Am. J. Hum. Genet. 41:331–349, 1987

Analysis of Genetic Interrelationship among HLA-associated Diseases

HAYDEH PAYAMI,* MUHAMMAD A. KHAN,† DAVID M. GRENNAN,‡
PAUL A. SANDERS,‡ PHILIP A. DYER,‡ AND GLENYS THOMSON*

*Department of Genetics, University of California, Berkeley; †Department of Medicine, Case Western Reserve University at Metropolitan General Hospital, Cleveland; ‡Rheumatic Disease Center, Hope Hospital, Salford, England

SUMMARY

We have developed a method to study the genetic relationship between any two HLA-associated diseases. We have considered the following hypotheses: (1) both diseases are caused by a common allele; (2) different alleles at the same locus predispose to the two diseases; (3) one disease is predisposed by two alleles, one of which can also lead to the second disease; and (4) different HLA-linked loci are involved in the etiology of each disease. For each hypothesis, we have derived the expected HLA haplotype–sharing distribution in sib pairs who are affected with two diseases. The comparison of the expectations indicate that, in many cases, the alternate hypotheses can be distinguished, if the sample size is appropriately large. The knowledge of the mode of inheritance of each disease is not usually necessary; however, it can greatly increase the power of the test. Analyses of data on pairwise combinations of rheumatoid arthritis (RA), autoimmune thyroid disease (ATD), and insulin-dependent (type I) diabetes mellitus (IDDM) suggest that (a) IDDM is predisposed by two HLA-linked alleles, one of which also predisposes to ATD, (b) one of the IDDM alleles also confers susceptibility to RA, and (c) although the HLA-linked susceptibilities to RA and ATD appear to be primarily due to distinct alleles, the ATD allele may also have a minor role in predisposition to RA.

Received July 14, 1986; revision received February 26, 1987.

Address for correspondence and reprints: Dr. Haydeh Payami, Department of Medical Genetics, L103, Oregon Health Sciences University, Portland, OR 97201.

INTRODUCTION

A number of diseases are not only associated with the same HLA antigen but also tend to cluster in families or individuals and are suspected to have related etiologies. Among such clusters are the B27-associated rheumatic disorders, notably ankylosing spondylitis and Reiter disease (Brewerton 1978; Parker 1980); DR4-associated insulin-dependent diabetes mellitus (IDDM) and rheumatoid arthritis (RA); and DR3-associated autoimmune diseases, including IDDM, autoimmune thyroid disease (ATD), and myasthenia gravis (Farid et al. 1980; Kahn and Flier 1980).

The question arises whether these clusters are results of common genetic predispositions. A common population association with an HLA antigen may be due either to a common susceptibility allele or to different alleles that are in linkage disequilibrium with the same HLA marker. Diseases showing different HLA association may be predisposed to by different alleles that may or may not be at the same locus. From these possibilities, we have developed a method to aid in an understanding of the basis of the disease clustering. The present paper presents the theoretical aspects of the method and its application to study the underlying genetic interrelationship among IDDM, RA, and ATD.

METHOD

The method is based on the HLA haplotype–sharing distribution in sib pairs in whom two diseases of interest have been diagnosed. The expected distribution of the parental HLA haplotypes in such sib pairs depends on the underlying genetic basis of the two diseases. We have considered four different genetic models, and for each model we have derived the expected haplotype-sharing distributions for the three possible disease states of affected sib pairs—namely, (1) that one sib is affected with the two diseases and the other with only one (type AB-A), (2) that both sibs are affected with the two diseases (type AB-AB), and (3) that each sib is affected with only one disease (type A-B). In all four genetic-interrelationship models, the disease locus is linked to HLA.

Model I assumes that a common allele, denoted by D, confers susceptibility to both diseases. We denote by d all other alleles at that locus. We consider intermediate modes of inheritance—that is, dd individuals are assumed to be not disease susceptible—but the penetrance of the Dd genotype can range from zero (recessive) to being equal to the penetrance of DD (dominant). We denote the penetrances of the genotypes DD and Dd by f_2 and f_1, respectively, for disease A and by g_2 and g_1, respectively, for disease B (table 1).

Model II assumes that disease A is predisposed by allele D_1, with penetrance values f_2 and f_1 for the genotypes D_1D_1 and D_1- ($-$ denotes absence of D_1). Disease B is predisposed by another allele at the same locus, D_2, with penetrance values g_2 and g_1 for the genotypes D_2D_2 and D_2- (table 1).

Model III is an extension of models I and II. It is based on the assumption that the two diseases are predisposed to by a common allele and that another allele at the same locus also predisposes to one (but not to the other) disease. This hypothesis is based on (1) the observed associations of DR3 with ATD, of

TABLE 1

PENETRANCES OF THE TWO DISEASES A AND B FOR THE APPROPRIATE GENOTYPES

DISEASE DESIGNATION	GENOTYPE PENETRANCES									
MODEL I										
	DD	Dd	dd							
A	f_2	f_1	0							
B	g_2	g_1	0							
MODEL II										
	D_1D_1	D_1D_2	D_2D_2	D_1d	D_2d	dd				
A	f_2	f_1	0	f_1	0	0				
B	0	g_1	g_2	0	g_1	0				
MODEL III										
	D_1D_1	D_1D_2	D_2D_2	D_1d	D_2d	dd				
A	f_2	f_1	0	f_1	0	0				
B	g_5	g_4	g_2	g_3	g_1	0				
MODEL IV										
	DE/DE	DE/De	De/De	DE/dE	DE/de	De/dE	De/de	dE/dE	dE/de	de/de
A	f_2	f_2	f_2	f_1	f_1	f_1	f_1	0	0	0
B	g_2	g_1	0	g_2	g_1	g_1	0	g_2	g_1	0

DR4 with RA, and of both DR3 and DR4 with IDDM and (2) the elevated risk of IDDM for the DR3/DR4 genotype. We denote by D_1 the allele that predisposes to both diseases A and B and denote by D_2 the allele that predisposes to B only. There are three different genotypes that can lead to disease A, and five that can lead to disease B. The genotypes and their respective penetrances for each disease are given in table 1. This model allows for synergistic interaction of the D_1 and D_2 alleles in predisposing to disease B.

Model IV is based on the hypothesis that the two diseases are predisposed to by alleles at different loci. We denote by D the allele that predisposes to A (d = all other alleles of the D locus) and denote by E the allele that predisposes to B (e = all other alleles of the E locus). The penetrance scheme is given in table 1. The four possible haplotypes and their frequencies (x) are as follows:

$$\begin{aligned} DE \ldots . \ x_1 &= p_D p_E + \Delta \\ De \ldots . \ x_2 &= p_D p_e - \Delta \\ dE \ldots . \ x_3 &= p_d p_E - \Delta \\ de \ldots . \ x_4 &= p_d p_e + \Delta \end{aligned}$$

where Δ is the coefficient of linkage disequilibrium between D and E. There are three specific cases of Δ values that are of interest:

1. If $\Delta = 0$, then alleles D and E assort randomly and the inheritance of one disease is independent of the other. This can be considered as the control case, in which there is no genetic relationship (either positive or negative) between the two diseases.

