Originals

Will a three-allele model of inheritance explain the HLA data for Type 1 (insulin-dependent) diabetes?

D.A.Greenberg

Wadsworth VA Medical Center, Neurology Service, Los Angeles, California, USA

Summary. The HLA data from nine published studies on Type 1 (insulin-dependent) diabetes were examined to see whether a three-allele model for the inheritance of Type 1 diabetes at the HLA-associated locus could be rejected. None of the data rejected the three-allele model. The data were also examined to see whether they would reject a recessive model. Out of the nine data sets, five rejected a recessive and four did not. The p value for all studies together rejected a recessive. Two of the data sets allowed us to test the hypothesis that multiplex and simplex families would exhibit different modes of inheritance. Multiplex data from both data sets rejected recessive.

There has been much discussion recently concerning a three-allele model of inheritance for Type 1 diabetes. Hodge et al. [1] were among the first to propose a threeallele model. Their rather complex model proposed three forms of Type 1 diabetes, each with an associated penetrance. In constructing their model, Hodge et al. used relative risks, affected sibling-pair data and HLA-B locus data (DR data being rare at the time). Because of its complexity, this model has never been tested adequately. Greenberg and Anderson [2] analyzed data from five different studies and found that, although all studies rejected a dominant model, some studies rejected a recessive mode of inheritance while others did not. They also examined whether a three-allele model could fit the existing HLA data and, although not statistically testing the data, found that a three-allele model could fit the HLA data but not the affected sibling pair data. Risch [3], using a similar analysis, showed that the HLA-DR locus data of Barbosa et al. [4] would fit a three-allele model but not a dominant, recessive or intermediate model.

The purpose of this study was to test whether the three-allele model would fit the HLA-DR locus data. The three-allele model we use is restrictive, i.e., only those patients with both disease alleles at the HLA-associated locus can have the disease. To test the model, sive inheritance while the multiplex data from only one data set also rejected three-allele inheritance. The results of assuming a recessive model and analyzing the data from simplex families led to different results from the two data sets. In addition, data from a non-European population were examined and found to reject both recessive and three-allele inheritance for Type 1 diabetes at the HLA-associated locus.

Key words: Type 1 diabetes, HLA, heterogeneity, genetic model, three-allele.

we examine the distribution of HLA-DR genotypes among the affected population. This method eliminates the need to consider simple penetrance or other genetic or environmental factors since it deals only with the affected population [5, 6].

Methods

The three-allele model that we used was the simplest possible. We assumed that there are three different alleles at the HLA-associated disease locus, one normal allele and two different disease alleles. Only people who carry the two different disease alleles at the HLA-associated locus can be affected. We examine only the affected population, as has been described elsewhere [5, 6].

We assume that there are six different haplotypes, each corresponding to a disease allele – HLA allele combination. The population frequencies of these haplotypes we designate as follows: Ds1DR3 = u, Ds2DR3 = v, Ds1DR4 = w, Ds2DR4 = x, Ds1DRx = y, Ds2DRx = z - where Ds1 = disease allele 1 and <math>Ds2 = disease allele 2. DRx is any HLA marker other than DR3 or DR4. There are thus six parameters that describe the frequencies of the haplotypes in the general population.

These haplotypes were then combined to produce the following genotypes which, under the three-allele model, can occur in affected subjects:

Table 1. Comparison of observed and expect	a distributions of HLA genotypes among probands assuming a three-affele model
Study	Genotype

