Sandwich
255	1j0s	A	00	Beta Barrel	Trefoil	256	2r82	A	02	Orthogonal Bundle	Up-down Bundle
257	1v5q	A	00	Roll	Roll	258	1y47	B	00	Orthogonal Bundle	Up-down Bundle
259	1dxl	B	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	260	2hwo	A	01	Roll	2-Layer Sandwich
261	1eaf	A	00	3-Layer(aba) Sandwich	2-Layer Sandwich	262	1nuf	A	02	3-Layer(aba) Sandwich	Alpha-Beta Complex
263	1oao	C	04	2-Layer Sandwich	3-Layer(aba) Sandwich	264	2c8u	B	01	2-Layer Sandwich	3-Layer(aba) Sandwich
265	2o7m	A	00	Orthogonal Bundle	Roll	266	3oeh	T	01	Roll	Beta Barrel
267	1nb3	J	00	2-Layer Sandwich	Roll	268	1qa7	D	01	Roll	Beta Barrel
269	3dcg	D	00	Orthogonal Bundle	2-Layer Sandwich	270	1xgn	A	02	2-Layer Sandwich	Orthogonal Bundle
271	3b7j	D	00	2-Layer Sandwich	Roll	272	2hz7	A	02	2-Layer Sandwich	Alpha-Beta Complex
273	1fs1	B	00	Orthogonal Bundle	2-Layer Sandwich	274	3iw4	B	01	Roll	2-Layer Sandwich
275	2xur	B	01	2-Layer Sandwich	Roll	276	3dgb	A	01	3-Layer(aba) Sandwich	2-Layer Sandwich
277	2qgq	B	01	3-Layer(aba) Sandwich	Alpha-Beta Horseshoe	278	1r0v	B	01	2-Layer Sandwich	3-Layer(aba) Sandwich
279	1f2l	C	00	2-Layer Sandwich	Beta Barrel	280	3ae7	D	00	Orthogonal Bundle	Up-down Bundle
281	1ay9	B	00	Sandwich	Ribbon	282	1te7	A	00	3-Layer(aba) Sandwich	Roll
283	1lln	A	02	2-Layer Sandwich	Irregular	284	1h6o	A	00	Up-down Bundle	Alpha Horseshoe
285	1bbn	A	00	Orthogonal Bundle	Up-down Bundle	286	1adv	A	02	2-Layer Sandwich	Alpha-Beta Complex
287	1qxh	A	00	Sandwich	3-Layer(aba) Sandwich	288	3r8n	M	02	Orthogonal Bundle	Irregular
289	1rxt	B	01	2-Layer Sandwich	3-Layer(aba) Sandwich	290	2ou6	A	00	Orthogonal Bundle	Up-down Bundle
291	2kh2	A	00	Sandwich	Trefoil	292	2nmt	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich
293	1he8	A	03	Up-down Bundle	Alpha Horseshoe	294	1jbr	A	00	Sandwich	Roll
295	1xcq	N	00	2-Layer Sandwich	Roll	296	1dko	A	02	Orthogonal Bundle	3-Layer(aba) Sandwich
297	1qu1	B	00	Orthogonal Bundle	Alpha-Beta Complex	298	2px7	B	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
299	2wcw	B	00	2-Layer Sandwich	3-Layer(aba) Sandwich	300	2hkt	A	00	Orthogonal Bundle	Up-down Bundle
301	3rk1	A	02	2-Layer Sandwich	Alpha-Beta Complex	302	1lxt	A	04	Roll	2-Layer Sandwich
303	1b3n	A	02	2-Layer	3-Layer(aba)	304	1shy	A	01	Sandwich	Beta Barrel

305	2ems	A	02	Sandwich Orthogonal Bundle	Sandwich Up-down Bundle	306	1s4b	P	03	2-Layer Sandwich	3-Layer(aba) Sandwich
307	3cpx	A	02	Roll	Beta Barrel	308	1t4z	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich
309	1kwk	A	03	Ribbon	Sandwich	310	2ogj	D	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel
311	3k43	B	00	Sandwich	Distorted Sandwich	312	1yre	D	00	2-Layer Sandwich	3-Layer(aba) Sandwich
313	3ofq	X	00	2-Layer Sandwich	Roll	314	2tpr	B	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
315	1vs8	S	00	2-Layer Sandwich	Alpha-Beta Complex	316	1gka	B	00	Sandwich	Beta Barrel
317	1egr	A	00	2-Layer Sandwich	3-Layer(aba) Sandwich	318	2zm6	N	00	Orthogonal Bundle	Irregular
319	2elc	D	01	Orthogonal Bundle	Up-down Bundle	320	1gqj	A	01	3-Layer(aba) Sandwich	2-Layer Sandwich
321	2xy5	A	02	Orthogonal Bundle	Up-down Bundle	322	2cyk	A	00	Orthogonal Bundle	Up-down Bundle
323	1ii4	B	00	Sandwich	Trefoil	324	1djn	A	03	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
325	1oao	B	03	Orthogonal Bundle	Up-down Bundle	326	3k86	B	00	Roll	Roll
327	1sh8	B	00	3-Layer(aba) Sandwich	Roll	328	1glc	G	02	3-Layer(aba) Sandwich	2-Layer Sandwich
329	1gts	A	04	2-Layer Sandwich	Alpha-Beta Complex	330	3ecy	A	00	Sandwich	Distorted Sandwich
331	1h6z	A	04	3-Layer(aba) Sandwich	Alpha-Beta Barrel	332	1jro	D	03	3-Layer(aba) Sandwich	2-Layer Sandwich
333	3gk6	A	00	Sandwich	Alpha-Beta Barrel	334	2d2z	A	01	2-Layer Sandwich	3-Layer(aba) Sandwich
335	1g9y	B	00	3-Layer(aba) Sandwich	Roll	336	2d26	A	01	3-Layer(aba) Sandwich	2-Layer Sandwich
337	1vz6	B	01	3-Layer(aba) Sandwich	4-Layer Sandwich	338	2iee	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
339	1t7s	A	00	Orthogonal Bundle	Up-down Bundle	340	3if9	A	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
341	1f7o	C	00	Sandwich	Distorted Sandwich	342	1t6p	H	02	Orthogonal Bundle	Up-down Bundle
343	3h39	B	01	3-Layer(aba) Sandwich	2-Layer Sandwich	344	2o8r	B	02	3-Layer(aba) Sandwich	2-Layer Sandwich
345	3cqy	B	01	3-Layer(aba) Sandwich	2-Layer Sandwich	346	1bt2	A	00	3-Layer(aba) Sandwich	Orthogonal Bundle
347	2f4d	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	348	2rfr	A	00	2-Layer Sandwich	Roll
349	3k75	E	03	Orthogonal Bundle	2-Layer Sandwich	350	3cd6	W	01	Roll	2-Layer Sandwich
351	3fgc	B	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel	352	2o30	A	00	Beta Barrel	Sandwich
353	1yq3	D	00	Orthogonal Bundle	Up-down Bundle	354	3knp	F	00	2-Layer Sandwich	3-Layer(bba) Sandwich
355	1vlw	A	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel	356	2zzz	C	01	Orthogonal Bundle	3-Layer(aba) Sandwich
357	1kie	B	00	Alpha-Beta Barrel	Alpha-Beta Complex	358	1sft	B	02	Sandwich	Beta Barrel
359	1mok	D	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	360	3b4v	E	00	2-Layer Sandwich	Ribbon
361	1r76	A	00	Alpha-Beta Barrel	Alpha/alpha barrel	362	1qoj	A	00	Orthogonal Bundle	Irregular
363	2rhs	D	02	Sandwich	Beta Barrel	364	1i8n	A	00	Beta Barrel	3-Layer(bba) Sandwich
365	3cet	A	02	3-Layer(aba) Sandwich	2-Layer Sandwich	366	3ggv	A	00	Sandwich	Beta Barrel
367	3llp	A	04	Sandwich	Trefoil	368	1ngl	A	01	Orthogonal Bundle	Up-down Bundle
369	3qfb	A	02	2-Layer Sandwich	3-Layer(bba) Sandwich	370	1ayz	A	00	2-Layer Sandwich	Roll
371	2wym	A	00	3-Layer(aba) Sandwich	4-Layer Sandwich	372	2dsj	B	01	Orthogonal Bundle	Up-down Bundle
373	2dw6	D	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel	374	2pn0	A	02	Roll	Roll
375	2qmh	I	00	2-Layer Sandwich	3-Layer(aba) Sandwich	376	1djg	A	03	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
377	1eeh	A	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	378	1iy6	A	00	Roll	2-Layer Sandwich
379	2rhs	B	01	Orthogonal Bundle	2-Layer Sandwich	380	1e01	A	00	Ribbon	Roll

