HLA-A amino acid polymorphism and delayed kidney allograft function

Malek Kamoun[a,1,2], John H. Holmes[b,1], Ajay K. Israni[c], Jane D. Kearns[a], Valerie Teal[b], Wei Peter Yang[b], Sylvia E. Rosas[d], Marshall M. Joffe[b], Hongzhe Li[b], and Harold I. Feldman[b,d]

Departments of [a]Pathology and Laboratory Medicine and [d]Medicine, and [b]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, Philadelphia, PA, 19104; and [c]Department of Medicine, Hennepin County Medical Center, Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN 55415

Communicated by Jean Dausset, Fondation Jean Dausset-CEPH, Paris, France, October 15, 2008 (received for review August 20, 2008)

Delayed allograft function (DGF) is a common adverse event in postrenal transplantation. The etiology of DGF is thought to include both nonimmunologic (donor age, cold ischemia time, and recipient race) and immunologic factors. We examined the association of DGF with amino acid mismatches at 66 variable sites of the HLA-A molecule in a prospective cohort study of 697 renal transplant recipients of deceased donors. Using a multivariate logistic regression model adjusted for nonimmunologic risk factors, we show that combinations of a few amino acid mismatches at crucial sites of HLA-A molecules were associated with DGF. In Caucasian recipients, a mismatch at position 62, 95, or 163, all known to be functionally important within the antigen recognition site, was associated with an increased risk for DGF. Furthermore, a decreased risk for DGF was associated with a mismatch at HLA-A family-specific sites (149, 184, 193, or 246), indicating that evolutionary features of HLA-A polymorphism separating HLA-A families and lineages among donor-recipient pairs may correlate with the magnitude of alloreactivity influencing the development of DGF. These findings suggest that amino acid polymorphisms at functionally important positions at the antigen recognition site of the HLA-A molecule have a significant influence on DGF.

HLA alleles | kidney transplantation | HLA mismatches | MHC

Major histocompatibility complex (MHC) genes were originally discovered (1) because of their role in tissue rejection in mammals (1, 2). In clinical transplantation, HLA alloantigens represent a formidable barrier to successful transplantation; HLA mismatches are associated with acute allograft rejection and allograft loss (3, 4). Historically, acute rejection was the principal clinical challenge after renal transplantation and a primary determinant of transplantation outcome. With the introduction of potent immunosuppressive treatment, the frequency of acute rejection episodes after kidney transplantation from deceased donors has been reported at less than 20% (4–6). Increasing attention is being focused on delayed allograft function (DGF), a more common continuing obstacle to successful renal transplantation. DGF is a major adverse event associated with an increased incidence of rejection and a reduced graft survival (7–9). Although the clinical severity of DGF varies, it is conventionally defined as the need for dialysis in the first 7 days after transplantation. The rate of DGF after kidney transplantation from deceased donors ranges between 15 and 50% (10).

The etiology of DGF is not well understood but is thought to include both immunologic and nonimmunologic components (10). The exact contribution of DGF to allograft loss is debatable, and although DGF has been found to be a predisposing factor for acute rejection and decreased allograft survival, the exact relationship between DGF and acute rejection remains controversial. There is evidence that immunologic events can up-regulate immune response and may increase alloreactivity, resulting in an increased risk for acute rejection (10). The main clinical risk factors for DGF are increasing donor age, recipient race, longer cold ischemia time (CIT), and the presence of anti-HLA antibodies (10).

In renal transplantation, organ alloreactivity may play a significant role in the development of DGF. However, the influence of the degree of HLA-A, -B and -DRB1 matching on DGF remains unclear. Because in renal transplantation HLA matching is only performed at the antigen level, the effect of HLA mismatches at the molecular amino acid level on organ alloreactivity and allograft function has remained largely unexplored. There are two major classes of HLA molecules, termed class I and class II. Class I molecules are heterodimers, composed of a single-membrane-spanning α chain paired with a soluble $\beta 2$ microglobulin protein. The α chain includes three distinct segments: $\alpha 1$, $\alpha 2$, and $\alpha 3$. The $\alpha 3$ region has an Ig-like fold, whereas the membrane distal $\alpha 1$ and $\alpha 2$ segments form a peptide-binding cleft consisting of two α helices overlying a floor comprised of eight antiparallel β-stranded sheets (11–13). HLA class II are also heterodimers comprised of two subunits, both contributing to the formation of the antigen-binding site. Genes encoding classical HLA molecules are extremely polymorphic, representing the most polymorphic set of genes in the human genome. Most classic HLA genes include a very large number of allelic variants, and allele distribution varies widely according to race and ethnicity. Most polymorphisms are associated with the peptide-binding residues of the HLA class I (HLA-A, -B, and -Cw) and HLA class II molecules (DR, DQ, and DP) (11). In humans, allelic variations among the classic HLA class I and class II gene products are the source of differential peptide binding, thymic selection, and alloreactivity (11).