Study		Genotype										
		DR3/DR4	DR3/x	DR4/x	DRx/x	DR3/DR3	DR4/DR4	χ^2				
Wolf et al. [18]	Expected Observed	62.4 62	13.5 15	25.3 27	3.8 3	9.7 9	7.4 6	0.78				
Brautbar et al. [10]	Expected Observed	13.7 14	3.7 3	15.7 16	1.8 2	1.8 2	13.3 13	0.18				
Anderson et al. [8] DR4/?=DR4/DR4	Expected Observed	34.3 34	14.2 15	24.0 24	7.1 7	2.9 3	14.5 14	(0.06) ^a				
Anderson et al. [8] DR4/?=DR4/blank	Expected Observed	33.8 34	13.0 15	34.3 36	9.3 7	3.7 3	2.8 2	1.32				
Ginsberg et al. [16]	Expected Observed	17.5 18	11.9 13	15.2 15	4.4 4	7.7 7	7.3 7	0.22				
Christy et al. [11]	Expected Observed	13.5 13	14.6 15	24.2 24	12.7 13	4.2 4	3.8 4	0.06				
Suciu-Foca et al. [17]	Expected Observed	13.0 14	14.5 10	11.6 11	6.9 9	7.3 9	3.6 4	2.50				
Farid et al. [15]	Expected Observed	9.4 10	8.5 8	11.2 11	7.7 8	1.1 1	2.1 2	0.09				
Barbosa et al. [4]	Expected Observed	33.8 35	17.8 19	55.6 55	7.5 7	7.7 7	11.6 11	0.26				
De Jongh et al. [7]	Expected Observed	30.7 32	10.1 8	18.2 18	3.8 5	5.9 6	6.3 6	0.86				

Each figure in the table represents the number of propositi with the corresponding genotype.

Overall $\chi^2 = 6.27$ with 9 degrees of freedom (not significant).

^a For the Anderson et al. data, two runs were carried out – one where all patients who typed only for DR4 but whose genotype could not be determined were treated as DR4/DR4, and one where they were treated as DR4/blank. Only the higher χ^2 value was used to compute the overall χ^2

The population frequency of Type 1 diabetes is then:

f = 2(uv + uz + uz + wv + wz + wz + yv + yz + yz)

The unknown population frequency of the two disease alleles in the general population, Ds1 and Ds2, we designate p_1 and p_2 . Therefore

 $f = 2 p_1 p_2$.

The above statements refer to population frequencies, but we now make the restriction that we are looking *only* at the affected population (including only unrelated people). By the requirements of the model, every affected person has both disease alleles at the disease locus. Therefore, in the affected population, $p_1 = p_2 = 0.5$. Also, we redefine the haplotype frequencies u, v, ..., z to be the frequencies of the haplotypes in the *affected population*. We can then set up the following equality

$$1 = \frac{f}{2 p_1 p_2} = \frac{2(uv + ux + uz + wv + wx + wz + yz + yv + yx)}{2(0.5) (0.5)}$$

Moreover, by the requirement that the frequency of the Ds1 allele be 0.5, and the same for the Ds2 allele, we can reduce the number of parameters by the equalities

$$u = 0.5 - w - y$$

 $v = 0.5 - x - z$,

since $u + w + y = p_1 = x + v + z = p_2 = 0.5$. There are thus only four parameters, given the restrictions of the model and the fact that we examine only the affected population.

The mutually exclusive and exhaustive categories of HLA genotypes DR3/DR4, DR3/x, DR4/x, DR3/DR3, DR4/DR4, and DRx/x constitute five independent observations (since one category is fixed once the others are known). We therefore have four independent parameters and five observations, leaving us one degree of freedom for testing. The best fit of the parameters to the data was found by going through all possible combinations of parameters (in steps of 0.01) to compute the theoretical proportions of HLA genotypes in a data set. (Again, the parameters are the theoretical proportions of the abovelisted haplotypes among affected individuals.) The theoretical proportions of HLA genotypes among affected individuals were computed and compared with the observed proportions, using a χ^2 statistic. The parameters which are reported were those that led to the best agreement between the theoretical proportions and the observed proportions. The theoretical values for the recessive model were computed similarly [2, 5].

Data were taken from nine different studies. Anderson et al. [8] (updated in [9]) reported that in 12 patients who typed positive for DR4 only, it was impossible to determine whether they were DR4/DR4 or DR4/blank. Thus, our analysis was done under both extreme cases: assuming all 12 patients were DR4/DR4 and assuming they were all DR4/blank. Brautbar et al. [10] separated their data into populations of Ashkenazi and non-Ashkenazi patients. We included only the Ashkenazi data in the main analysis, since the other studies report data mostly from populations of European origin. The "prospective" data of Christy et al. [11] were used for the main analyses. By "prospective" is meant "randomly ascertained in a clinic," which it is reasonable to assume come mostly from simplex families.

Results

The expected and observed numbers of each of the six HLA genotypes under the three-allele model are shown in Table 1. The fit is good for all of the data sets. When all the χ^2 values are added together and the degrees of freedom adjusted accordingly (1 degree of freedom added for each data set), the total χ^2 value for all observations is not significant.