381	1aqt	A	01	Beta Barrel	Sandwich	382	2ush	A	02	2-Layer Sandwich	Alpha-Beta Complex Sandwich
383	1ngv	D	03	Irregular	Roll	384	2p2e	A	00	Roll	Sandwich
385	3jrl	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	386	1ryi	D	01	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich
387	3i5r	A	00	Sandwich	Roll	388	1sgk	A	01	2-Layer Sandwich	Alpha-Beta Complex
389	1mss	B	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel	390	3mpp	G	02	Sandwich	Trefoil
391	1w7p	B	01	3-Layer(aba) Sandwich	Orthogonal Bundle	392	2d1x	C	00	Beta Barrel	Roll
393	2h6r	F	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel	394	1dee	G	00	Orthogonal Bundle	Up-down Bundle
395	1e1k	A	02	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	396	1qoj	B	00	Orthogonal Bundle	Irregular
397	1edl	A	00	Orthogonal Bundle	Up-down Bundle	398	2zr1	D	01	Sandwich	Trefoil
399	1qpp	C	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel	400	2pyy	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich
401	2tpt	A	03	2-Layer Sandwich	Alpha-Beta Complex	402	1zsv	B	01	Beta Barrel	Alpha-Beta Complex
403	3aeg	A	03	Orthogonal Bundle	Up-down Bundle	404	1g3i	T	02	Up-down Bundle	Orthogonal Bundle
405	3oo2	A	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel	406	1fch	A	00	Orthogonal Bundle	Alpha Horseshoe
407	1llz	A	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel	408	1poi	C	02	Roll	2-Layer Sandwich
409	3gkx	B	00	Orthogonal Bundle	3-Layer(aba) Sandwich	410	1ka8	B	00	2-Layer Sandwich	Orthogonal Bundle
411	1qd1	B	01	3-Layer(aba) Sandwich	2-Layer Sandwich	412	2og9	A	02	3-Layer(aba) Sandwich	Alpha-Beta Barrel
413	1y6d	A	00	Orthogonal Bundle	Up-down Bundle	414	2oji	A	02	3-Layer(aba) Sandwich	Orthogonal Bundle
415	2ghg	A	02	3-Layer(aba) Sandwich	Orthogonal Bundle	416	2oyq	A	02	3-Layer(aba) Sandwich	2-Layer Sandwich
417	1bcm	A	01	3-Layer(aba) Sandwich	2-Layer Sandwich	418	2hnd	A	05	3-Layer(aba) Sandwich	2-Layer Sandwich
419	1cxq	A	00	3-Layer(aba) Sandwich	2-Layer Sandwich	420	2vwk	A	02	3-Layer(aba) Sandwich	2-Layer Sandwich
421	1x1p	A	01	Orthogonal Bundle	2-Layer Sandwich	422	3oju	A	01	Orthogonal Bundle	2-Layer Sandwich
423	2is3	C	00	3-Layer(aba) Sandwich	2-Layer Sandwich	424	1rch	A	00	3-Layer(aba) Sandwich	2-Layer Sandwich
425	3qio	A	00	3-Layer(aba) Sandwich	2-Layer Sandwich	426	1bl3	C	00	3-Layer(aba) Sandwich	2-Layer Sandwich
427	2bg9	D	01	Sandwich	Distorted Sandwich	428	2zju	E	00	Sandwich	Distorted Sandwich
429	1i9b	D	00	Sandwich	Distorted Sandwich	430	2ohd	D	00	3-Layer(aba) Sandwich	2-Layer Sandwich
431	2nr4	A	02	Orthogonal Bundle	Up-down Bundle	432	2iml	B	02	Orthogonal Bundle	Up-down Bundle
433	2p0l	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	434	1nvf	A	02	Orthogonal Bundle	Up-down Bundle
435	1q60	A	00	Orthogonal Bundle	Alpha-Beta Complex	436	2hay	D	00	Orthogonal Bundle	3-Layer(aba) Sandwich
437	2pju	D	02	Orthogonal Bundle	3-Layer(aba) Sandwich	438	1lp9	E	02	Beta Barrel	Sandwich
439	1ktk	F	02	Ribbon	Sandwich	440	1rfh	A	01	Irregular	2-Layer Sandwich
441	2d05	A	01	Orthogonal Bundle	2-Layer Sandwich	442	1icf	I	00	2-Layer Sandwich	Irregular
443	1mdy	B	00	Orthogonal Bundle	Irregular	444	2kid	A	00	Sandwich	Beta Barrel
445	1t2w	A	00	Sandwich	Beta Barrel	446	2bpn	A	00	2-Layer Sandwich	Alpha-Beta Complex
447	1gyo	A	00	2-Layer Sandwich	Alpha-Beta Complex	448	3ch0	A	00	3-Layer(aba) Sandwich	Alpha-Beta Barrel
449	1ap9	A	00	Orthogonal Bundle	Up-down Bundle	450	1ey3	F	01	3-Layer(aba) Sandwich	Alpha-Beta Complex
451	1pix	A	03	3-Layer(aba) Sandwich	Alpha-Beta Complex	452	3p2l	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
453	3hln	Q	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	454	1od2	A	01	3-Layer(aba) Sandwich	Alpha-Beta Complex
455	2iex	C	01	3-Layer(aba) Sandwich	Alpha-Beta Complex	456	2cby	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
457	1hnu	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex	458	3ff6	B	03	Alpha-Beta Barrel	Alpha-Beta Complex