2. If $x_2 = x_3 = 0$, then $p_D = p_E$ and Δ takes its maximum positive value; in this case, $\Delta_{max} = p_D p_e = p_d p_E$ and the alleles D and E always appear together. This special case of Δ_{max} corresponds to model I, with the frequency of the disease-susceptibility haplotype of model IV (x_1) being equivalent to the common-susceptibility-allele (i.e., disease-allele) frequency (p_D) in model I.

3. If $x_1 = 0$, then Δ takes its maximum negative value ($\Delta_{max} = p_D p_E$) and this special case corresponds to model II (different alleles of the same locus) with the frequencies of the disease-susceptibility haplotypes of model IV (x_2 and x_3) being equivalent to the disease-allele frequencies p_{D_1} and p_{D_2} of the two disease alleles in model II.

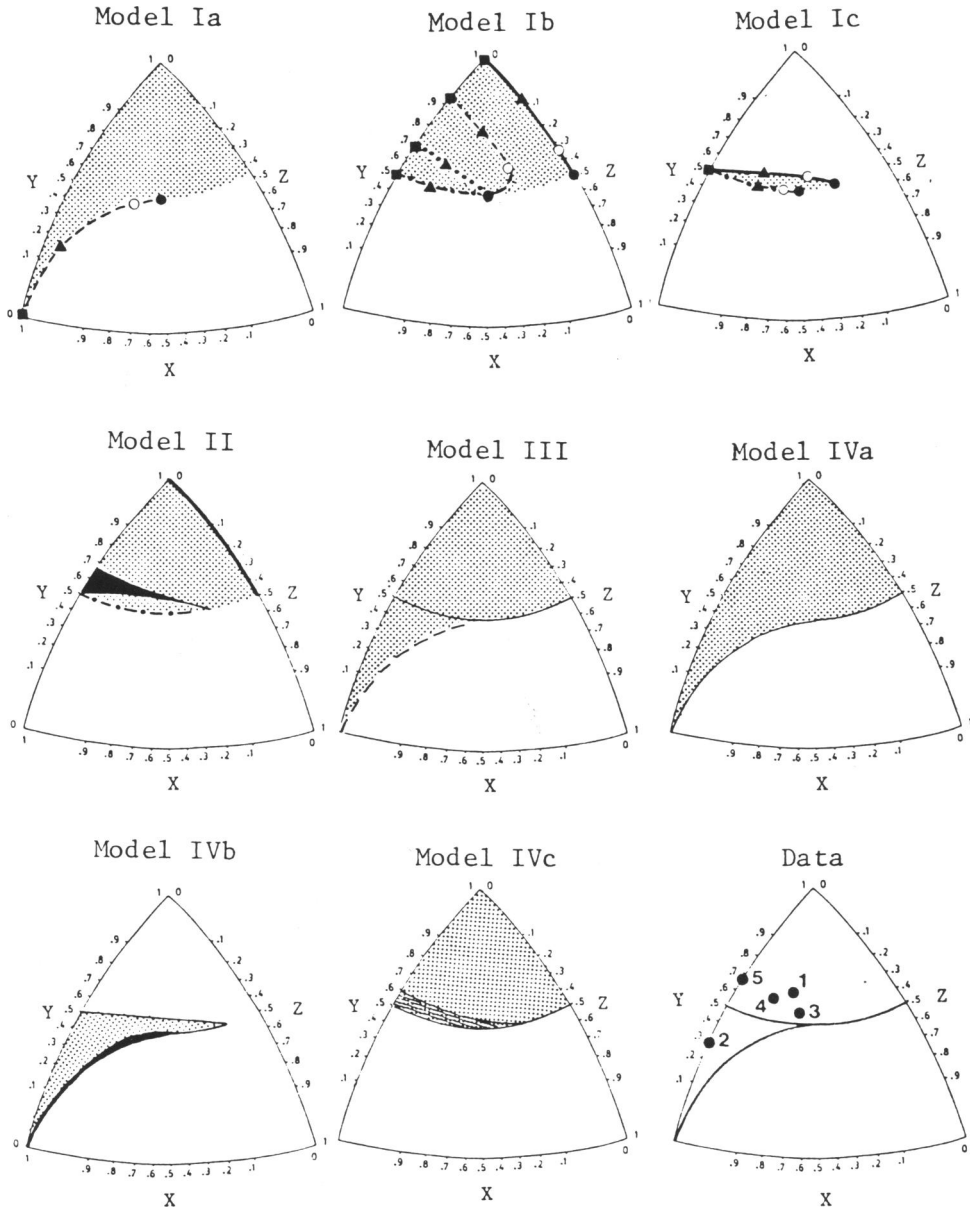
The expected haplotype-sharing distributions in affected sib pairs were derived conditional on parental mating types and based on the penetrance schemes outlined in table 1. The general derivation procedure and, as an example, the equations for model II, are given in the Appendix. (The derivations for other models will be made available on request.) The expected haplotype-sharing proportions were calculated for a range of disease-allele frequencies and penetrance values and plotted on trinomial graphs. We will present for each combination of disease status of affected sib pairs the expected distributions under the different genetic models. We will define the total expectation space for each case and discuss in more detail the special cases in which the diseases A and B follow recessive, additive, or dominant modes of inheritance. The expectations under different models will then be compared to determine the discriminatory power of this method. Data analysis by this method simply involves plotting the data against the expectations for each genetic model and using 2-SE circles (Edwards 1971) for the significance of deviation (see Application below).

ONE SIB AFFECTED WITH A AND B, THE OTHER WITH A ONLY (Type AB-A)

Model I (Common Allele)

For this case, the haplotype-sharing distribution is a function of p_D, the mode of inheritance of disease A ($\lambda_A = f_1/f_2$), and the absolute values of g_2 and g_1. The probabilities of sharing two, one, and zero haplotypes (denoted by X, Y, and Z, respectively) are plotted in figure 1 (*Model Ia*), where the shaded area repre-

FIG. 1.—The expectations for AB-A sibs under models I–IV. X, Y, and Z values are plotted on the X, Y, and Z axes, respectively. *Model Ia:* All expectations fall in the dotted space; when A is recessive, the distribution (broken curve) is only a function of p_D. If A is additive and B is recessive (*Model Ib*) or additive (*Model Ic*), the distribution is a function of p_D and g_2. p_D values: 1 (●); .5 (○); .1 (▲); and 0 (■). g_2 values: 1 (———); .9 (— — —); .5 (.....); and 0 (—·—·—). *Model II:* In this case neither disease can be recessive. In other cases the expectations fall on or above the Y



= .5 curve. The space is defined for when A is additive for g_1 values ranging from 0 (—·—·—) to 1 (———). For $0 < g_1 < 1$, each value forms a space determined by p_D, e.g., the shaded space represents $g_1 = .5$. *Model III:* When A is recessive, the distribution is only a function of the frequency of the allele that predisposes to both A and B (broken curve). In other cases, a given mode of inheritance of A forms a space determined by the penetrances of B and the p_D values, e.g., if A is additive, the expectations occupy the space above the Y = .5 curve. *Model IVa:* Total distribution; when A and B are both recessive, the expectations occupy all of this space. *Model IVb:* The distribution when A is recessive and B is additive (shaded) or dominant (dotted). *Model IVc:* The distribution when A is additive and B is recessive or dominant (dotted) or additive (shaded). *Data:* The haplotype-sharing frequencies in IDDM,ATD-ATD (point 1), IDDM,ATD-IDDM (point 2), RA,ATD-RA (point 3), RA,ATD-ATD (point 4), and RA,IDDM-RA (point 5) sib pairs.

sents the total expectation space for all values of p_D, $0 \leq \lambda_A \leq 1$, and $0 \leq (g_2, g_1) \leq 1$ with $g_2 \geq g_1$. Subregions of the space that present the expectation for specific values of p_D, λ_A, g_2, and g_1 have been defined in detail.

