Capítulo 2

PLANTEAMIENTO DEL PROBLEMA

En el presente capítulo se planteará el problema de que trata esta tesis. En la sección 2.1 se describirá el problema biológico que fue la principal motivación para la realización de este trabajo. Luego, en la sección 2.2 describiremos el modelo matemático que emplearemos para resolver el problema que nos concierne. Y, finalmente, en la sección 2.3 se presentará el contenido de los siguientes capítulos.

2.1 PLANTEAMIENTO DEL PROBLEMA BIOLÓGICO

Supongamos por un lado que un grupo de personas se encuentra en lista de espera para recibir el transplante de un órgano vital. Por otro lado, se presenta una persona con la posibilidad de donar dicho órgano a algún miembro de este grupo de receptores. Nuestro problema central es establecer, en el menor tiempo posible, cuál de ellos es genéticamente compatible con el donador para que el transplante se lleve a cabo exitosamente. Y para esto se requiere que los genes del receptor y del donador, encargados de aceptar o rechazar un órgano ajeno, sean compatibles.

Estos genes se ubican en una región del código genético humano ubicada en el brazo corto del cromosoma 6, denominada antígenos de leucocitos humanos (HLA) [7]. Están divididos en varias clases, de las cuales sólo analizaremos la clase I. Dicha clase se subdivide en tipos (A, B y C) y concentraremos nuestro estudio únicamente al tipo A. Al mismo tiempo, cada tipo está dividido en subclases (denominadas especificidades en Biología), las cuales están identificadas por un número. Estas subclases son conjuntos de segmentos de cadena genética todos diferentes entre sí, pero con ciertas características comunes que los hacen pertenecer a una subclase u otra. Un segmento de cadena de información genética recibe el nombre de alelo. Esto es, un alelo es una versión de un gen o segmento de cadena genética. Y en nuestro caso, una versión de un gen HLA, clase I, tipo A.

Los alelos se representan mediante sucesiones o filas de 544 caracteres de las letras A, G, C y T, que corresponden a las bases nitrogenadas Adenina, Guanina, Citosina y Timina, respectivamente. O sea que los genes HLA-I-A son un conjunto de filas de este tipo,

dividido en subclases, como mencionábamos en el párrafo anterior. Este conjunto (HLA-I-A) es una lista que se actualiza periódicamente, pues reúne la información conocida hasta el momento sobre los alelos en cuestión y los agrupa en su correspondiente subclase. Para esta tesis utilizaremos la lista vigente a febrero de 2003, que consta de 251 alelos y 21 subclases.

Ahora bien, para que un transplante de órgano no presente rechazo, no es necesario que el donador y el receptor posean exactamente los mismos alelos del gen HLA en su información genética, basta con que dichos alelos pertenezcan a la misma subclase [4].

Por otro lado, todo individuo posee dos alelos para cada uno de sus genes, pues reciben uno de su madre y otro de su padre, y cada uno de estos alelos pertenece a una subclase. Por lo que, para llevar a cabo un transplante exitoso entre dos personas, es necesario que el par de subclases a las que pertenecen los alelos de una, coincida exactamente con el par de subclases a las que pertenecen los alelos de la otra; sin importar el origen de cada alelo (materno o paterno).

En otras palabras, nuestro problema se reduce a determinar el par de subclases correspondiente al donador y el par correspondiente a cada uno de los receptores. De esta forma, aquel receptor cuyo par de subclases coincida con el del donador es un candidato seguro a recibir el órgano en cuestión.

Ahora es necesario comprender de qué forma es posible encontrar las subclases a las que pertenecen los alelos HLA de un ser humano. Primero, para simplificar la situación, empecemos por analizar cómo es posible hallar la subclase a la que pertenece un solo alelo.

Como se mencionó anteriormente, un alelo es una cadena de letras (A, G, C y T), y lo que diferencia un alelo de otro es el orden en el que aparecen estas letras en la sucesión de 544 caracteres de longitud. Los caracteres de la cadena están ordenados en 544 posiciones enumeradas. Así, dos alelos son distintos entre sí, si tienen letras diferentes en al menos una posición.

Es posible determinar, mediante un análisis molecular, si una letra se encuentra en una determinada posición o no. Es decir, mediante el uso de reactivos específicos podemos "preguntar" al código genético de una persona si en una posición dada existe una letra específica [4]. También puede determinarse si existen ciertas letras en 20 posiciones consecutivas del código mediante una sola pregunta, pero este tipo de preguntas en múltiples posiciones no se utilizarán en este trabajo.