459	3ibb	A	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	460	3h81	B	01	3-Layer(aba) Sandwich	Alpha-Beta Complex
461	2hr3	D	03	Up-down Bundle	Orthogonal Bundle	462	2ckp	B	02	Orthogonal Bundle	Alpha-Beta Complex
463	2ppq	A	02	Orthogonal Bundle	Alpha-Beta Complex	464	2hwj	E	01	2-Layer Sandwich	Alpha-Beta Complex
465	2d1k	A	04	Orthogonal Bundle	Alpha-Beta Complex	466	2yjf	D	03	Orthogonal Bundle	Alpha-Beta Complex
467	3ldp	B	04	Orthogonal Bundle	Alpha-Beta Complex	468	2v7z	B	04	Orthogonal Bundle	Alpha-Beta Complex
469	2yjf	E	03	Orthogonal Bundle	Alpha-Beta Complex	470	2yjf	A	03	Orthogonal Bundle	Alpha-Beta Complex
471	1nlv	A	03	Orthogonal Bundle	Alpha-Beta Complex	472	3a5n	C	03	Orthogonal Bundle	Alpha-Beta Complex
473	1ijj	A	04	Orthogonal Bundle	Alpha-Beta Complex	474	3iuc	A	04	Orthogonal Bundle	Alpha-Beta Complex
475	3f9m	A	01	Alpha-Beta Barrel	3-Layer(aba) Sandwich	476	1wr0	A	01	Orthogonal Bundle	Up-down Bundle
477	2a69	E	00	Orthogonal Bundle	Alpha-Beta Complex	478	3fki	F	00	2-Layer Sandwich	Alpha-Beta Complex
479	1uw4	B	00	Orthogonal Bundle	Alpha Horseshoe	480	3sr2	C	01	2-Layer Sandwich	Beta Complex
481	1mvx	A	00	Alpha-Beta Complex	Beta Complex	482	2gju	D	00	3-Layer(aba) Sandwich	4-Layer Sandwich
483	2y14	C	01	3-Layer(aba) Sandwich	2-Layer Sandwich	484	1pgv	A	00	3-Layer(aba) Sandwich	Alpha-Beta Horseshoe
485	1koh	B	00	3-Layer(aba) Sandwich	Alpha-Beta Horseshoe	486	2ca6	B	00	Alpha-Beta Barrel	Alpha-Beta Horseshoe
487	1j9z	A	03	Orthogonal Bundle	Up-down Bundle	488	1mo4	A	02	Orthogonal Bundle	2-Layer Sandwich
489	1u98	A	02	Orthogonal Bundle	2-Layer Sandwich	490	1bgj	A	01	3-Layer(aba) Sandwich	2-Layer Sandwich
491	2r87	C	03	Beta Barrel	2-Layer Sandwich	492	1wr6	A	00	Orthogonal Bundle	Up-down Bundle
493	1dq8	B	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	494	3fc4	A	03	Beta Barrel	Alpha-Beta Complex
495	2zd2	A	02	Up-down Bundle	Orthogonal Bundle	496	2vpq	B	01	2-Layer Sandwich	3-Layer(aba) Sandwich
497	1i7n	B	02	2-Layer Sandwich	3-Layer(aba) Sandwich	498	3qrr	A	02	Beta Barrel	2-Layer Sandwich
499	1a79	A	02	2-Layer Sandwich	3-Layer(aba) Sandwich	500	1gdf	A	00	Sandwich	Distorted Sandwich
501	1rrz	A	00	Orthogonal Bundle	Up-down Bundle	502	1mhq	B	00	Up-down Bundle	Alpha Horseshoe
503	1jpl	B	00	Orthogonal Bundle	Alpha Horseshoe	504	1hx8	A	01	Orthogonal Bundle	Alpha Horseshoe
505	3dll	E	01	2-Layer Sandwich	Alpha-Beta Complex	506	3r8t	G	01	2-Layer Sandwich	Alpha-Beta Complex
507	3i1p	G	01	2-Layer Sandwich	Alpha-Beta Complex	508	1ybz	A	00	Orthogonal Bundle	Up-down Bundle
509	1lrr	D	00	Orthogonal Bundle	Up-down Bundle	510	2y13	R	00	2-Layer Sandwich	Alpha-Beta Complex
511	1m08	B	00	2-Layer Sandwich	Alpha-Beta Complex	512	2gze	B	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
513	3b4s	F	02	Orthogonal Bundle	Up-down Bundle	514	2r6g	F	01	Orthogonal Bundle	Up-down Bundle
515	1mfs	A	00	Ribbon	Irregular	516	2bos	A	00	Roll	Beta Barrel
517	3hk0	A	01	2-Layer Sandwich	Roll	518	2ijf	A	03	Orthogonal Bundle	Up-down Bundle
519	3kmq	A	03	Orthogonal Bundle	Up-down Bundle	520	3kna	A	03	Orthogonal Bundle	Up-down Bundle
521	1it4	A	00	Orthogonal Bundle	Up-down Bundle	522	1kp4	A	00	Orthogonal Bundle	Up-down Bundle
523	1dik	A	03	3-Layer(aba) Sandwich	3-Layer(bba) Sandwich	524	1y0o	C	00	Sandwich	Roll
525	1q6i	A	02	Beta Barrel	Roll	526	2azw	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
527	1q33	A	02	3-Layer(aba) Sandwich	Alpha-Beta Complex	528	3cou	A	00	3-Layer(aba) Sandwich	Alpha-Beta Complex
529	1elu	A	01	3-Layer(aba) Sandwich	Alpha-Beta Complex	530	3lg0	A	01	3-Layer(aba) Sandwich	Alpha-Beta Complex
531	2ay8	A	01	3-Layer(aba) Sandwich	Alpha-Beta Complex	532	1ax4	C	01	3-Layer(aba) Sandwich	Alpha-Beta Complex
533	3qbo	A	02	Orthogonal Bundle	Alpha-Beta Complex	534	3g4f	A	02	Orthogonal Bundle	Alpha/alpha barrel
535	3m00	A	02	Orthogonal Bundle	Alpha/alpha barrel	536	3g4d	B	02	Orthogonal Bundle	Alpha/alpha barrel