When the shared disease A is recessive ($\lambda_A = 0$), the expected haplotype-sharing distribution is independent of the mode of inheritance of the unshared disease and is only a function of p_D. This is expected, since the recessive mode of inheritance of the shared disease means that all the affected sibs must be homozygous DD and whether they express disease B is irrelevent to the haplotype-sharing distribution. This distribution forms a single curve determined by $4XZ \geq Y^2$, $X \geq .25$ (fig. 1, *Model Ia*). (This curve is the same as the recessive expectations curve in the affected-sib-pair method when only one disease is under study [Louis et al. 1983].) If $p_D < .5$, the value of X is always greater than the value of Y or Z.

The haplotype-sharing distribution when A is additive ($\lambda_A = .5$) and B is recessive is a function of p_D and g_2 of the unshared recessive disease (see fig. 1, *Model Ib*). This space is defined by $Y \geq .5$, $X \leq .5$. One of the main features of the distribution is that the expected frequency of Y is $>.5$, unlike the case above in which A is recessive, in which Y is always $<.5$. The larger the g_2 of the unshared recessive disease, the larger the expected value of Y.

If both diseases follow an additive mode of inheritance, then the value of X is expected to range from .1 to .5 (depending on p_D and the penetrance of B) whereas the value of Y is always close to .5 ($.5 \leq Y \leq .58$), as shown in figure 1 (*Model Ic*). This distribution is in practice indistinguishable from that occurring when A is additive and B is dominant (in which case $Y = .5$ and $.5 \geq X \geq .25$) or when both A and B are dominant ($.48 \leq Y \leq .5$ and $.5 \geq X \geq .25$).

Model II (Different Alleles)

The haplotype-sharing distribution in this model is independent of the penetrance of the unshared disease B in D_2D_2 individuals, since the D_2D_2 genotype cannot confer susceptibility to the shared disease A. Thus, the distribution is only a function of the mode of inheritance of A (measured by λ_A), g_1, and p_{D_1} and p_{D_2} and is independent of the mode of inheritance of B.

Depending on the disease parameters, the value of X may range from 0 to .5 and that of Y may range from .46 to 1 (fig. 1, *Model II*). The expected distribution when A is additive is given in detail in figure 1 (*Model II*). The general form of haplotype sharing for all modes of inheritance of A except recessive is similar to that when A is additive, with a general decrease in the value of X for smaller λ_A values. In general, when g_1 is low or the frequency of p_{D_1} is high, the value of X is expected to be smaller.

For model II, it is not possible to observe sib pairs of the type AB-A if either the shared disease A or the unshared disease B or both are recessive. Since one sib must have both diseases, this individual's genotype must be D_1D_2 (see table 1), so only diseases that have a nonzero penetrance in heterozygous individuals can form sib pairs of type AB-A under model II.

Model III (A Common Allele and a Second Allele for B)

When the unshared disease B is predisposed by two alleles, D_1 and D_2 (allowing for synergistic interaction), and the shared disease A is predisposed by D_1 only, the haplotype-sharing distribution in AB-A sib pairs is a function of the mode of inheritance of A ($\lambda_A = f_1/f_2$), p_{D_1} and p_{D_2}, and the penetrance of B for genotypes that have the D_1 allele (g_5, g_4, and g_3). (Since both sibs are affected with A, neither one can have the D_2D_2 or D_2d genotype; therefore, the distribution is independent of g_2 and g_1). We have examined the expectations when A is recessive, additive, or dominant.

When A is recessive, the haplotype sharing is independent of the disease B parameters. The distribution is identical to that expected under model I: it forms a single curve as a function of p_{D_1} (fig. 1, *Model II*).

Assuming an additive (or dominant) mode of inheritance for A, we considered $.001 \leq g_4 \leq .9999$ with $g_4 \geq g_5$ for the three cases of (1) $g_3 = 0$ (B recessive), (2) $g_3 = g_5/2$ (B additive), and (3) $g_3 = g_5$ (B dominant). In all three cases $X \leq .5$ and $Y \geq .5$ (or, if A is dominant, $Y \geq .46$) as shown in figure 1 (*Model III*).

Model IV (Different Loci)

The haplotype-sharing distribution in this case is a complex function of the following six independent factors: the mode of inheritance of A, g_2 and g_1, p_D and p_E, and Δ. Owing to the complexity of this model, the expectations were calculated by numerical analysis, assuming that $\Delta = 0$. (For positive values of Δ_{max}, see model I; for negative values of Δ_{max}, see model II.)

Figure 1 (*Model IVa*) shows the space that all expected haplotype-sharing points occupy when the shared disease A is recessive. We have considered, in more detail, the special cases in which the unshared disease B is recessive, additive, or dominant. When B is recessive, depending on the penetrance values and p_D, the expectation may fall anywhere in the space shown in figure 1 (*Model IVa*). When B is dominant, the expected points occupy only a subregion of this space, in which $Y < .52$ (fig. 1, *Model IVb*). If B is additive, the expectation space is reduced drastically, as shown in figure 1 (*Model IVb*).

If A is additive and B is either recessive or dominant, X can range from .25 to .5 and Y can range from .5 to 1. If both diseases are additive, however, the ranges are reduced to $.5 \leq Y \leq .6$ and $.16 \leq X \leq .5$ (fig. 1, *Model IVc*).

The haplotype sharing for the cases in which A is dominant are very similar to those for the additive case, with the exception that the lower limit for Y is .48 instead of .5.

Distinguishability of the Models

Depending on the observed frequencies of haplotype sharing in AB-A sib pairs, it may be possible to discriminate between the different alleles and a common-allele model even if the modes of inheritance of the two diseases are unknown. For example, if the frequencies X and Y in affected sib pairs AB-A

are 0.9 and 0.1, respectively, then the data would be compatible with the expectations for model I but not with those for model II (fig. 1, *Model Ia*, *Model II*), and a sample size of 10 sib pairs would be adequate to statistically reject the different-allele hypothesis, using the 2-SE circles. The prior knowledge of the mode of inheritance of at least one disease can increase the discriminatory power of the analysis.