For this article, we examined the influence of the degree of HLA-A, -B and -DRB1 matching on DGF in a cohort of recipients of deceased donor renal transplants. Based on our results, we further evaluated in a multivariate logistic regression, adjusted for significant nonimmunological covariates, the association of amino acid mismatches at individual sites of an HLA-A molecule with DGF. We discuss our findings in the context of structural and functional correlates of amino acid polymorphism of HLA-A molecules, and the potential application of this approach to the complexity of HLA polymorphisms as they relate to clinical transplantation.

IMMUNOLOGY

This study was presented in part at the meetings of the American Transplant Congress, May 2008 (Toronto, Canada) and May 2007 (San Francisco, CA) and at the European Federation of Immunogenetics, April 2008 (Toulouse, France).

Author contributions: M.K., J.H.H., A.K.I., and H.I.F. designed research; J.H.H., J.D.K., V.T., and W.P.Y. performed research; J.H.H. and H.L. contributed new reagents/analytic tools; M.K., J.H.H., A.K.I., S.E.R., M.M.J., H.L., and H.I.F. analyzed data; and M.K., J.H.H., A.K.I., and H.L. wrote the paper.

The authors declare no conflict of interest.

[1]M.K. and J.H.H. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: malekkam@mail.med.upenn.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0810308105/DCSupplemental.

Table 1. Association of HLA-A allele mismatches with delayed graft function: Results of a multivariate logistic regression analysis*

	Unadjusted		Adjusted for other risk factors	
HLA mismatches	OR (95% CL)	*P* Value	OR (95% CL)	*P* Value
Antigen-level[1]				
HLA-A	1.25 (0.94–1.59)	0.07	1.12 (0.82–1.54)	0.44
HLA-B	1.08 (0.83–1.42)	0.54	1.40 (0.98–1.99)	0.06
HLA-DR	1.26 (0.97–1.65)	0.08	1.00 (0.72–1.39)	0.98
Additional mismatches detected at allele-level[2]				
HLA-A	3.66 (1.72–7.76)	0.0007	3.09 (1.20–7.94)	0.02
HLA-B	1.12 (0.69–1.81)	0.62	1.40 (0.79–2.47)	0.24
HLA-DR	1.04 (0.70–1.59)	0.82	0.87 (0.53–1.41)	0.57
Recipient race (AA vs. non-AA)			2.16 (1.37–3.39)	0.001
Pre-transplant dialysis	—	—	4.56 (1.63–12.72)	0.004
Previous transplant	—	—	1.91 (0.99–3.66)	0.05
Donor age (year)	—	—	1.02 (1.00–1.03)	0.002
Cold ischemia time (hours)	—	—	5.10 (1.23–21.12)	0.02

*Subjects include both non-AA and AA. Other non significant covariates that were included in this multivariate logistic regression model are: recipient age, cause of the end-stage renal disease, diabetes, recipient sex, previous transplant, pre-transplant PRA, and donor race and sex.

[1]A numerical form of HLA mismatches (0, 1) was used for this analysis as described in *Materials and Methods*.

[2] Additional mismatches detected at the allele-level (non imputed alleles) in recipients that were matched at the antigen level (1 or 2 antigens match) as described in *Materials and Methods*.

Results

Recipient and Donor Characteristics. A total of 705 donor-recipient pairs were enrolled into a longitudinal cohort study of transplantation, as previously described (14, 15). The racial distribution was as follows: 498 (70.6%) were non-African Americans (AA) and 207 were AA (29.4%). Of the non-AA, 1.99% were Asians, 1.56% were Hispanics, 0.52% were Asian Indians, and 0.17% were of other races. Eight were excluded from this analysis because of missing data for delayed allograft function, yielding 697. The characteristics of the study population are shown in supporting information (SI) Table S1.