537	3qan	C	01	Alpha-Beta Barrel	3-Layer(aba) Sandwich	538	3or9	R	00	Orthogonal Bundle	Irregular
539	1fzg	A	00	Orthogonal Bundle	Up-down Bundle	540	2hod	G	00	Orthogonal Bundle	Up-down Bundle
541	1fp7	A	03	2-Layer Sandwich	Roll	542	1o7f	A	03	2-Layer Sandwich	Sandwich
543	2bgc	E	01	Beta Barrel	Sandwich	544	2oyz	A	00	Beta Barrel	Sandwich
545	2vhd	A	02	3-Layer(aba) Sandwich	Orthogonal Bundle	546	1o7l	C	02	2-Layer Sandwich	Beta Barrel
547	1gs9	A	00	Orthogonal Bundle	Up-down Bundle	548	1fpo	C	02	Orthogonal Bundle	Up-down Bundle

Table S2C. Domain classification conflicts between CATH and automatic classification of protein fold space using Class, Architecture and Topology criteria.

Index	PDB ID	Chain ID	Domain ID	Predicted class	Annotated class	Index	PDB ID	Chain ID	Domain ID	Predicted class	Annotated class
1	1dif	B	00	Immunoglobulin-like	Cathepsin D, subunit A; domain 1	2	3ypi	B	00	Rossmann fold	TIM Barrel
3	1fuq	B	02	Retinoid X Receptor	Fumarase C; Chain A, domain 2	4	1o5z	A	02	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A
5	2im8	A	00	Glutathione S-transferase Yfyf (Class Pi); Chain A, domain 2	Four Helix Bundle (Hemerythrin (Met), subunit A)	6	1ava	C	00	Immunoglobulin-like	Trefoil (Acidic Fibroblast Growth Factor, subunit A)
7	1uj1	B	01	Immunoglobulin-like	Thrombin, subunit H	8	2fuw	A	00	Rossmann fold	Molybdenum Cofactor Biosynthetic Enzyme; Chain A
9	2ebd	B	01	Rossmann fold	Peroxisomal Thiolase; Chain A, domain 1	10	1q06	A	00	Arc Repressor Mutant, subunit A	A; Domain 2
11	1wfk	A	01	Factor Xa Inhibitor	Herpes Virus-1	12	3p9c	A	02	TIM Barrel	Rossmann fold
13	1ymp	A	00	Peroxidase; domain 2	Serine Threonine Protein Phosphatase 5, Tetratricopeptide repeat	14	1thz	B	02	Rossmann fold	Cytidine Deaminase; domain 2
15	1y9u	A	02	Transferase (Phosphotransferase); domain 1	D-Maltodextrin-Binding Protein; domain 2	16	1mc0	A	02	D-Maltodextrin-Binding Protein; domain 2	Beta-Lactamase
17	1cgi	E	02	SH3 type barrels.	Thrombin, subunit H	18	3itj	C	02	Immunoglobulin-like	FAD/NAD (P)-binding domain
19	1hrz	A	00	Helix Hairpins	DNA Binding (I), subunit A	20	1fc9	A	02	Immunoglobulin-like	Pdz3 Domain
21	2ofj	B	02	Rossmann fold	TIM Barrel	22	1vjj	A	02	FAD/NAD(P)-binding domain	Coagulation Factor XIII; Chain A, domain 2
23	1s3s	B	04	DNA polymerase; domain 1	Helicase, RuvA Protein; domain 3	24	3ru7	B	02	Alpha-Beta Plaits	UDP-galactose 4-epimerase; domain 1
25	2pys	B	00	Immunoglobulin-like	HIV-inactivating Protein, Cyanovirin-n	26	1urd	A	02	Rossmann fold	D-Maltodextrin-Binding Protein; domain 2
27	3cax	A	02	Immunoglobulin-like	Beta-Lactamase	28	2cx8	A	02	Rossmann fold	Alpha/beta knot
29	2z43	B	01	Arc Repressor Mutant, subunit A	DNA polymerase; domain 1	30	2wss	L	01	SH3 type barrels.	Elongation Factor Tu (Ef-tu); domain 3
31	2igx	A	02	Immunoglobulin-like	Cathepsin D, subunit A; domain	32	2z9h	B	00	Immunoglobulin-like	OB fold (Dihydrolipoamide)