Ahora bien, dado que todos los alelos son distintos entre sí, siempre es posible encontrar una lista de preguntas que nos diga de qué alelo se trata. Sin embargo, como se explicó anteriormente, nosotros no requerimos saber de qué alelo específico se trata, sino a qué subclase pertenece dicho alelo. Por lo tanto, necesitamos encontrar una lista de preguntas que, dependiendo de las respuestas arrojadas, nos indique a qué subclase pertenece el alelo en cuestión.

No obstante, en la realidad, el problema es más complicado. Como se dijo, los seres humanos poseemos un par de alelos para cada gen. Estos alelos están ligados entre sí de la siguiente manera. Si queremos saber si existe una letra determinada en una cierta posición, debemos hacer la "pregunta" a ambos alelos simultáneamente. La respuesta obtenida sería afirmativa si la letra existe en la posición dada en al menos uno de los alelos, sin saber en cual; la respuesta sería negativa si ninguno de los alelos tuviera la letra buscada en la posición que se preguntó. El lector debe notar que si la respuesta es afirmativa, no se sabe si es porque un solo alelo tiene la letra en la posición determinada o porque ambos la tienen. Esto representa una pérdida de información, lo cual complica enormemente la tarea de encontrar una lista de preguntas que nos permita encontrar el par de subclases a las que pertenecen los alelos de un individuo. Más aún, si queremos encontrar una lista de preguntas mínima, que reduzca lo más posible el tiempo (y el costo en materiales de laboratorio) de la clasificación de alelos en subclases.

Debido a las características de este problema, descritas en esta primera sección, requerimos de un modelo matemático adecuado que nos permita obtener una lista de preguntas que determine la clasificación buscada. En la siguiente sección de este capítulo describiremos el modelo cuya solución es el problema central de este trabajo de tesis.

2.2 DESCRIPCIÓN DEL MODELO MATEMÁTICO

La información total sobre los alelos HLA-I-A, como se explicó en la sección anterior, es una lista de 251 sucesiones de letras (A, G, C y T) con una longitud de 544 caracteres. Esto es, un conjunto **R** de 251 sucesiones cada una de 544 elementos en el conjunto **L** = {A, G, C, T}. Si el *n-ésimo* término ($1 \leq \mathbf{n} \leq 544$) de una sucesión es una letra $X \in \mathbf{L}$ diremos que la letra X se encuentra en la posición *n* del alelo que representa esta sucesión. Debido a la equivalencia entre alelo y sucesión, nos referiremos a cada elemento de **R** mediante el alelo que representa. Los alelos pertenecientes al conjunto **R** están agrupados en 21 subclases o subconjuntos que no se intersectan entre sí.

Según lo que se dijo en la sección 2.1, trataremos solamente con los genes HLA, clase I, tipo A. Luego entonces, dado que no haremos referencia a otras clases en HLA, reservaremos el término "clase" para referirnos a una subclase del conjunto en cuestión. Esto lo hacemos para simplificar la notación, pues más adelante se introducirá el concepto de biclase, evitando nombres más complicados.

En estos términos, la información a la que nos referimos consiste del conjunto **R** de cardinalidad igual a 251, cuyos elementos se denominan alelos (sucesiones) y poseen 544 posiciones (elementos de la sucesión). Sin embargo, los datos que nos fueron proporcionados ya habían sido depurados. Es decir, fueron eliminadas las posiciones que no nos daban información valiosa, o sea las posiciones en las que todos los alelos del conjunto **R** tenían la misma letra. Esta depuración eliminó 426 posiciones, dejando un total de 118.

Así, los datos originales constan de un conjunto que denotaremos por la letra **S**, cuyos elementos son 251 alelos con 118 posiciones, agrupados en 21 clases. Además, el conjunto

de clases forma una partición **P** de **S**. O sea que **P** es una familia de 21 subconjuntos disjuntos de **S** denotados por números, con **P** = {01, 02, 03, 11, 23, 24, 25, 26, 29, 30, 31, 32, 33, 34, 36, 43, 66, 68, 69, 74, 80}.

Al ser todos los alelos de la misma longitud (118), podemos construir a partir de **S** una matriz cuyas filas y columnas son los alelos y posiciones, respectivamente. De manera que tenemos una matriz **M** de 251 × 118, con entradas en el conjunto **L** = {A, G, C, T}.