Study		Genotype	Genotype									
		DR3/DR4	DR3/x	DR4/x	DRx/x	DR3/DR3	DR4/DR4	χ^2				
Wolf et al. [18]	Expected Observed	40.8 62	16.7 15	19.3 27	4 3	17.6 9	23.6 6	31.83				
Brautbar et al. [10]	Expected Observed	11.8 14	4.8 3	12.9 16	2.6 2	2.2 2	15.7 13	2.51				
Anderson et al. [8] DR4/?=DR4/DR4	Expected Observed	25.6 34	15.1 15	22.2 24	6.6 7	8.7 3	18.8 14	(7.91) ^a				
Anderson et al. [8] DR4/?=DR4/blank	Expected Observed	22.0 34	16.2 15	25.3 36	9.3 7	7.1 3	17.1 2	27.47				
Ginsberg et al. [16]	Expected Observed	16.6 18	12.5 13	13.3 15	5.0 4	7.8 7	8.8 7	1.02				
Christy et al. [11]	Expected Observed	11.2 13	15.4 15	20.6 24	14.1 13	4.2 4	7.5 4	2.60				
Suciu-Foca et al. [17]	Expected Observed	11.8 14	14.8 10	11.2 11	7.0 9	7.8 9	4.5 4	2.76				
Farid et al. [15]	Expected Observed	6.7 10	8.7 8	10.8 11	7.1 8	2.7 1	4.1 2	4.02				
Barbosa et al. [4]	Expected Observed	27.7 35	21.2 19	38.0 55	1 4. 6 7	7.7 7	24.8 11	21.43				
De Jongh et al. [7]	Expected Observed	21.9 32	11.7 8	14.8 18	4.0 5	8.7 6	13.9 6	12.04				

Table 2. Comparison of observed and expected distributions of HLA genotypes among probands assuming a recessive model

Overall $\chi^2 = 105.68$ with 9 degrees of freedom, $p < 10^{-5}$. Explanation and key as in footnote to Table 1

Table 3. Analysis by family type from the data of De Jongh et al. [7]

Family type		Genotype									
		DR3/DR4	DR3/x	DR4/x	DRx/x	DR3/DR3	DR4/DR4	χ^2	р		
Three-allele model											
All propositi	Expected Observed	30.7 32	10.1 8	18.2 18	3.8 5	5.9 6	6.3 6	0.86	NS		
Propositi from simplex families	Expected Observed	24.1 23	6.7 8	12.6 13	2.5 2	3.2 3	3.9 4	0.45	NS		
Propositi from multiplex families	Expected Observed	7.4 9	3.9 0	5.2 5	1.6 3	2.1 3	1.8 2	5.76	< 0.025		
Recessive model											
All data	Expected Observed	21.9 32	11.7 8	14.8 18	4 5	8.7 6	13.9 6	12.04	< 0.0001		
Simplex only	Expected Observed	16 23	8.2 8	10 13	2.6 2	6.5 3	9.8 4	9.43	< 0.005		
Multiplex only	Expected Observed	5.7 9	4.3 0	4.6 5	1.7 3	2.7 3	3.0 2	7.59	< 0.01		

Explanation as for Table 1

For comparison, the expected and observed numbers of each of the HLA genotypes among probands, assuming a simple recessive model, are shown in Table 2. Here, the χ^2_{τ} values for five of the data sets are significant, as is the overall χ^2 , indicating a rejection of the simple recessive model. Note, however, that the results for the recessive model appear to be dichotomous; that is, while five data sets clearly reject the recessive (those with a $\chi^2 > 3.84$) the other four do not. This dichotomy has been noted previously [2]. Anderson et al. [8] could not determine in all cases whether patients who typed for DR4 only were DR4/DR4 or DR4/blank. We did the calculations assuming the extremes of counting all patients who typed only for DR4 successively as DR4/DR4 and then as DR4/blank. Table 1 shows that, for the three-allele model, assuming all DR4/? are DR4/DR4 or assuming they are all DR4/blank leads to a non-significant χ^2 , although the χ^2 is higher when all are called DR4/blank. For the recessive model, in contrast, both assumptions