33	1mjt	A	02	Alpha-Beta Plaits	Nitric Oxide Synthase;Heme Domain; Chain A, domain 2	34	1r6q	C	00	Arc Repressor Mutant, subunit A	Acetyltransferase, E2P) A,
35	1ptt	A	00	TIM Barrel	Protein-Tyrosine Phosphatase; Chain A	36	1txo	B	00	Glutamine Phosphorylphosphate, subunit 1, domain 1	Phosphatase 2c; domain 1
37	1i4j	B	00	Alpha-Beta Plaits	Ribosomal Protein L22; Chain A	38	1kfy	A	02	Rossmann fold	Flavocytochrome C3; Chain A, domain 1
39	1kly	A	02	Jelly Rolls	Beta-galactosidase; Chain A, domain 5	40	1gqy	A	03	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A
41	1c3p	A	00	TIM Barrel	Arginase; Chain A	42	1vgu	B	00	Rossmann fold	Spore Coat Polysaccharide Biosynthesis Protein SpsA; Chain A
43	2xmr	A	00	TIM Barrel	Rossmann fold	44	2fl4	A	02	Phosphorylase Kinase; domain 1	Aminopeptidase
45	3b40	A	02	Arc Repressor Mutant, subunit A	Helix Hairpins	46	1pvq	A	01	Fumarase C; Chain B, domain 1	hpl Integrase; Chain A
47	3kom	B	01	TIM Barrel	Rossmann fold	48	1ky9	B	02	Immunoglobulin-like	Thrombin, subunit H
49	3lab	A	00	Rossmann fold	TIM Barrel	50	3hiv	A	01	Rossmann fold	Ricin (A subunit); domain 1
51	1jpd	X	02	Rossmann fold	TIM Barrel	52	3ck5	B	02	Rossmann fold	TIM Barrel
53	1t09	B	00	TIM Barrel	Isopropylmalate Dehydrogenase	54	2hav	A	03	Rossmann fold	D-Maltodextrin-Binding Protein; domain 2
55	1ekj	D	00	Rossmann fold	Beta-carbonic Anhydrase; Chain A	56	1o8o	B	00	Rossmann fold	Molybdenum Cofactor Biosynthetic Enzyme; Chain A
57	1g1m	A	00	TIM Barrel	Rossmann fold	58	3afe	D	02	Immunoglobulin-like	Butyryl-CoA Dehydrogenase, subunit A; domain 2
59	1q05	A	00	Arc Repressor Mutant, subunit A	A; Domain 2	60	2z0y	B	01	Rossmann fold	Alpha/beta knot
61	1vz8	A	01	UDP-n-acetylglucosamine1-carboxyvinyl-transferase; Chain Translation	L-amino peptidase D-ALA esterase/amidase	62	3ilq	P	00	OB fold (Dihydroli poamide Acetyltransferase, E2P)	A;
63	1zzk	A	00		A, domain 1	64	1de8	A	00	BirA Bifunction	Deoxyribonuclease I;

65	1vil	0	00	Initiation Factor IF3 Alpha-Beta Plaits	Severin	66	2r5s	B	02	al Protein; domain 2 DNA polymerase; domain 1	Chain A Serine Threonine Protein Phosphatase 5, Tetratricopeptide repeat
67	1r1n	I	02	Rossmann fold	D-Maltodextrin-Binding Protein; domain 2	68	2ozg	A	01	Alpha-Beta Plaits	Aminopeptidase
69	3b93	B	00	Immunoglobulin-like	Jelly Rolls	70	2dbn	A	00	TIM Barrel	Jelly Rolls
71	1zov	B	01	Rossmann fold	FAD/NAD (P)-binding domain	72	1c9r	B	01	Alpha-Beta Plaits	HIV Type 1 Reverse Transcriptase; Chain A, domain 1
73	3i2x	B	00	Jelly Rolls	Trefoil (Acidic Fibroblast Growth Factor, subunit A)	74	2eir	B	00	Rossmann fold	Glutaredoxin
75	3h02	A	02	Helix Hairpins	Lyase 2-enoyl-coa Hydratase; Chain A, domain 2	76	1vpa	B	00	Rossmann fold	Spore Coat Polysaccharide Biosynthesis Protein SpsA; Chain A
77	1b98	A	00	Immunoglobulin-like	Cystine Knot Cytokines, subunit B	78	1gwc	B	01	Arc Repressor Mutant, subunit A	Glutaredoxin
79	1ej8	A	00	Trefoil (Acidic Fibroblast Growth Factor, subunit A)	Immunoglobulin-like	80	1ea9	D	04	CD59	Immunoglobulin-like
81	1gql	A	01	Rossmann fold	Chitobiase; domain 2	82	1ikb	0	00	TIM Barrel	Isopropylmalate Dehydrogenase
83	1oyd	A	04	Alpha-Beta Plaits	Multidrug efflux transporter AcrB TolC docking domain; DN and DC subdomains	84	3cjd	A	00	Ferritin;	Tetracycline Repressor; domain 2
85	1y47	A	00	Helix Hairpins	Single alpha-helices involved in coiled-coils or other helix-helix interfaces	86	2c46	D	00	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A
87	1oz0	B	02	Cytochrome Bc1 Complex; Chain A, domain 1	Cytidine Deaminase; domain 2	88	2eo0	B	00	Nucleotidyltransferase; domain 5	A, domain 1
89	1i6p	A	00	Rossmann fold	Beta-carbonic	90	1cx1	A	00	Immunoglobulin-	Jelly Rolls

91	3bf2	A	00	A,	Anhydrase; Chain A Double Stranded RNA Binding Domain	92	1dkg	A	01	like	
93	2h8a	A	00	Helix Hairpins	Four Helix Bundle (Hemerythr in (Met), subunit A)	94	1ewy	A	01	Thrombin, subunit H	Hemaggluti nin Ectodomai n; Chain B
95	2yrf	B	01	Tetracycli ne Repressor ; domain 2	Four Helix Bundle (Hemerythr in (Met), subunit A)	96	1anx	B	04	DNA polymeras e; domain 1	Elongation Factor Tu (Ef-tu); domain 3
97	1um2	A	02	Arc Repressor Mutant, subunit A	Endonuclea se I-creI	98	2gwo	A	00	Rossmann fold	Annexin V; domain 1
99	3g7t	A	01	Immunogl obulin- like	Tick-borne Encephaliti s virus Glycoprote in; domain 1	100	2qt3	B	01	Immunogl obulin- like	Protein- Tyrosine Phosphatas e; Chain A Urease, subunit C; domain 1
101	3kcf	C	01	A,	Phosphoryl ase Kinase; domain 1	102	1woc	D	00	Immunogl obulin- like	OB fold (Dihydrolip oamide Acetyltrans ferase, E2P)
103	1e0t	C	03	Immunogl obulin- like	M1 Pyruvate Kinase; Domain 3	104	1vz8	C	01	Rossmann fold	L-amino peptidase D-ALA esterase/am idase
105	2dgc	A	00	Helix Hairpins	Double Stranded RNA Binding Domain	106	1b89	A	00	Farnesyl Diphosph ate Synthase	Serine Threonine Protein Phosphatas e 5, Tetratricop eptide repeat
107	1lee	A	02	Immunogl obulin- like	Cathepsin D, subunit A; domain 1	108	1fpz	D	00	Rossmann fold	Protein- Tyrosine Phosphatas e; Chain A Four Helix Bundle (Hemerythr in (Met), subunit A)
109	1xmt	A	00	Alpha- Beta Plaits	Aminopept idase	110	3dww	C	00	Serum Albumin; Chain A, Domain 1	Serum Albumin; Chain A, Domain 1
111	1za3	S	01	Laminin	Tumor Necrosis Factor Receptor, subunit A; domain 2	112	3lu7	A	06	Arc Repressor Mutant, subunit A	Serum Albumin; Chain A, Domain 1
113	1ojl	B	02	DNA polymeras e; domain 1	Helicase, Ruva Protein; domain 3	114	1a1z	A	00	Recoverin ; domain 1	Death Domain, Fas
115	2i7f	A	00	Immunogl obulin- like	Rieske Iron-sulfur Protein	116	1a6s	A	00	Arc Repressor Mutant, subunit A	DNA polymerase ; domain 1
117	1b8m	A	00	Immunogl obulin- like	Cystine Knot Cytokines, subunit B	118	2b0w	B	03	Rossmann fold	Cytidine Deaminase; domain 2
119	1ev9	D	01	Alpha- Beta Plaits	Glutaredox in	120	1kwo	C	02	Arc Repressor Mutant, subunit A	Recoverin; domain 1