Para aclarar esta idea, analicemos un ejemplo. La siguiente tabla presenta un extracto pequeño de los datos, presentados en la forma en que nos fueron entregados para esta tesis.

Posición	5	8	17	19	24	25
A*010101	C	C	G	A	A	T
A*010102	C	C	G	A	A	T
A*0102	C	C	G	A	A	C
A*020101	T	A	T	G	T	T
A*020102	C	A	T	G	T	T
A*020103	T	A	C	A	T	T
A*020104	T	A	A	C	T	T
A*110101	C	T	G	T	C	A
A*110102	C	T	G	C	C	A
A*4301	C	C	G	A	T	A
A	0	4	1	5	3	3
G	0	0	6	2	0	0
C	7	4	1	2	2	1
T	3	2	2	1	5	6

Tabla 1: Datos originales resumidos.

En la Tabla 1 observamos un conjunto de 10 alelos (A*010101, A*010102,..., A*4301) agrupados en 4 clases (01, 02, 11, 43). La letra A que aparece antes del número de cada alelo, indica que estamos tratando con el grupo A, por lo tanto todos los alelos de la tabla (y de nuestro problema general) comienzan por esta letra. Los siguientes dos dígitos determinan el número de la clase a la que pertenece el alelo, es decir, tenemos 3, 4, 2 y 1 alelos en las clases 01, 02, 11 y 43 respectivamente. Los dígitos restantes representan el número del alelo dentro de la clase, pero esta información no es relevante para nuestro problema.

La primera línea de esta tabla, nos indica el número de la posición. De este modo, el alelo A*010101 tiene la letra C en las posiciones 5 y 8, la letra G en la posición 17, etcétera. Note que el número de las posiciones no es consecutivo. Esto se debe a la depuración de los datos, en donde se eliminaron las columnas o posiciones que tenían la misma letra para todos los alelos. Finalmente, en la parte inferior de la tabla aparecen las cuatro letras del conjunto **L** y en cada columna la frecuencia con que éstas aparecen en la posición correspondiente.

De la misma manera se presentan los datos originales de este trabajo, sólo que contamos con 251 alelos y 21 clases con distintos números de elementos, desde un solo alelo en algunas clases hasta 65 alelos en la clase más grande.

En la sección 2.1 se explicó la manera en que pueden hacerse preguntas al código genético. En una posición, se pregunta si hay una letra específica y la respuesta sólo puede ser *Sí* o *No*. Si la respuesta es negativa, no podemos saber cuál otra letra se ubica en la posición referida, sino mediante una nueva pregunta.

Como ya se dijo en el planteamiento del problema biológico, lo que queremos es determinar el par de clases a las que pertenecen los alelos de cualquier persona. Esto puede lograrse haciendo preguntas al código genético. Así, llamaremos solución al problema al conjunto de preguntas que nos permitan determinar el par de clases a las que pertenecen los dos alelos HLA de un ser humano.

Ahora deseamos transformar la información de tal modo que podamos estudiarla más fácilmente. Dicha transformación está dividida en tres pasos que se presentan a continuación.

Primer Paso

A cada posición le haremos corresponder tantas columnas como letras diferentes haya en esa posición en todos los datos. Es decir, si en la posición *n* sólo aparecen dos letras distintas en todo el conjunto de alelos, le haremos corresponder dos columnas a esta posición; si aparecen tres letras distintas, le corresponden tres columnas, etcétera.

Posición	5 5	8 8 8	1 1 1 1 7 7 7 7	1 1 1 1 9 9 9 9	2 2 2 4 4 4	2 2 2 5 5 5
A*010101	C C	C C C	G G G G	A A A A	A A A	T T T
A*010102	C C	C C C	G G G G	A A A A	A A A	T T T
A*0102	C C	C C C	G G G G	A A A A	A A A	C C C
A*020101	T T	A A A	T T T T	G G G G	T T T	T T T
A*020102	C C	A A A	T T T T	G G G G	T T T	T T T
A*020103	T T	A A A	C C C C	A A A A	T T T	T T T
A*020104	T T	A A A	A A A A	C C C C	T T T	T T T
A*110101	C C	T T T	G G G G	T T T T	C C C	A A A
A*110102	C C	T T T	G G G G	C C C C	C C C	A A A
A*4301	C C	C C C	G G G G	A A A A	T T T	A A A
A	0 0	4 4 4	1 1 1 1	5 5 5 5	3 3 3	3 3 3
G	0 0	0 0 0	6 6 6 6	2 2 2 2	0 0 0	0 0 0
C	7 7	4 4 4	1 1 1 1	2 2 2 2	2 2 2	1 1 1
T	3 3	2 2 2	2 2 2 2	1 1 1 1	5 5 5	6 6 6

Tabla 2: Primer paso de la transformación.