When disease A and/or disease B is recessive, no sib pairs of type AB-A can be found under model II. Therefore, the hypothesis of different alleles must be rejected if at least one of the diseases is known to be recessive and sib pairs of the type AB-A are observed.

If the shared disease A is additive, then the power of discrimination depends on the mode of inheritance and the penetrance of the unshared disease B. If B is close to recessive, then for model II $Y = \sim.5$ and $.25 < X < .50$ (fig. 1, *Model II*), whereas for model I X ranges from .5 when g_2 is low to 0 when g_2 is very large (fig. 1, *Model Ib*). Conversely, when the unshared disease B is additive with low penetrance, the expected distribution under model II is quite distinct from those under model I, a difference that allows discrimination between the two hypotheses. If both A and B are known to be dominant and B is highly penetrant, again the two models can be easily distinguished (see above).

Model III is an extension of models I and II; therefore it is not possible to overrule model III in favor of model I or model II. However, the reverse may be possible. If nothing is known about the predisposition to the individual diseases, it is still possible to overrule model II (but not model I) in favor of model III, if $Y < .5$ (fig. 1, *Model II, Model III*). Owing to the complexity of model III, the knowledge of the mode of inheritance of the shared disease may not by itself increase the power of the test. But the values of the penetrance(s) or p_D(s) can aid in specifying the appropriate expectation subregions of the three models, subregions that may be distinct and allow differentiation of model III from models I and II.

When testing the different susceptibility-allele models, the one-locus hypothesis (model II) and the two-loci hypothesis (model IV) may be distinguished even in the absence of any information about the modes of inheritance. If the two alleles are at the same locus, X is always $\leq .5$, whereas if two loci are involved, the value of X may be as great as 1 (fig. 1, *Model II*, *Model IVa*). If at least one disease is known to be recessive, then model II must be rejected if sib pairs of type AB-A are observed; however, sib pairs of this type can be formed under model IV (fig. 1, *Model II*, *Model IVb*). If both diseases are additive or dominant, high values of Y will lead to the rejection of model IV but not of model II (fig. 1, *Model II*, *Model IVc*).

The total expectation space for model IV is almost completely overlapping with those for models I and III (fig. 1, *Model Ia*, *Model III*, *Model IVa*); therefore it would not be possible to distinguish these models, unless some of the disease parameters are known. For example, if the shared disease is known to be recessive, the majority of the expected points for model IV would be very distinct from those for models I and III, a result that would allow differentiation of model IV from the latter models (fig. 1, *Model Ia*, *Model III*, *Model IVa*).

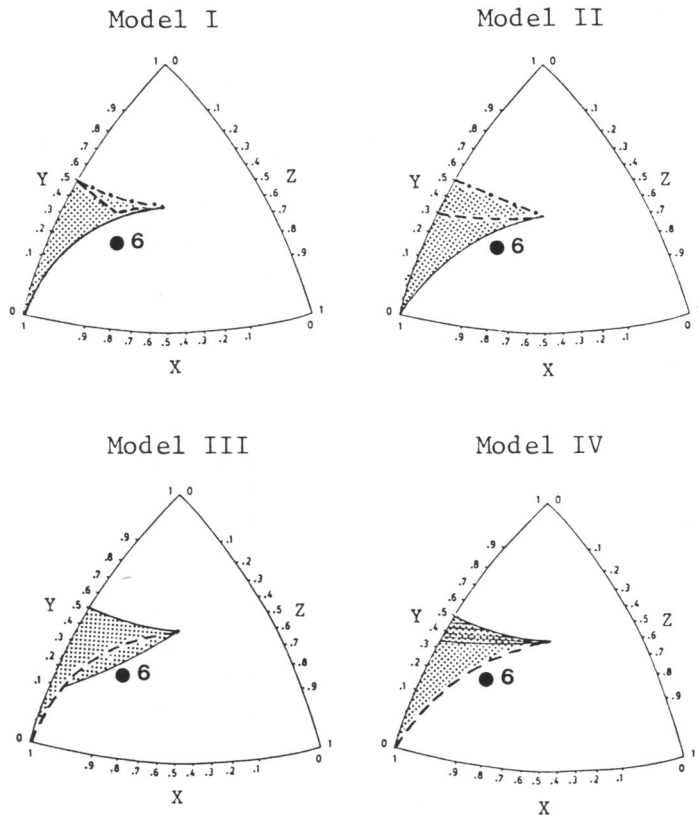


FIG. 2.—The expected distribution for AB-AB sibs under models I–IV. The data point for IDDM,ATD-IDDM,ATD (point 6) is plotted directly against the expectations. *Model I:* The total expectation space (dotted); when A and/or B is recessive (———); A and B are additive (— — —); or one disease is additive and the other is dominant (—·—·—). *Model II:* The distribution is only a function of p_{D_1} and p_{D_2}. $p_{D_1} = p_{D_2}$ (———); $p_{D_1} + p_{D_2} = 1$ (— — —); $p_{D_1} = .4$ with $p_{D_2} = 0$–.6 (—·—·—). *Model III:* The expectations when A is recessive (broken curve) and A is dominant or additive (dotted space). *Model IV:* The expectation when both A and B are recessive (broken curve), A is recessive and B is additive (dotted space), and when both A and B are additive (shaded space).

Similarly, if both diseases are additive, there is considerable discriminatory power between models III and IV—in this case, against model IV.

BOTH SIBS AFFECTED WITH A AND B (Type AB-AB)

Model I (Common Allele)

In this case the haplotype-sharing distribution is a symmetric function of the modes of inheritance of the two diseases and p_D. For each set of modes of inheritance, a single curve is formed as a function of p_D (fig. 2, *Model I*). The curve $4XZ = Y^2$, $X \geq .25$ represents the expectations when one or both diseases are recessive. The curve $Y = .5$, $X \geq .25$ represents the expectations when one disease is additive and the other is dominant. These two curves form

the boundaries of the space where the expected curves for all other modes of inheritance fall.

In general, for a given p_D, Y is lowest when one or both diseases are recessive and increases as a function of the modes of inheritance to the limit of .5 when one disease is additive and the other is dominant. However, when the modes of inheritance are ranged from additive/dominant to dominant/dominant, the value of Y is slightly decreased. The curve for the dominant/dominant expectation falls below, but is very close to, the additive/dominant curve.

Model II (Different Alleles)

The haplotype-sharing distribution in this case is only a function of the p_D values, and is independent of the modes of inheritance. However, sib pairs of type AB-AB cannot be formed under this model if one of the diseases is recessive, because it is not possible for an individual to be affected with both a recessive disease and another disease that requires predisposition by another allele at the same locus. The total expectation space is given in figure 2 (*Model II*). The boundaries are the expected curves for the cases in which $p_{D_1} + p_{D_2} = 1$ and $p_{D_1} = p_{D_2}$.