Association of HLA Allele Mismatches with DGF. In this cohort, the frequency of HLA-A, -B, and -DRB1 antigen mismatches in donor-recipient pairs and the degree and distribution of allele mismatches among recipients with no mismatch or one antigen mismatch were previously reported (14). In recipients with a zero antigen (serologic) mismatch at HLA-A, -B, or -DR, the frequency of one or two HLA allele (high resolution) mismatches were 10.7%, 28.7%, and 47.2%, respectively, in the entire cohort. Among recipients with one antigen mismatch at HLA-A, -B, or -DR, additional allele mismatches were observed in 9.6%, 21.9%, and 33% respectively (14). None of three ordinal terms for HLA antigen mismatches at the A, B, and DR loci was associated with DGF in an unadjusted logistic regression model (Table 1). However, HLA-A mismatches detected at the allele level, adjusted for matching at the antigen level, were significantly associated with DGF [Odds ratio (OR), 3.66; confidence interval (CI), 1.7–7.7; $P = 0.0007$]. This association was attenuated, but remained significant (OR, 3.09; CI, 1.20–7.94; $P = 0.02$) in a multivariable logistic regression model adjusted for recipient race, dialysis pretransplant, previous transplant, donor age, and CIT (see Table 1). In contrast, HLA-B or -DRB1 mismatches at the allele level adjusted for matching at the antigen level were not significantly associated with DGF (see Table 1). Although the number of additional HLA-A mismatches detected at the allele level was relatively small (37 out of 348), their effect, as measured by the OR (and accompanying *P*-values) was significant, suggesting a potentially important impact of these allele mismatches. These findings remained similar after modeling HLA antigen mismatches categorically, as 0, 1, or 2 (data not shown). Interestingly, we have previously reported in this cohort that HLA-A amino acid residues at peptide contact sites were frequently mismatched among donor-recipient pairs who have one or two mismatches at the allele level, but are matched at the antigen level (14). These findings suggested that amino acid mismatches at functionally important sites may be associated with DGF and prompted us to evaluate these potential associations further.

Identification of Candidate Amino Acid Mismatch Sites that Are Associated with DGF. Amino acid sequence profiles covering the antigen-presenting domain (α1, α2) as well as the α3 domains were created from 100 imputed HLA-A allele datasets, owing to the absence of complete data on the allele level for the HLA-A locus. In this study, the imputations and the statistical analyzes were restricted to Caucasian recipients. Over the 66 individual amino acid sites of the HLA-A molecule, mismatches were most frequently observed at the peptide binding sites (9, 77, 95, 97, 114, 116, 152, and 156) and T-cell receptor (TCR) contact sites (62 and 76). In addition, other frequent mismatches (found in 25–35% of the datasets) were observed at several positions in the α1 and α2 domains that are not peptide-binding or TCR contact sites, as well as in the α3 domain. Even though HLA-A alleles were imputed for many subjects, on the whole, the variance in the proportion of mismatched sites was relatively small over the 100 imputed datasets (Fig. S1).

We used an unadjusted logistic regression (Table 2) and several machine-learning feature selection methods (Fig. S2) to evaluate the association of single mismatched sites with DGF. Dichotomously coded (0, 1) mismatches at individual sites were evaluated independently of each other for each of the 100 imputed datasets. This analysis was done separately for all recipients and for those who were panel-reactive antibody (PRA)-negative before the transplant. Notably, five of the seven positions that were significantly associated with DGF belong to functionally important sites of the α1 and α2 domains. The selected sites included positions 9, 62, 95, 156, and 163 (see Table 2). Variation in the *P*-values observed for a few positions, such as 95 and 156, could be because of the effect of the partial imputation of HL-A alleles. These results were further confirmed using other univariate methods that included data-

Table 2. Association of HLA-A amino acid mismatches with delayed graft function for Caucasian recipients: Results of univariate analyses

			Logistic regression				
Domain	Position	Potential Contact	OR Mean (SD)	OR Range	Mean *P*-value (SD)	*P* Range	Mean % of MM
All recipients							
α1	9	P	1.80 (0.061)	1.65–1.97	0.0062 (0.0029)	0.0015–0.0185	0.55
	62	P-TCR	1.67 (0.014)	1.62–1.72	0.0193 (0.0021)	0.0128–0.0273	0.61
	90	Other	1.90 (0.049)	1.79–2.03	0.0058 (0.0019)	0.0021–0.0117	0.26
α2	105	Other	1.58 (0.091)	1.35–1.84	0.0496 (0.0291)	0.0061–0.1929	0.28
	156	P	1.46 (0.065)	1.31–1.63	0.0749 (0.0325)	0.0197–0.1929	0.53
	163	P-TCR	1.81 (0.038)	1.72–1.90	0.0115 (0.0028)	0.0061–0.0200	0.25
PRA-negative recipients							
α1	9	P	1.97 (0.080)	1.80–2.20	0.0061 (0.0028)	0.0013–0.0150	0.54
	62	P-TCR	1.58 (0.011)	1.53–1.58	0.0660 (0.0045)	0.0644–0.0901	0.61
	90	Other	1.82 (0.052)	1.72–1.95	0.0251 (0.0066)	0.0114–0.0388	0.25
α2	95	P	1.52 (0.080)	1.32–1.85	0.0864 (0.0414)	0.0099–0.2390	0.50
	163	P-TCR	1.77 (0.042)	1.70–1.87	0.0330 (0.0068)	0.0182–0.0462	0.24

A dichotomous amino acid mismatches coding was used. MM, mismatch; P, peptide; TCR, T- cell receptor.

mining tools, such as feature selection methods (see Fig. S2). A similar assortment of mismatched sites was observed using a trichotomous coding (0, 1, or 2) of amino acid mismatching (data not shown).