Para aclarar ideas, consideremos la Tabla 1. En la posición 5, si observamos verticalmente, encontramos sólo las letras C y T para todos los alelos ahí representados. En este caso, a la posición 5 le asignamos dos columnas. En la posición 8, hay tres letras distintas (A, C y T), de modo que le asignaremos tres columnas a esta posición. A la posición 17 le corresponden cuatro columnas, pues aparecen las cuatro letras en ella. Continuamos de esta forma y obtenemos la matriz que aparece en la Tabla 2.

Segundo Paso

El siguiente paso consiste en construir, a partir de esta nueva matriz, otra de igual tamaño cuyas entradas sean sólo 0 ó 1. Para hacer esto, primero insertaremos una nueva fila debajo de la que representa la posición, que nos servirá de encabezado. Ahí escribiremos las diferentes letras que aparecen en cada posición; por esto le asignamos varias columnas a cada una. Después, para construir nuestra matriz le asignamos a cada entrada el número 1 si en la misma entrada de la matriz anterior se encuentra la letra que está escrita en la nueva fila y un 0 en el caso contrario.

La matriz obtenida mediante este proceso tiene entradas en el conjunto {0,1} y esos números representan respectivamente la respuesta *No* y S*i* a la pregunta que aparece en el encabezado. Aplicando este método a la matriz de la Tabla 2, se construyó la matriz que aparece en la siguiente tabla:

Posición	55		888			1717				1919				244			255		
Pregunta	**C**	**T**	**A**	**C**	**T**	**A**	**G**	**C**	**T**	**A**	**G**	**C**	**T**	**A**	**C**	**T**	**A**	**C**	**T**
A*010101	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1
A*010102	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1
A*0102	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0
A*020101	0	1	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1
A*020102	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1
A*020103	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1
A*020104	0	1	1	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1
A*110101	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	0	0
A*110102	1	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	1	0	0
A*4301	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0

Tabla 3: Segundo paso de la transformación.

Ahora interpretemos esta matriz. Los alelos ya no vienen representados como sucesiones de letras, sino a través de las respuestas que dan a las preguntas del encabezado. En la primera columna aparecen las respuestas a la pregunta ¿existe una C en la posición 5?. Así, el alelo A*010101 tiene un 1 (Sí) en esta columna y el alelo A*020101 tiene un 0 (No). En efecto, si observamos en la Tabla 2, el alelo A*010101 tiene una C en la posición 5 y el alelo A*020101 una T en esa posición. Tal como se había explicado antes, el cero representa simplemente una respuesta negativa y no nos habla de la letra que sí existe en esa posición. En esta forma interpretamos la información que aparece en la matriz de la Tabla 3.

Al aplicar los dos pasos de transformación descritos anteriormente a la información total de nuestro problema, se obtuvo una matriz de 0's y 1's con 265 columnas y 251 filas. El lector podrá notar que hasta este punto no se ha afectado el número de filas con las que se cuenta; sólo se aumentó el número de columnas de 118 a 265.

Hasta ahora sólo hemos tratado con la información en forma sencilla. Es decir, cada fila de la matriz del segundo paso representa un solo alelo. Sin embargo, según lo que se dijo en la sección 2.1, toda persona posee dos alelos en su código genético. Debido a nuestra naturaleza y a los métodos empleados en la actualidad para hacer "preguntas" este código doble, una parte de la información queda oculta, como se explicará más adelante. Lo cual representa una primera complicación. La segunda complicación es el "tamaño" del problema. El hecho de tratar con dos alelos aumenta considerablemente los datos a analizar, pues debemos considerar todos los pares posibles, incluyendo los que se forman a partir de uno solo. Esta última situación debe ser tomada en cuenta, pues no podemos descartar la posibilidad de que el alelo materno sea igual al alelo paterno. Si tenemos *m* alelos, la cantidad de pares que se pueden formar a partir de ellos es m(m+1)/2. Nosotros contamos con 251 alelos, o sea que el número de pares de alelos posibles asciende a 31626.