Model III (Common Allele and a Second Allele for B)

In this case the haplotype-sharing distribution is a function of the p_D values, the mode of inheritance of disease A, and the penetrances of D_1- genotypes for disease B. However, if A is recessive, then the genotypes of both sibs must be D_1D_1 and the haplotype-sharing distribution between them is only a function of p_{D_1} (fig. 2, *Model III*).

Owing to the complexity of this model, the expectations for the cases in which A is additive or dominant were calculated numerically, using values of .001–.9 for the p_D's and .001–1 for the penetrances of the D_1- genotype for disease B, with the penetrance of D_1D_2 being equal to or greater than the others (synergistic effect). The expectations for these two cases are very similar, as shown in figure 2 (*Model III*).

Model IV (Different Loci)

The expectations in this case are determined by the modes of inheritance of diseases A and B and the frequencies of the four possible haplotypes for the two loci, which are in turn a function of the p_D and Δ values.

We have considered the case in which $\Delta = 0$ (i.e., in which there is a random assortment of disease alleles) for the recessive, additive, and dominant modes of inheritance of A and B with p_D values ranging from 10^{-6} to .9. (For positive values of Δ_{max}, see the distribution under model I; for negative values of Δ_{max}, see model II.)

If both diseases are recessive, the distribution is only a function of x_1, forming the curve given by $4XZ = Y^2$, $X \geq .25$ (fig. 2, *Model IV*). If one disease is recessive and the other additive, the expectations fall in the space bounded by the $4XZ = Y^2$, $X \geq .25$ and $Y = .5$, $X \geq .25$ curves. (A similar space is

obtained for the recessive/dominant case). When both diseases are additive or dominant, the expectations occupy a small region where $.32 \leq Y \leq .5$ and $.25 \leq X \leq .68$ (fig. 2, *Model IV*).

Distinguishability of the Models

It is difficult to distinguish between model I and model II by means of AB-AB sib pairs, unless the modes of inheritance are known. If one of the diseases is recessive, sib pairs of type AB-AB cannot form under model II. If the two diseases have modes of inheritance close to additive, there is considerable discriminatory power between the expectations of the two models if p_{D_1} and p_{D_2} for model II are small and close in value (fig. 2, *Model I, Model II*). In other cases it may be difficult to distinguish between the expectations of the two models by means of AB-AB sib pairs.

The total expectation spaces under model III and model IV overlap considerably (fig. 2, *Model III, Model IV*). If it is known that only A is recessive, then there is considerable discriminatory power against model III. Similarly, if both diseases are known to be additive or dominant, the majority of the expectations under model III are distinct from those under model IV and allow discrimination against the latter model.

As discussed earlier, it is not possible to reject model III without rejecting models I and II. However, it is possible to reject model I in favor of model III when the disease A that is predisposed by only one allele is additive or dominant (fig. 2, *Model I, Model III*).

The total expectation spaces for models II and IV are similar. However, certain modes of inheritance form distinct distributions. For example, if both diseases are additive or dominant, X under model II could be as great as 1, whereas for model IV it cannot be $>.68$. If one disease is recessive, sib pairs of type AB-AB cannot form under model II, whereas they can under model IV.

EACH SIB AFFECTED WITH A DIFFERENT DISEASE (Type A-B)

Model I (Common Allele)

The haplotype-sharing distribution for this model is a symmetric function of the penetrances of the two diseases in homozygotes and heterozygotes, and of p_D.

When both diseases are recessive, the distribution is only a function of p_D. If $p_D < .5$, then X is expected to be greater than Y or Z (fig. 3, *Model Ia*).

If one disease is recessive and the other is additive, the haplotype-sharing distribution falls in the space defined by $Y \geq .5$, $X \leq .5$ (fig. 3, *Model Ib*). In this case higher values of Y are expected when the penetrance of the recessive disease is high.

When both diseases are additive, the haplotype-sharing distribution falls in a small space where $Y = .42–.5$ (fig. 3, *Model Ib*). When both diseases are dominant, then the X, Y, and Z values fall below but very close to the $Y = .5$, $X \geq .25$ curve.