D. A. Greenberg: A three-allele model for Type 1 diabetes

Family type		Genotype									
		DR3/DR4	DR3/x	DR4/x	DRx/x	DR3/DR3	DR4/DR4	χ^2	р		
Three-allele model All propositi	Expected Observed	23.1 23	18.4 18	30.6 31	15.7 16	4 .7 5	5.5 5	0.09	NS		
"Prospective" group	Expected Observed	13.5 13	14.6 15	24.2 24	12.7 13	4.2 4	3.8 4	0.06	NS		
"Familial" group	Expected Observed	9 10	3.9 3	7.3 7	2.4 3	1 1	1.4 1	0.58	NS		
Recessive model All propositi	Expected Observed	17.3 23	20.4 18	26.7 31	15.7 16	6.6 5	11.3 5	6.79	< 0.01		
"Prospective" group	Expected Observed	11.2 13	15. 4 15	20.6 24	14.1 13	4.2 4	7.5 4	2.60	NS		
"Familial" group	Expected Observed	6.2 10	4.5 3	5.8 7	2.1 3	2.4 1	4.0 1	6.53	< 0.025		

Table 5. Data from non-Ashkenazi population of Brautbar et al. [10]

Model		Genotype									
		DR3/DR4	DR3/x	DR4/x	DRx/x	DR3/DR3	DR4/DR4	χ^2	p		
Three-allele	Expected Observed	20.4 23	4. 7 1	8.5 7	1.4 3	2.6 3	4.5 5	5.34	< 0.025		
Recessive	Expected Observed	14.0 23	5.6 1	6.8 7	1.4 3	5.7 3	8.5 5	14.32	< 0.0001		

Table 6. Calculated haplotype frequencies among probands

Study	Ds1DR3	Ds2DR3	Ds1DR4	Ds2DR4	Ds1DRx	Ds2DRx
Wolf et al. [18]	0.33	0.06	0.04	0.38	0.13	0.06
Brautbar et al. [10]	0.15	0.06	0.17	0.39	0.18	0.05
Anderson et al. $[8]$ (?=4)	0.25	0.03	0.11	0.34	0.14	0.13
Anderson et al. [8] $(?=blank)$	0.24	0.04	0.02	0.36	0.24	0.10
Ginsberg et al. [16]	0.20	0.15	0.11	0.26	0.19	0.09
Christy et al. [11]	0.16	0.09	0.05	0.26	0.29	0.15
Suciu-Foca et al. [17]	0.23	0.14	0.08	0.20	0.19	0.16
Farid et al. [15]	0.22	0.03	0.05	0.26	0.23	0.21
Barbosa et al. [4]	0.16	0.09	0.06	0.36	0.28	0.05
De Jongh et al. [7]	0.28	0.07	0.06	0.35	0.16	0.08

lead to a significant χ^2 . In other words, the statistical conclusions are the same, independent of how the DR4/? patients are interpreted.

Discussion

It appears from our results that a three-allele model can adequately account for the observed distribution of HLA types among probands in different populations. One aspect of the model we use is that it is simple, i. e., only people with each of the two disease alleles can be affected. The fact that the model is restrictive, yet successfully explains the distribution of HLA types in different populations, is encouraging. Risch [3] published a model similar to the one presented previously [2, 5] and to this one. He also found that the data of Barbosa et al. [4] fitted the three-allele model. One difference between Risch's model and the present one is that the latter is more restrictive, having only four parameters and one degree of freedom instead of seven parameters and two degrees of freedom. The model appears to provide a good description of the data in spite of the additional restrictions.

We had previously proposed that there may be two forms of Type 1 diabetes, one expressed more (that is, less penetrant) in simplex families (with only one affected member) and one that showed up more in multiplex families (more than one affected member) [2]. We suggested that the multiplex form was three-allele and the