121	3nm9	A	00	Helix Hairpins	DNA Binding (I), subunit A	122	lysa	D	00	Single alpha-helices involved in coiled-coils or other helix-helix interfaces	Double Stranded RNA Binding Domain
123	2j86	B	00	Rossmann fold	Phosphatase 2c; domain 1	124	3cwg	A	04	Phosphorylase Kinase; domain 1	SHC Adaptor Protein
125	1wpp	A	01	Rossmann fold	inorganic pyrophosphatase (n-terminal core)	126	2iph	B	02	Immunoglobulin-like	Thrombin, subunit H
127	3dv0	E	00	TIM Barrel	Rossmann fold	128	3hfr	A	02	main proteinase (3clpro) structure, domain 3	Rossmann fold
129	2nv5	A	00	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A	130	3e2x	A	00	Rossmann fold	Beta-carbonic Anhydrase; Chain A
131	1q11	A	01	Retinoid X Receptor	Rossmann fold	132	3e04	A	03	Arc Repressor Mutant, subunit A	Ribonucleotide Reductase Protein R1; domain 1
133	2opp	B	04	Rossmann fold	Alpha-Beta Plaits	134	1iyl	D	02	Rossmann fold	Aminopeptidase
135	1qha	B	03	Rossmann fold	Nucleotidyl transferase; domain 5	136	3cme	J	00	Rossmann fold	A;
137	2pop	A	00	TIM Barrel	Phosphatase 2c; domain 1	138	1xm2	B	00	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A
139	1d4v	A	01	Laminin	Tumor Necrosis Factor Receptor, subunit A; domain 2	140	1yr2	A	01	TIM Barrel	Rossmann fold
141	1eyr	B	00	Rossmann fold	Spore Coat Polysaccharide Biosynthesis Protein SpsA; Chain A	142	3dru	C	01	Rossmann fold	Antithrombin; Chain I, domain 2
143	2ogj	A	01	Immunoglobulin-like	Urease, subunit C; domain 1	144	1yj5	C	00	Immunoglobulin-like	Tumour Suppressor Smad4
145	1mk9	D	02	Thrombin, subunit H	PH-domain like	146	2khi	A	00	Immunoglobulin-like	OB fold (Dihydrolipamide Acetyltransferase, E2P)
147	1hru	A	00	Nucleotidyltransferase; domain 5	DHBP synthase	148	1htt	B	02	Alpha-Beta Plaits	Rossmann fold
149	2zy9	A	03	Globin-like	Tetracycline Repressor; domain 2	150	3itj	D	01	Rossmann fold	FAD/NAD (P)-binding domain
151	3cnt	B	01	Rossmann fold	FAD/NAD (P)-binding domain	152	2w8o	A	02	Rossmann fold	Aldehyde Dehydrogenase; Chain A, domain 2

153	3aff	A	02	Immunoglobulin-like	Butyryl-CoA Dehydrogenase, subunit A; domain 2	154	2oc3	A	00	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A
155	3llp	B	04	Immunoglobulin-like	Trefoil (Acidic Fibroblast Growth Factor, subunit A)	156	1r5d	A	00	Thrombin, subunit H	P-30 Protein
157	1kyi	E	02	Single alpha-helices involved in coiled-coils or other helix-helix interfaces	Helicase, Ruva Protein; domain 3	158	2phm	A	00	TIM Barrel	Phenylalanine Hydroxylase
159	1kh2	C	02	Rossmann fold	Argininosuccinate synthetase, chain A, domain 2	160	1bdf	D	02	OB fold (Dihydroli poamide Acetyltransferase, E2P)	RNA Polymerase Alpha Subunit; Chain A, domain 2
161	1d2e	C	03	Immunoglobulin-like	Elongation Factor Tu (Ef-tu); domain 3	162	1v5b	B	00	TIM Barrel	Isopropylmalate Dehydrogenase
163	1ary	A	01	Transferase (Phosphotransferase); domain 1	Peroxidase; domain 1	164	3m8u	A	03	Rossmann fold	Dipeptide-binding Protein; domain 3
165	2h4r	A	01	Rossmann fold	Antithrombin; Chain I, domain 2	166	1rf5	D	01	Rossmann fold	UDP-n-acetylglucosamine 1-carboxyvinyl-transferase; Chain
167	1j70	A	01	Glutaredoxin	Rossmann fold	168	2d2r	A	00	Rossmann fold	Undecaprenyl pyrophosphate synthetase
169	1exm	A	02	Immunoglobulin-like	Elongation Factor Tu (Ef-tu); domain 3	170	1tr1	A	00	Arc Repressor Mutant, subunit A	Neutral Protease; domain 2
171	1yrr	A	01	Immunoglobulin-like	Urease, subunit C; domain 1	172	1v58	A	02	Transferase (Phosphotransferase); domain 1	Glutaredoxin
173	1w17	B	00	Rossmann fold	TIM Barrel	174	1xx5	C	01	2,3-Dihydroxybiphenyl 1,2-Dioxygenase; domain 1	Pathogenesis-related Protein p14a
175	1xm2	E	00	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A	176	1fon	A	02	PH-domain like	Thrombin, subunit H
177	1pwa	A	00	Immunoglobulin-like	Trefoil (Acidic Fibroblast Growth Factor, subunit A)	178	2qr6	A	01	Single alpha-helices involved in coiled-coils or other helix-	Tumor Necrosis Factor Receptor, subunit A; domain 2