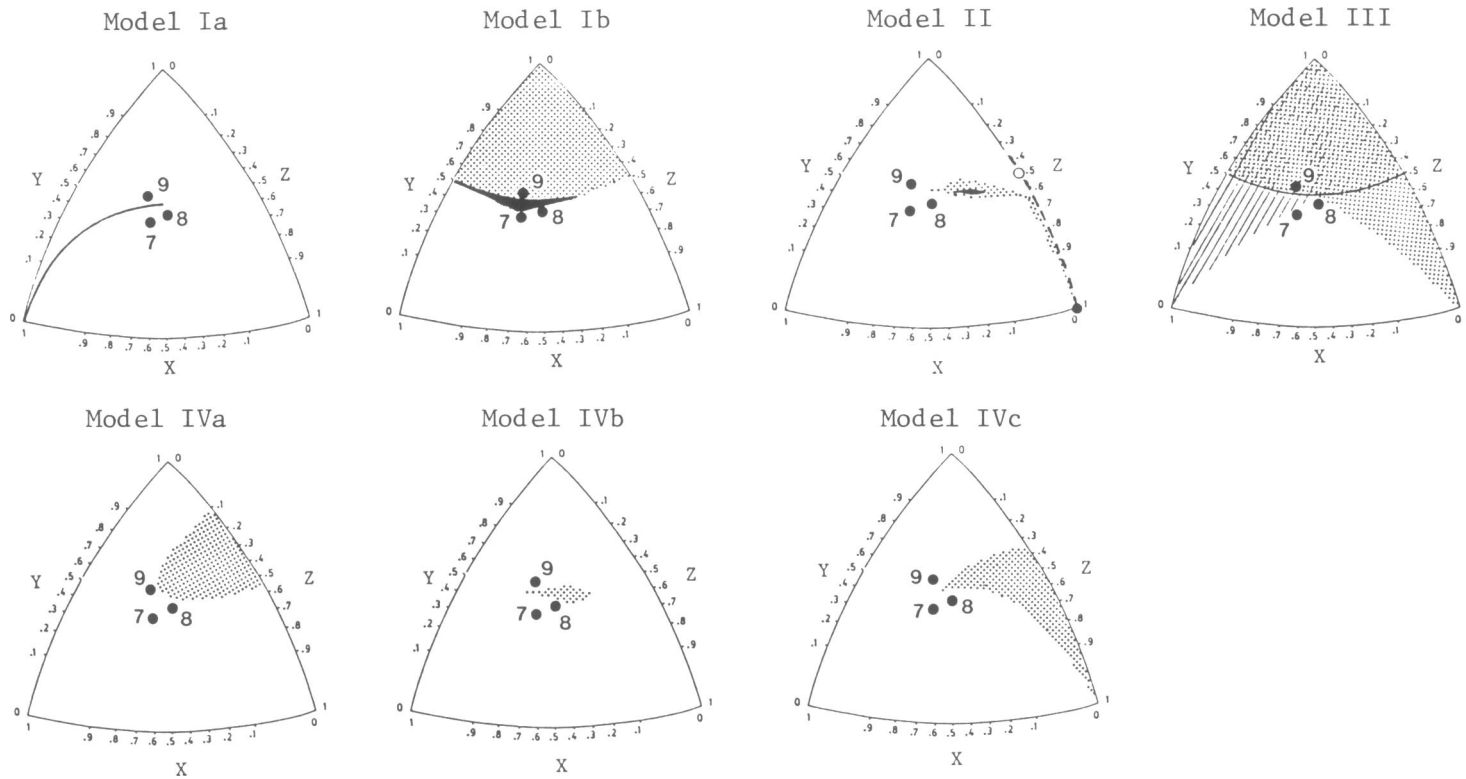


FIG. 3.—The expectations for A-B sibs under models I–IV. The data points for IDDM-ATD (point 7), RA-ATD (point 8), and IDDM-RA (point 9) are plotted against the expectations. *Model Ia:* The expectations when both diseases are recessive. *Model Ib:* The expectation when one disease is recessive and the other additive (dotted space) and when both are additive (shaded area). *Model II:* The broken curve includes all expectations when one disease is recessive. If the other disease is also recessive, then $Z = 1$ (●); if it is additive, then $Y = Z = .5$ (○). The shaded area represents the situation when both diseases are additive. The dotted area includes all other expectations. *Model III:* The expectations form the shaded area if A which is predisposed by only one allele is recessive, the dotted area if A is dominant. The overlapping region (shaded and dotted) includes all expectations for the situation when A is additive. *Model IVa:* The expectation when one disease is recessive. *Model IVb:* The expectation when both diseases are additive. *Model IVc:* The expectation when both diseases are dominant.

Model II (Different Alleles)

This distribution is a function of p_{D_1} and p_{D_2} and of the penetrances of the two diseases in homozygotes and heterozygotes.

If both diseases are recessive, then the genotypes of the individuals affected with A or B are D_1D_1 or D_2D_2, respectively. Therefore, affected sib pairs of type A-B cannot share any haplotypes; that is, $Z = 1$.

If one of the diseases, say A, is recessive, then the affected sib pairs of type A-B can share only one or no haplotype under model II, regardless of the mode of inheritance of B. This is expected, since, because the sib that is affected with the recessive disease must carry two copies of the disease-susceptibility allele and the one affected with the other disease must carry at least one copy of the other disease-susceptibility allele, the sib pairs therefore cannot share two haplotypes. For example, if the other disease is additive, half of the sib pairs are expected to share one haplotype and the other half are expected to share none (independent of p_D). Similarly, if A is recessive and B is dominant, more than one-half of the sib pairs would be expected to share one haplotype and the rest would be expected to share none (the exact value of Y and Z depend on the p_D of the dominant disease) (fig. 3, *Model II*). When both diseases are additive, the haplotype-sharing distribution occupies a small space, where $.059 \leqslant X \leqslant .167$ and $.471 \leqslant Y \leqslant .5$, (see fig. 3, *Model II*).

Owing to the complexity of this model, it is difficult to define the boundaries of the total haplotype-sharing space. Numerical analysis, however, indicates that the majority of X, Y, and Z points fall within or near the space defined for the case in which both diseases are additive. The distribution of the expected points (fig. 3, *Model II*) is obtained for $f_2,g_2 = 0$–1 (in increments of .2) and for $p_{D_1},p_{D_2} = 0$–1 (in increments of .2).

Model III (Common Allele and a Second Allele for B)

The expected haplotype-sharing distribution in A-B sib pairs is a function of nine parameters: the seven penetrances for diseases A and B and the two p_D's. We have calculated the haplotype-sharing probabilities for $p_D = .001$–$.9$, and for penetrances for diseases A and B ranging from .001 to 1, allowing for synergistic interaction of the two disease alleles for B.

The expectations for this model include nearly all possible haplotype-sharing configurations for sib pairs (fig. 3, *Model III*). If A is recessive, the expectations can range from $X = 1$ to $X = Z = .25$, $Y = .5$ to $Y = Z = .5$ to $Y = 1$. If A is additive, then Y is always $>.4$, whereas X and Z range from 0 to .5. When A is dominant, the haplotype sharing may range from $Y = 1$ to $X = Y = .5$ to $X = Z = .25$, $Y = .5$ to $Z = 1$.

Model IV (Different Loci)

We have used numerical analysis to examine this distribution for recessive, additive, or dominant modes of inheritance of each disease, considering penetrance values ranging from 10^{-6} to 1 and p_D values ranging from 10^{-6} to 1 and assuming that the two disease alleles assort independently ($\Delta = 0$). (For max-

imum positive and negative Δ values, see the distributions under models I and II.)

When one disease, say A, is recessive, the X is always $<.3$, whereas Y and Z are .42–.9 and .1–.58, respectively (fig. 3, *Model IVa*). If B is also recessive, then $.5 \leq Y \leq .9$; if B is additive, $.42 \leq Y \leq .82$; and if B is dominant, $.46 \leq Y \leq .72$.

If both diseases are additive, the expectations occupy a small space with $.1 \leq X \leq .4$ and $.4 \leq Y \leq .52$ (fig. 3, *Model IVb*). If one disease is additive and the other is dominant, X and Y can be 0–.34 and .32–.42, respectively. When both diseases are dominant, Z may be as great as 1 (fig. 3, *Model IVc*).

Distinguishability of the Models

The expectations under the common-allele and the different-alleles models (models I and II, respectively) are in many cases distinct and can be differentiated even if the modes of inheritance are unknown (fig. 3, *Model I, Model II*). In general, when the two diseases are predisposed by a common allele, the alternate model can be ruled out if the p_D is low. It is difficult to determine the exact conditions under which model I can be rejected when different alleles (model II) predispose to the diseases, except when either both diseases are recessive or one disease is recessive and the other follows a mode of inheritance that is closer to recessive than dominant. The distributions for these cases fall on the $X = 0$, $Y < .5$ curve, which is distinct from the model I expectations.

The total expectations under models I and III occupy the same space; therefore, only if the mode of inheritance of at least one disease is known can the two models be distinguished (fig. 3, *Model I, Model III*). (It is only possible to eliminate model I in favor of model III, not vice versa.) If both disease A and disease B are recessive (allowing for negative complementation for B), the haplotype-sharing distributions under the two models are quite distinct and can allow discrimination against model I. (The expectations under model I fall on a boundary of the large expectation space for model III). When both diseases are additive or dominant, for model I X is always $\geq .25$ and Y is very close to, but never greater than, .5. For model III, however, X may be as low as zero and Y can be between .33–.66. Therefore, a large region of the model III expectation space does not overlap with model I expectations and can be useful in discriminating between the two models.