simplex form was recessively inherited. In addition, we suggested that there was a form that was not HLA associated. The data of de Jongh et al. [7] gave us an opportunity to test further part of the hypothesis. As can be seen from Table 3, both the simplex and multiplex data reject a recessive mode of inheritance, which is contrary to our hypothesis. But this result is also different from the result using the data of Christy et al. [11], where we found that the "prospective" data did not reject a recessive [2]. However, the multiplex data of de Jongh et al. also reject a three allele mode of inheritance while none of the data-type categories from the data of Christy et al. reject a three-allele model (Table 4). In the de Jongh et al. multiplex data, the major contribution to the χ^2 comes from the DR3/x category. There were no observed DR3/x genotypes among the multiplex propositi. Interestingly, the data of de Jongh et al. indicate a haplotype sharing among affected siblings (i.e., multiplex families) of 0.83:0.17:0 (12 pairs of siblings observed) which would be in line with a three-allele or a recessive model, but which is a higher proportion of siblings with two alleles in common than had previously been reported [11]. It therefore appears that, while the de Jongh et al. and the Christy et al. data sets lead to different conclusions vis-a-vis the two models we are considering, both data sets point to differences between data from simplex and from multiplex families.

The data of Brautbar et al. [10] are broken down into Ashkenazi and non-Ashkenazi groups. The analyses reported in Tables 1 and 2 reflect only the Ashkenazi group. If we look at the non-Ashkenazi group (Table 5), we see that the data reject both the recessive and threeallele models. In fact, the results look similar to those found in multiplex families by de Jongh et al. [7] discussed above, with the major contribution to the χ^2 coming from the DR3/x category, and with fewer DR3/x propositi than predicted by the three-allele model. While the data are sparse, this could be construed as further evidence for another form of Type 1 diabetes.

Some cautions must be raised in any work such as this: First, the calculated frequencies of the haplotypes carrying the disease alleles vary widely among the data sets (Table 6). This most likely reflects the wide variation from study to study of the proportions of the different HLA genotypes among probands and control subjects. The explanation offered by Risch [12], that such variation represents differences in ascertainment schemes, is unsatisfactory. The stated ascertainment schemes were similar. Had patients been chosen on the basis of their HLA type, then there would be a bias in the data. None of the studies used such an ascertainment scheme. [By ascertainment, we do not mean the term as it is used in segregation analysis (i.e., correcting data for unobserved at-risk matings). Such ascertainment plays no role in the methods being used here. We mean ascertainment in the more general sense of how the patients were brought into the studies.]

Assuming that the mode of inheritance of Type 1 diabetes at the HLA-associated locus is three-allele, it may not be surprising that some data sets reject the recessive and some do not. Three-allele inheritance can be viewed as a more general case of the recessive. It is possible that in some data sets the distribution of HLA types among probands would mimic recessive inheritance. Falk [13] has pointed out that different genotypes may exhibit a variety of penetrances depending on the environment. Thus, if there were, in fact, two disease susceptibility alleles, the penetrance of a homozygote for one of the alleles may be greater in one environment than in another. Such an hypothesis could explain our observations when we attempted to fit the data to a recessive model; there, some data sets strongly rejected the recessive and some did not. Falk points out that the three-allele model would be the same as a recessive, for example, when the penetrances of all genotypes at the disease locus were the same, and more "three-allelelike" when the heterozygote was more penetrant than the other genotypes. (In our notation, the disease genotypes would be Ds1-Ds1, Ds2-Ds2, Ds1-Ds2.) Our analysis indicated that such a strong dichotomy existed between those that accepted and those that rejected, a recessive that, taking Falk's approach, we might expect there to be only two different environments.

Secondly, the affected sibling-pair data appear incompatible with a three-allele model [2]. The methods used to calculate the expected proportions of affected siblings sharing both, one, or neither HLA haplotype make assumptions such as no diminished survival or reproductive capacity due to disease, and no significant recombination between HLA and the disease allele. However, Payami et al. [14] have shown that these factors may not cause large differences in the results in the case of the recessive and, by analogy, this applies to the three-allele model.

Thirdly, it is important to remember that a good statistical fit does not prove that a model is correct. We have shown that the fit of the three-allele model to the association data is good and that, based on those data, this model cannot be rejected.

Finally, the calculations done here assume a uniform disease population with no heterogeneity. We have shown that a three-allele model will explain the distribution of HLA types in most studies, but this does not explain the affected-sib-pair rate. In addition, the de Jongh et al. [7] multiplex data, the Brautbar et al. [10] non-Ashkenazi data and the Christy et al. [11] data hint at another form of Type 1 diabetes.

Acknowledgements. This work was supported in part by the Veterans Administration, by grant AM31775 from NIH and by NIH contract number N01-NS-2332.