179	3fd3	A	01	Alpha-Beta Plaits	D-Maltodextrin-Binding Protein; domain 2	180	1lpj	A	00	helix interfaces Thrombin, subunit H	Lipocalin
181	1gsj	A	00	Rossmann fold	Carbamate kinase	182	1pkl	D	03	Immunoglobulin-like	M1 Pyruvate Kinase; Domain 3
183	1j7m	A	00	Complement Module; domain 1	Seminal Fluid Protein PDC-109 (Domain B)	184	3h7q	A	00	Rossmann fold	Ribosomal Protein L22; Chain A
185	1dgc	A	00	Single alpha-helices involved in coiled-coils or other helix-helix interfaces	Double Stranded RNA Binding Domain	186	1d7y	A	02	Rossmann fold	FAD/NAD (P)-binding domain
187	1h41	A	01	Rossmann fold	Chitinase; domain 2	188	2pf5	D	00	Cytidine Deaminase; domain 2	Mannose-Binding Protein A; Chain A
189	2kwf	A	00	DNA polymerase; domain 1	Serum Albumin; Chain A, Domain 1	190	2q47	B	00	Rossmann fold	Protein-Tyrosine Phosphatase; Chain A
191	3e5y	A	00	Rossmann fold	Alpha/beta knot	192	1w1j	A	02	Ribosomal Protein L14	Uridine Diphosphon-acetylenolpyruvylglucosamine Reductase; domain 3
193	2y69	M	00	Single alpha-helices involved in coiled-coils or other helix-helix interfaces	Cytochrome C Oxidase; Chain M	194	1y8g	A	02	Rossmann fold	Transferase (Phosphotransferase); domain 1
195	1hyz	A	00	Rossmann fold	Nucleotidyl transferase; domain 5	196	1wye	A	00	TIM Barrel	D-glutamate ligase
197	1esq	C	00	TIM Barrel	D-glutamate ligase	198	3iax	A	01	Beta Polymerase; domain 2	Rossmann fold
199	1nox	A	00	Transferase (Phosphotransferase); domain 1	NADH Oxidase	200	1i2k	A	01	Alpha-Beta Plaits	D-amino Acid Aminotransferase; Chain A, domain 1
201	1xrb	A	02	Alpha-Beta Plaits	GMP Synthetase; Chain A, domain 3	202	1upp	K	00	Alpha-Beta Plaits	Ribulose 1,5 Bisphosphate Carboxylase/Oxygenase
203	1svd	M	00	Alpha-Beta Plaits	Ribulose 1,5 Bisphosphate	204	2gz4	D	00	Rossmann fold	Hypothetical protein af1432

205	1ey3	F	01	Rossmann fold	Carboxylase/Oxygenase 2-enoyl-CoA Hydratase; Chain A, domain 1	206	2iex	A	01	Rossmann fold	2-enoyl-CoA Hydratase; Chain A, domain 1
207	1q52	L	01	Rossmann fold	2-enoyl-CoA Hydratase; Chain A, domain 1	208	1y7o	G	00	Rossmann fold	2-enoyl-CoA Hydratase; Chain A, domain 1
209	3gf3	A	02	Rossmann fold	2-enoyl-CoA Hydratase; Chain A, domain 1	210	3gr9	C	01	Rossmann fold	Aspartate Aminotransferase; domain 2
211	3drd	A	02	Rossmann fold	Aspartate Aminotransferase; domain 2	212	3aat	A	02	Rossmann fold	Aspartate Aminotransferase; domain 2
213	1z3z	A	02	Rossmann fold	Aspartate Aminotransferase; domain 2	214	1mdz	A	01	Rossmann fold	Aspartate Aminotransferase; domain 2
215	2g6w	A	02	TIM Barrel	Aspartate Aminotransferase; domain 2	216	1m32	B	02	Rossmann fold	Aspartate Aminotransferase; domain 2
217	3gr9	B	01	Rossmann fold	Aspartate Aminotransferase; domain 2	218	1gc0	A	01	Rossmann fold	Aspartate Aminotransferase; domain 2
219	1j7u	B	02	Rossmann fold	A, domain 2	220	2q83	A	02	Transferase(Phosphotransferase); domain 1	A, domain 2
221	1poi	D	00	Rossmann fold	Glutaconate Coenzyme A-transferase; Chain B	222	2nrb	D	02	Rossmann fold	Glutaconate Coenzyme A-transferase; Chain B
223	1xr4	A	02	Spore Coat Polysaccharide Biosynthesis Protein SpsA; Chain A	Glutaconate Coenzyme A-transferase; Chain B	224	1yxr	A	01	Helix Hairpins	Methane Monooxygenase Hydroxylase; Chain G, domain 1
225	2vtx	H	00	Immunoglobulin-like	Jelly Rolls	226	3mez	B	00	Immunoglobulin-like	Agglutinin, subunit A
227	1dlp	D	01	Immunoglobulin-like	Agglutinin, subunit A	228	1bwu	P	00	Immunoglobulin-like	Agglutinin, subunit A
229	1z1v	A	00	Recoverin; domain 1	DNA polymerase; domain 1	230	1b4f	A	00	Arc Repressor Mutant, subunit A	DNA polymerase; domain 1
231	3d1v	A	00	TIM Barrel	Rossmann fold	232	1t8y	F	02	TIM Barrel	Rossmann fold
233	3t02	A	01	TIM Barrel	Alkaline Phosphatase, subunit A	234	115j	A	03	Rossmann fold	Aconitase; Domain 2
235	3cjt	F	02	Alpha-Beta Plaits	Arc Repressor Mutant, subunit A	236	3d66	B	02	TIM Barrel	Aminopeptidase
237	3d68	C	02	TIM Barrel	Aminopeptidase	238	1a3x	B	01	Rossmann fold	A, domain 1
239	2fy8	C	03	OB fold (Dihydroli	Alpha-Beta Plaits	240	1rrz	A	00	Recoverin; domain 1	Transferase,