Comparison of the different disease-allele models (models II and IV) shows that some regions of the model IV expectations do not overlap with model II expectations and can allow discrimination against the latter without prior knowledge of the modes of inheritance (fig. 3, *Model II, Model IV*). If the mode of inheritance of at least one disease is known, the power of discrimination increases greatly. For example, if both diseases are recessive and are predisposed by alleles of the same locus (model II), sib pairs of type A-B cannot share any haplotype. But if the disease alleles are at different loci, X can be as great as .3 and $Y = .5–.9$. Similarly, if one disease is recessive and the other is

TABLE 2

OBSERVED FREQUENCIES OF HAPLOTYPE SHARING IN CAUCASOID SIB PAIRS AFFECTED WITH COMBINATIONS OF IDDM, RA, AND ATD

SIB 1-SIB 2 CONDITION (*N*)	NO. OF HAPLOTYPES SHARED		
	2	1	0
1. IDDM,ATD-ATD (27)	0.22	0.70	0.08
2. IDDM,ATD-IDDM (7)	0.71	0.29	0.00
3. RA,ATD-RA (7)	0.29	0.57	0.14
4. RA,ATD-ATD (18)	0.28	0.67	0.05
5. RA,IDDM-RA (3)	0.33	0.67	0.00
6. IDDM,ATD-IDDM,ATD (9)	0.67	0.22	0.11
7. IDDM-ATD (10)	0.40	0.40	0.20
8. RA-ATD (20)	0.30	0.45	0.25
9. IDDM-RA (7)	0.29	0.57	0.14

NOTE.—Numbers on the left designate data sets (which are also plotted in figs. 1–3).

additive, under model II both Y and Z equal .5, whereas under model IV X may be as great as .34.

The expected haplotype-sharing distributions under models III and IV are in many cases distinct (fig. 3, *Model III*, *Model IV*). For all modes of inheritance of A and B that have been considered, model III can result in haplotype-sharing distributions that are not expected under model IV. If these distributions are observed, the hypothesis of different loci can be eliminated. In the special case in which both diseases are dominant, model IV can result in very high Z values, which are not possible under—and allow rejection of—model III.

APPLICATION

The method described in the present paper provides a simple test to differentiate alternate hypotheses of genetic relationship between any two HLA-associated diseases. The test involves plotting the observed haplotype-sharing frequencies in sib pairs, who have the diseases of interest, against the expectations under each hypothesis and then using Edwards's 2-SE circles to determine the significance of deviation. If none of the disease parameters are known, the total expectation spaces must be used for testing. On the other hand, if mode of inheritance, penetrance, and/or allele frequency for one or both diseases is (are) available, only the subregion of the total expectations that correspond to the known parameter(s) need be used.

We have pooled the data in the literature (Farid et al. 1980; Grennan et al. 1983; Bertrams et al. 1984; Torfs et al. 1986), as well as additional data from the laboratories of Khan and Grennan, on the haplotype-sharing frequencies in Caucasian sib pairs having combinations of IDDM, RA, and ATD. The data are shown in table 2 and plotted in figures 1–3. We have compared each data set with the respective expected distributions under the four models and tested the significance of the deviations by means of Edwards's 2-SE circles.

IDDM and ATD

If IDDM is predisposed by only one allele, as assumed in models I, II, and IV, then its mode of inheritance must be assumed to be almost strictly recessive (Payami et al. 1985). Model II must therefore be rejected because of the existence of the IDDM,ATD-IDDM sib pairs, sib pairs that would not be possible if the two diseases were predisposed by different alleles of the same locus one of which was recessive. The data are most compatible with model III: two alleles for IDDM, one of which also predisposes to ATD. An additive, or dominant, model for ATD, as suggested by Torfs et al. (1986), is compatible with this hypothesis. A recessive model for ATD is rejected under all models except model IV.

IDDM and RA

The two data sets are small (seven and three sib pairs) and do not allow rejection, with statistical significance, of any of the hypotheses. However, regardless of the sample size, model II must be rejected if either RA or IDDM is recessive, owing to the presence of RA,IDDM-RA sib pairs. Both data sets are more compatible with the hypotheses of a common-susceptibility allele (models I and III) than with the different-alleles hypotheses (models II and IV), especially if the suggested additive mode of inheritance and low penetrance of RA (Payami et al. 1986) are considered (points 5 and 9 in figs. 1, 3).

RA and ATD

All three data sets fall within or close to regions that are common for different models; therefore, discrimination is difficult. However, considering additive or dominant models for the two diseases, the RA-ATD data (point 8) is closer to models I and III, whereas the RA,ATD-ATD data (point 4) shows considerable deviation from model I but not from models II and III. These results, although not statistically significant, favor model III for RA and ATD. The compatibility of the data with model III implies that one allele predisposes to both diseases and that another allele predisposes to one of the diseases. If this hypothesis is true, then the population associations with different DR antigens would suggest that susceptibility to the disease that has two alleles is mostly due to the unshared allele.

The data are also compatible with some of the expectations under model IV. Therefore, until the expectation space is studied in more detail for specific sets of disease parameters, it is not possible to eliminate the possibility of different loci.

DISCUSSION

One possible explanation of the clustering of the HLA-associated diseases is that having one disease merely increases an individual's susceptibility to a second disease, there being no genetic association in the predisposition to the two diseases. In that case, the expected haplotype-sharing distribution is the same as that for different loci having no linkage disequilibrium. If one disease

increases the individual's susceptibility to the second disease, the distribution will still follow the expectations for the appropriate genetic model, with a higher penetrance for the second disease.

This method allows detection of a common-susceptibility allele in complex situations in which such information cannot be easily deduced from the patterns of disease-associated haplotypes. Alternatively, diseases that exhibit common population associations with HLA but are not predisposed by a common allele can be identified. Furthermore, the method shows that the expected haplotype-sharing distributions are more complex than some investigators have assumed. For example, Torfs et al. (1986) based their study of RA and ATD on the following assumptions: (1) that if both diseases are dominant and do not share a common allele, then the distribution in RA-ATD sib pairs should be random (.25, .5, .25) and (2) that the expected X value for RA,ATD-ATD sib pairs is less than that for the RA-ATD sib pairs. Our derivations reveal that both of these speculations are too simplistic and wrong.

The genetic-interrelationship models could also aid in the understanding of individual diseases. For example, in a number of cases, the expectations for the models considered here fall outside the limits of the haplotype-sharing space for when only one disease is considered (Louis et al. 1983). It follows, then, that if the haplotype-sharing values for a particular disease fall outside the single-disease-expectation limits, the deviation could be interpreted as being due to disease heterogeneity.

The observed distribution of IDDM and ATD data are in close agreement with the specific subregions of model III that correspond to (1) the suggested recessive mode of inheritance for the DR3-associated allele and the dominant mode of inheritance for the DR4-associated allele that predispose to IDDM (Louis and Thomson, in press; MacDonald et al., in press; Thomson et al., in press) and (2) a dominant (additive) model for the DR3-associated allele that predisposes to ATD.

Although both IDDM and RA are associated with antigen DR4, the increased frequency of the DR4.3 subtype among IDDM, but not among RA, patients (Tait et al. 1984) implies the existence of different disease-susceptibility alleles for the two diseases. Considering the compatibility of the sib-pair data with model III, we suggest that there are at least two alleles involved, one predisposing to IDDM and the other predisposing to both RA and IDDM. It is possible that in the study by Tait et al. (1984)—and perhaps in the general population—the IDDM patients predominantly carry the allele that does not predispose to RA, thus causing the putative common allele to not be detected. In the present study, on the other hand, there may be an overrepresentation of the common allele, since the families were probably selected for the joint occurrence of IDDM and RA.