References

 Hodge SE, Rotter JI, Lange KL (1980) A three-allele model for heterogeneity in juvenile-onset insulin-dependent diabetes. Ann Hum Genet 43: 399–412

- 2. Greenberg DA, Anderson E (1983) The search for heterogeneity in insulin dependent diabetes mellitus: evidence for familial and non-familial forms. Am J Med Genet 14: 487–499
- 3. Risch N (1983) A general model for disease-marker association, Ann Hum Genet 47: 245-252
- 4. Barbosa J, Bach FH, Rich SS (1982) Genetic heterogeneity of diabetes and HLA. Clin Genet 21: 25–32
- 5. Greenberg DA, Hodge SE, Rotter JI (1982) Evidence for recessive and against dominant inheritance at the HLA "linked" locus in coeliac disease. Am J Hum Genet 34: 263–277
- 6. Thomson G (1983) Investigation of the mode of inheritance of the HLA associated diseases by the method of antigen genotype frequencies among diseased individuals. Tissue Antigens 21: 81–104
- de Jongh BM, Bruining GJ, Schreuder GMTh, Schuurman RKB, Radder JK, van Loghem E, Kahn PM, Hauptmann G, van Rood JJ (1984) HLA ands GM in insulin dependent diabetes in the Netherlands: report on a combined multiplex and family population study, Hum Immunol 10: 5–21
- Anderson CE, Hodge SE, Rubin R, Rotter JI, Terasaki PI, Irvine WJ, Rimoin DL (1983) A search for heterogeneity in insulin dependent diabetes mellitus: HLA and autoimmune studies in simplex, multiplex and multigenerational families, Metabolism 32: 471-477
- Hodge SE, Anderson CE, Neiswanger K, Rubin R, Sparkes RS, Sparkes MC, Crist M, Spence MA, Terasaki PI, Rimoin DL, Rotter JI (1983) Association studies between Type 1 (insulin-dependent) diabetes and 27 genetic markers: lack of association between Type 1 diabetes and Kidd blood group. Diabetologia 25: 343–347
- Brautbar C, Karp M, Amar A, Cohen I, Cohen O, Sharon R, Topper E, Levene C, Cohen T (1981) Genetics of insulin dependent diabetes mellitus in Israel: population and family study. Hum Immunol 3: 1–12
- Christy M, Green A, Christau B, Kormann H, Nerup J, Platz P, Thomson M, Ryder LP, Svejgaard A (1979) Studies of the HLA

system and insulin dependent diabetes mellitus. Diabetes Care 2: 209-214

- Risch N (1984) Segregation analysis incorporating linkage markers. I. Single-locus models with an application to Type 1 diabetes, Am J Hum Genet 36: 363–386
- Falk CT (1984) A two-susceptibility-allele model for genetic disease and associated marker loci: differences and similarities to a one-s-allele model. Ann Hum Genet 48: 87–95
- Payami H, Thomson G, Louis EJ (1984) The affected sibling method. III. Selection and recombination, Am J Hum Genet 36: 352-362
- Farid NR, Sampson L, Noel P, Barnard JM, Davis AJ, Hillman DA (1979) HLA-D-related (DRw) antigens in juvenile diabetes mellitus. Diabetes 28: 552–557
- 16. Ginsberg-Fellner F, Dobersen MJ, Witt ME, Rayfield EJ, Rubenstein P, Notkins AL (1982) HLA antigens, cytoplasmic islet cell antibodies, and carbohydrate tolerance in families of children with insulin dependent diabetes mellitus, Diabetes, 31: 292–298
- Suciu-Foca N, Rubenstein P, Nicholson J, Susinno E, Weiner J, Godfrey M, Hardy M, Rayfield E, Reemtsma K (1979) Juvenile diabetes mellitus and the HLA system. Transplant Proc xi: 1309-1313
- Wolf E, Spencer KM, Cudworth AG (1983) The genetic susceptibility to Type 1 (insulin-dependent) diabetes: analysis of the HLA-DR association. Diabetologia, 24: 224–230

Received: 31 August 1984 and in revised form: 13 December 1984

Dr. David A. Greenberg Wadsworth VA Medical Center Room 3405 Los Angeles, CA 90073 USA