A continuación se describe el tercer y último paso para transformar la información a la forma en que será analizada para solucionar el problema. Nos resta formar todos los pares posibles de alelos a partir de la lista de 251 que tomamos como información inicial. Esto debe hacerse tomando en consideración la pérdida de información de la que hablábamos anteriormente.

Tercer Paso

Un par de alelos generarán lo que llamaremos un *bialelo*. Un bialelo se forma tomando dos alelos y sumándolos posición por posición, como si sus entradas fueran elementos de un álgebra booleana. Es decir, bajo la regla **0 + 0 = 0, 0 + 1 = 1 + 0 = 1** y **1 + 1 = 1**. Como ejemplo tomemos el siguiente par de alelos y el bialelo que forman:

A*010102	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1
A*020103	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1
Bialelo 01x 02	**1**	**1**	**1**	**1**	**0**	**0**	**1**	**1**	**0**	**1**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**1**

Tabla 4: Construcción de un bialelo.

Como podemos ver, el bialelo se formó sumando verticalmente los elementos de los alelos originales, utilizando la regla establecida en el párrafo anterior. Notemos que el primer alelo de este pequeño ejemplo pertenece a la clase 01 y el segundo a la 02. Por esta razón, diremos que el bialelo que forman pertenece a la *biclase 01×02*. Es decir, llamaremos *biclase X×Y* al conjunto de los bialelos formados al sumar todos los alelos de la clase X con todos los de la clase Y, y como la suma es conmutativa, las biclases también lo son. Esto es, la biclase X×Y es equivalente a la biclase Y×X, pero por convención llamaremos a una biclase X×Y cuando $X \leq Y$. Además, si la clase X es de tamaño *m* y la clase Y de tamaño *n*, la biclase X×Y tiene *m x n* elementos (bialelos).

En resumen, el tercer paso de la transformación de los datos consiste en construir una matriz de bialelos y biclases a partir de la matriz que se obtuvo en el segundo paso. A continuación se muestra la matriz de bialelos que genera la matriz de alelos de la Tabla 3. Debido a que dicha tabla cuenta con 10 alelos y éstos generan 55 bialelos, sólo se presenta un resumen de la tabla generada a manera de ejemplo.

Posición	55		88			117				119				224			225		
Pregunta	**C**	**T**	**A**	**C**	**T**	**A**	**G**	**C**	**T**	**A**	**G**	**C**	**T**	**A**	**C**	**T**	**A**	**C**	**T**
01x 01	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1
01x 01	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1
01x 01	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	1
01x 01	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	1
01x 01	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	1
01x 01	1	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0
01x02	1	1	1	1	0	0	1	0	1	1	1	0	0	1	0	1	0	0	1
01x02	1	0	1	1	0	0	1	0	1	1	1	0	0	1	0	1	0	0	1
01x02	1	1	1	1	0	0	1	1	0	1	0	0	0	1	0	1	0	0	1
01x02	1	1	1	1	0	1	1	0	0	1	0	1	0	1	0	1	0	0	1
01x02	1	1	1	1	0	0	1	0	1	1	1	0	0	1	0	1	0	0	1
01x02	1	0	1	1	0	0	1	0	1	1	1	0	0	1	0	1	0	0	1
01x02	1	1	1	1	0	0	1	1	0	1	0	0	0	1	0	1	0	0	1
01x02	1	1	1	1	0	1	1	0	0	1	0	1	0	1	0	1	0	0	1
01x02	1	1	1	1	0	0	1	0	1	1	1	0	0	1	0	1	0	1	1
01x02	1	0	1	1	0	0	1	0	1	1	1	0	0	1	0	1	0	1	1
01x02	1	1	1	1	0	0	1	1	0	1	0	0	0	1	0	1	0	1	1
01x02	1	1	1	1	0	1	1	0	0	1	0	1	0	1	0	1	0	1	1
⋮	⋮																		
11x43	1	0	0	1	1	0	1	0	0	1	0	0	1	0	1	1	1	0	0
11x43	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	1	1	0	0
43x43	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0

Tabla 5: Tercer y último paso de la transformación: Matriz de bialelos.