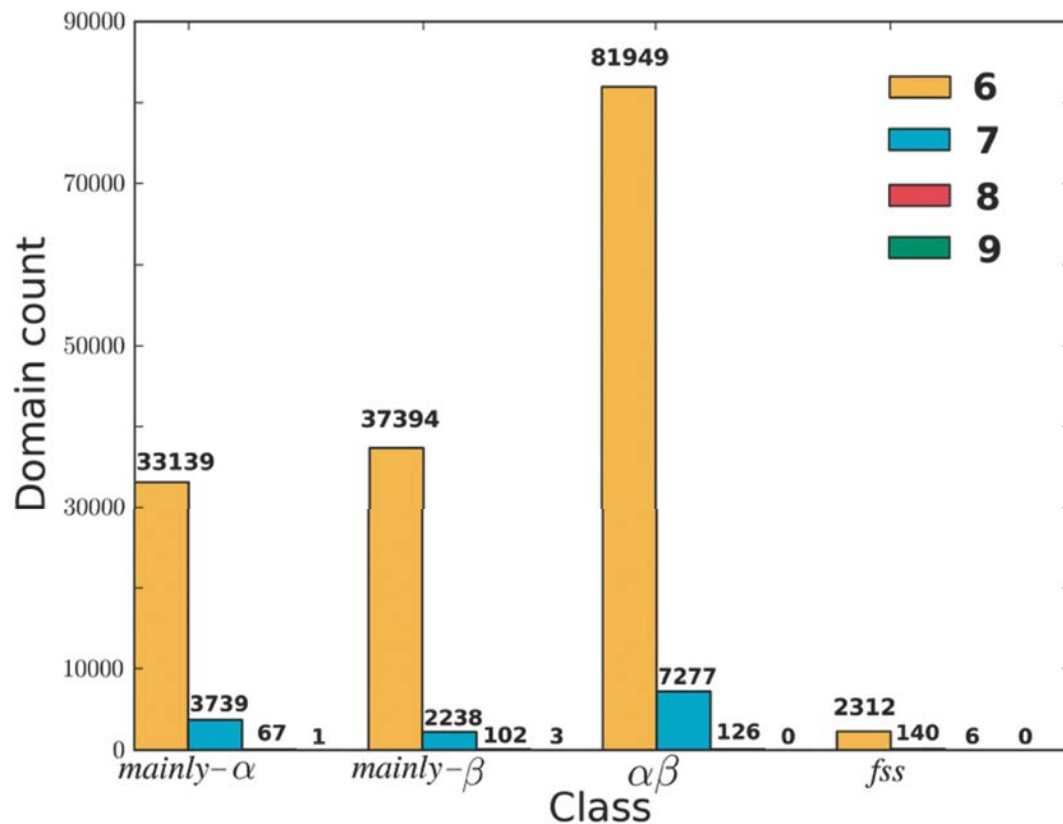
					poamide Acetyltran sferase, E2P)						Pyrimidine Nucleoside Phosphoryl ase; Chain A, domain 3
241	1sza	A	00	Globin- like	Serine Threonine Protein Phosphatas e 5, Tetratricop eptide repeat	242	2vhn	G	01	Alpha- Beta Plaits	Outer Surface Protein A; domain 3
243	1mww	A	00	Alpha- Beta Plaits	Macrophag e Migration Inhibitory Factor	244	1bcp	C	02	Immunogl obulin- like	OB fold (Dihydroli p oamide Acetyltrans ferase, E2P)
245	1et6	A	02	Immunogl obulin- like	OB fold (Dihydroli p oamide Acetyltrans ferase, E2P)	246	2tpt	A	02	Rossmann fold	Pyrimidine Nucleoside Phosphoryl ase; Chain A, domain 2
247	1vqu	A	02	Rossmann fold	Pyrimidine Nucleoside Phosphoryl ase; Chain A, domain 2	248	1kgz	B	02	Rossmann fold	Pyrimidine Nucleoside Phosphoryl ase; Chain A, domain 2
249	1gxb	C	02	TIM Barrel	Pyrimidine Nucleoside Phosphoryl ase; Chain A, domain 2	250	1s3s	G	00	Alpha- Beta Plaits	Ubiquitin- like (UB roll)
251	1t11	A	03	Immunogl obulin- like	Chitinase A; domain 3	252	1fks	A	00	Thrombin, subunit H	Chitinase A; domain 3
253	3zr0	A	00	Immunogl obulin- like	Nucleoside Triphospha te Pyrophosp hohydrolas e	254	1kt9	A	00	A;	Nucleoside Triphospha te Pyrophosp hohydrolas e
255	3dku	E	00	Mannose- Binding Protein A; Chain A	Nucleoside Triphospha te Pyrophosp hohydrolas e	256	1xi9	C	01	Rossmann fold	Aspartate Aminotrans ferase, domain 1
257	1bs0	A	01	Rossmann fold	Aspartate Aminotrans ferase, domain 1	258	1iax	A	01	Rossmann fold	Aspartate Aminotrans ferase, domain 1
259	1w8z	A	00	Immunogl obulin- like	Jelly Rolls	260	1ipk	B	01	Lipocalin	Jelly Rolls
261	2i45	H	00	Immunogl obulin- like	Jelly Rolls	262	1q9c	H	01	Ferritin;	Histone, subunit A
263	3dwo	X	00	Neuramin idase	Porin	264	1t11	B	00	Methylam ine Dehydrog enase; Chain H	Porin
265	1k1f	E	00	Single alpha- helices involved in coiled- coils or other helix- helix	MYOD Basic- Helix- Loop-Helix Domain, subunit B	266	2iuk	B	01	Cathepsin D, subunit A; domain 1	Lipoxygen ase-1

267	1rp1	A	02	interfaces Immunogl obulin- like	Lipoxygen ase-1	268	2wy6	A	02	Lipocalin	Lipoxygen ase-1
269	1n8s	A	02	Immunogl obulin- like	Lipoxygen ase-1	270	2qmx	A	03	Phosphory lase Kinase; domain 1	Alpha-Beta Plaits

Table S3. SCOP classification and SAS score agreement. The table shows the fraction of domain pairs aligned in each SAS level (29, 35, 50) that belongs to the same SCOP classification.

SAS level	Family	Superfamily	Fold	Class
20	0.1288	0.3837	0.8417	0.9198
35	0.0140	0.0571	0.1116	0.7640
50	0.0032	0.0138	0.0311	0.4437

Figure S1. Maximum clusters size observed in proteins



Frequency of occurrence for RCCs of size 6, 7, 8 and 9 found in domains present in CATHALL are presented in the histogram. On top of each bar is the total number of domains containing each specified RCC. RCCs of size six are observed in all CATH classes, but that is not the case for larger RCC sizes.