The haplotype-sharing distributions in sib pairs with RA and ATD are closer to the expectations under model III than to those under models I and II. This preliminary result suggests that although the two diseases exhibit strong population associations with different HLA antigens, there may be a common component in their etiologies. It has recently been suggested that, in addition to the

DR4 association, a DR3- and/or DR1-associated allele may also be involved in the etiology of RA (Grennan et al. 1986). This putative DR3-associated allele may be the common genetic factor predisposing to both RA and ATD that is suggested by the present study.

ACKNOWLEDGMENT

This research was supported by grant HD 12731 from the National Institutes of Health and grant AM-20618, Northeast Ohio Multi-purpose Arthritis Center, from the National Institute of Arthritis, Diabetes and Digestive and Kidney Diseases.

APPENDIX

We have derived the HLA haplotype–sharing expectations for 12 cases, considering four different interrelationship models for each of the three disease statuses of sib pairs. The expectations for each model are derived conditional on the parental mating types: First we list all possible parental mating types; then, for each type, we derive the probability that two offspring would be affected in the disease status of interest (using penetrance schemes represented in table 1) and share 2, 1, or 0 haplotypes. The overall haplotype-sharing expectations for the population are then calculated by multiplying the above probabilities by the probability of their parental mating type and normalizing the sum of the probabilities for each haplotype-sharing configuration over the total. Consider model II as an example. As shown in table 1, there are six genotypes and 21 mating types. On the basis of these mating types, the expected distribution for AB-A sib pairs are derived as follows:

$$X = \lambda_A(1 - g_1)/\Sigma \ ;$$

$$Y = \{\lambda_A[1 - g_1(1 - p_d)] + p_{D_1}\}/\Sigma \ ;$$

$$Z = p_{D_1}\{2\lambda_A[(1 - g_1)p_{D_2} + p_d] + p_{D_1}\}/\Sigma \ ;$$

$$\Sigma = \lambda_A[2(1 + p_{D_1} - p_{D_1^2}) - g_1(2 + 2p_{D_1}p_{D_2} - p_d)] + p_{D_1}(1 - p_{D_1}) \ .$$

For AB-AB sibs the expectations are only a function of allele frequencies:

$$X = 1/\Sigma \ ;$$

$$Y = (p_{D_1} + p_{D_2})/\Sigma \ ;$$

$$Z = 2p_{D_1}p_{D_2}/\Sigma \ ;$$

$$\Sigma = 1 + p_{D_1} + p_{D_2} + 2p_{D_1}p_{D_2} \ .$$

For A-B sib pairs the expectations are determined by six parameters:

$$X = a/\Sigma \ ;$$

$$Y = b/\Sigma \ ;$$

$$Z = c/\Sigma \ ;$$

$$\Sigma = a + b + c \ ;$$

$$a = 2f_1(1 - f_1)g_1(1 - g_1) \ ;$$

$$b = 2\{(1 - f_1)g_1p_{D_1}[f_1(1 - g_1) + f_2] + f_1(1 - g_1)p_{D_2}[(1 - f_1)g_1 + g_2] + f_1g_1p_d(3 - f_1 - g_1)\} \ ;$$

$$c = \{2f_1[(1 - g_1)p_{D_2} + p_d] + f_2p_{D_1}][2g_1[(1 - f_1)p_{D_1} + p_d] + g_2p_{D_2}\} \ .$$

REFERENCES

Bertrams, J., M. P. Baur, F. Clerget-Darpoux, M. C. King, C. Kollmar, M. Neugebaur, and H. Payami. 1984. Insulin dependent diabetes mellitus. Pp. 348–358 *in* E. D. Albert, ed. Histocompatability testing. Springer, Berlin and Heidelberg.

Brewerton, D. A. 1978. Inherited susceptibility to rheumatic disease. J. R. Soc. Med. **71**:331–338.

Edwards, A. W. F. 1971. Distance between populations on the basis of gene frequencies. Biometrics **27**:873–881.

Farid, N., H. Moens, B. Larson, R. Payne, K. Saltman, F. Fifield, and D. Ingram. 1980. HLA haplotypes in familial Graves disease. Tissue Antigens **15**:492–500.

Grennan, D. M., P. Dyer, R. Clegue, W. Dodds, I. Smeaton, and R. Harris. 1983. Family studies in RA—the importance of HLA DR4 and genes for autoimmune thyroid disease. J. Rheumatol **10**:584–589.

Grennan, D. M., P. Sanders, P. Dyer, and R. Harris. 1986. HLA haplotype sharing by siblings with rheumatoid arthritis: evidence for genetic heterogeneity. Ann. Rheum. Dis. **45**:126–129.

Kahn, C. R., and J. S. Flier. 1980. Immunologic aspects of endocrine disease. Pp. 815–866 *in* C. W. Parker, ed. Clinical immunology. Vol. **2**. W. B. Saunders, Philadelphia.

Louis, E. J., and G. Thomson. The three allele synergistic mixed model for IDDM. Diabetes (in press).

Louis, E. J., G. Thomson, and H. Payami. 1983. The affected sib method. II. The intermediate model. Ann. Hum. Genet. **47**:225–243.

MacDonald, M. J., J. Gottschall, J. B. Hunter, K. L. Winter, S. D. Jhonson, J. L. Blank, E. P. Mason, and S. Maby. 1987. In insulin dependent diabetes the expression of an HLA-DR4 related susceptibility gene is dominant but the expression of a DR3 related gene is dependent on the DR4 related gene. Ann. NY Acad. Sci. (in press).

Parker, C. W. 1980. Clinical immunology. Vols. **1, 2**. W. B. Saunders, New York.

Payami, H., G. Thomson, M. A. Khan, D. M. Grannan, P. Sanders, P. Dyer, and C. Dostal. 1986. Genetics of rheumatoid arthritis. Tissue Antigens **27**:57–63.

Payami, H., G. Thomson, U. Motro, E. J. Louis, and E. Hudes. 1985. The affected sib method. IV. Sib trios. Ann. Hum. Genet. **49**:303–314.

Tait, B. D., A. Boyle, S. Solty, T. Cunningham, T. Mandel, F. I. R. Martin, and T. Doran. 1984. DR4 related antisera pattern differences in insulin dependent diabetes mellitus and rheumatoid arthritis. Tissue Antigens **24**:228–233.

Thomson, G., W. Klitz, E. J. Louis, S. K. Low, J. Bertrams, M. Baur, and M. Neugebaur. HLA and IDDM: new aspects. Genet. Epidemiol. (in press).

Torfs, C. P., M.-C. King, B. Huey, J. Malmgren, and F. C. Grumet. 1986. Genetic interrelationship between insulin-dependent diabetes mellitus, the autoimmune thyroid diseases, and rheumatoid arthritis. Am. J. Hum. Genet. **38**:170–187.