Para elaborar esta matriz, primero se genera la biclase 1×1, luego la 1×2,..., 1×43. Después se genera la biclase 2×2,..., 2×43. Se continúa de este modo hasta llegar a la última biclase, la 43×43.

La información que aparece en la Tabla 5 pone de manifiesto las características del problema que nos ocupa. Por un lado, podemos observar que los datos a considerar aumentaron al formar todos los bialelos posibles. Además, éstos fueron agrupados en biclases cuyos nombres se refieren a las dos clases de las que provienen. Así, al identificar

la biclase a la que pertenece un bialelo, se posee toda la información necesaria para establecer la compatibilidad que estamos buscando. Por otro lado, se toma en cuenta la forma en que se hacen las preguntas, aceptando por respuestas 1 (Si) y 0 (No), como sucede en la realidad.

Por último, es importante observar que en nuestro modelo se tomó en consideración la pérdida de información que se explicó anteriormente. Esto se demuestra con el siguiente ejemplo. Analicemos el tercer y quinto bialelos de la biclase 01×01 en la Tabla 5. Como podrá notar el lector, estos bialelos son idénticos. Sin embargo, ellos fueron construidos usando pares de alelos distintos. Es decir, si se hicieran todas las preguntas que aparecen en la tabla (19 en total), las respuestas arrojadas no nos permitirían distinguir si se trata de un bialelo o el otro. En este caso, la pérdida de información no es importante, pues ambos bialelos pertenecen a la misma biclase y, según lo que se explicó en la sección 2.1, estos bialelos son equivalentes; no es necesario distinguirlos, ya que provienen del mismo par de clases. No obstante, pueden presentarse casos en que dos bialelos idénticos provengan de pares de clases distintos, lo cual representa un problema de ambigüedad.

Debido a la cantidad de información con la que contamos para nuestro problema, fue necesario recurrir a la programación computacional para construir rápidamente la matriz de bialelos total. Esta matriz tiene 251(252)/2 = 31626 bialelos agrupados en 21(22)/2 = 231 biclases y 265 columnas. El programa fue hecho en el lenguaje de programación JAVA y el código aparece en la sección 3.3 de esta tesis.

Consideraremos como solución al modelo matemático al conjunto de preguntas que clasifiquen cualquier bialelo en su biclase. Es decir, debemos seleccionar un conjunto de preguntas dentro de las 265 posibles. Los problemas de este tipo, en los que se debe elegir entre un número finito de opciones se ubican, como nuestro problema, dentro del campo de la optimización discreta [6]. Como es frecuente hacer en los problemas de optimización discreta, emplearemos métodos heurísticos para encontrar una solución, o sea una lista de preguntas adecuada [3].

Hasta ahora sólo hemos explicado cómo se modeló matemáticamente el problema, transformando la información original. En el próximo capítulo se darán a conocer los métodos heurísticos empleados para solucionar el problema de obtener una lista de preguntas que identifiquen o clasifiquen el par de alelos (bialelo) de un ser humano.

2.3 CONTENIDO DE LOS SIGUIENTES CAPÍTULOS

Contamos en este momento con el planteamiento del problema biológico que se hizo en la sección 2.1. Además, en la sección 2.2 describimos el modelo matemático que se empleará para resolver el problema de esta tesis.

En la sección 3.1 hablaremos sobre la solución inicial al problema de clasificación que nos ocupa. Después, en la sección 3.2, mejoraremos esta solución inicial mediante una lista de preguntas que llamaremos la primera etapa y que representa una solución parcial al problema. También construiremos una matriz de biclases en la sección 3.3, que nos ayudará

a compactar la información para hacerla más manejable. El código del programa utilizado para crear esta matriz de biclases y el diagrama de flujo para explicarlo aparecen también en la sección 3.3. Luego, en la sección 3.4, generaremos una segunda etapa de preguntas que complementará a la primera, alcanzando una solución a nuestro problema.

En el capítulo 4, describiremos los pasos que deben seguirse para aplicar los resultados de esta tesis a la clasificación de códigos genéticos. En la sección 4.1 se explicará cómo emplear las preguntas de la primera etapa. Y en la sección 4.2 se darán criterios para pasar a la segunda etapa y escoger conjuntos de preguntas en la misma.

En el capítulo 5 se exponen las conclusiones generales de este trabajo de tesis. Y, por último, en el apéndice se anexan algunas de las matrices obtenidas durante el desarrollo de esta investigación además de la base de datos como apoyo al capítulo 4.