We also used a rule discovery method (JRip) to identify sets of positions that together were associated with increased risk for DGF. Each rule included specific combinations of mismatched and matched sites that, together, classified DGF. Amino acid sites identified by the rule discovery analysis included those shown in Table 2, as well as additional candidate sites (Table 3). Interestingly, listings of the HLA-A alleles in donor-recipient pairs tracked by several rule sets revealed a significant overlap, indicating the existence of a strong association between mismatches at different sites (data not shown). Indeed, a pair-wise correlation analysis between mismatches at various sites that were associated with DGF revealed multiple significant associations (Table S2). These pair-wise associations may be a reflection of the nature of HLA-A amino acid polymorphism and of the evolutionary features of the HLA-A molecule. Because of the small size of our sample, we were unable to evaluate further the interaction terms for relationships between mismatches at different sites in a multivariable logistic regression model.

Joint Effects of HLA-A Amino Acid Mismatches at Individual Sites on DGF. To assess the joint effect of the amino acid mismatches on DGF, we first performed stepwise logistic regression analysis that included the 22 aa sites that were selected by means of the machine-learning and statistical methods discussed above. We included the following risk factors as covariates in the logistic regression models: cold ischemia time, recipient age, donor age, and donor cause of death. All of these clinical variables were associated with DGF based on unadjusted analysis at $P \leq 0.05$ levels. To minimize the effect of preemptive transplantation on the OR, only recipients who had pretransplant dialysis (407 subjects, 87.2%) were included in this analysis. Logistic models were created separately for PRA-negative recipients and all recipients; models could not be created for the PRA-positive recipients-only because of the small number in this group.

In the stepwise logistic regression model, using liberal inclusion/exclusion criteria (*P* value to enter <0.25), candidate amino acid sites were selected based on a frequency of 20% or higher over the 100 imputations, with $P < 0.05$. Six of the 22 candidate amino acid sites were selected in the group of recipients that included both PRA-positive and PRA-negative subjects. These sites included position 9, 62, 90, 127, 149, 184, and 246; they were selected in 73%, 29%, 100%, 29%, 35%, 23%, and 65% of the imputed datasets, respectively. In the PRA-negative recipient group, six out of 22 mismatch sites were selected and included position 9, 90, 95, 127, 149, and, 193. These sites were selected in the regression models in 87%, 100%, 93%, 95%, 100%, and, 99% of the 100 imputed datasets, respectively. The other mismatched sites either remained in the logistic regression model in a small percentage of the datasets (3–12%) or did not remain in the model (data not shown).

The mismatch sites that were selected in the stepwise logistic model were then simultaneously evaluated in a fixed logistic regression model to assess their joint effects on DGF (Table 4).

Table 3. Additional HLA-A amino acid polymorphic sites associated with delayed graft function: Results of a classification rule algorithm[1]

Position*	66	70	74	76	77	97	114	127	144	149	152	166	184	193	246
Contact Site	α1 P/T	α1 P	α1 P	α1 T	α1 P	α2 P	α2 P	α2 Other	α2 Other	α2 T	α2 P	α2 T	α3	α3	α3
Mean % MM	0.32	0.31	0.30	0.48	0.39	0.54	0.54	0.33	0.30	0.10	0.47	0.27	0.31	0.31	0.24
MM in the rule sets[2]	NO	YES	YES	YES	NO	NO	YES	NO	NO	NO	YES	YES	NO	NO	NO

*Amino acid sites were displayed based on the functional α1, α2, and α3 domains of the HLA-A molecule. MM, mismatch, P, peptide contact site; T, TCR site; P/T, a peptide and/or a TCR contact site.

[1]A rule discovery algorithm, JRip was used to identify mismatches and matches at various sites that appear to work in concert with others in classifying DGF as described in *Materials and Methods*. A dichotomous amino acid mismatches coding was used in this algorithm. The population examined consisted of Non-AA recipients. Importantly, the presence of a "no mismatch" site at a specific position does not necessarily indicate a match at the HLA-A antigen level. However, "no mismatch" at a few particular amino acid sites can separate HLA-A antigens into HLA-A families. (Supporting Information, see Table S3).

[2]Note that the rule sets also include the amino acid mismatched sites identified in Table 2.

Table 4. Association of HLA-A amino acid mismatches with delayed graft function: Results of the multivariate logistic regression analysis for Caucasian recipients[1]

	All recipients					PRA-negative recipients				
Position	OR (SD)	LCL (SD)	UCL (SD)	*P*-value (SD)	*P*-value range	OR (SD)	LCL (SD)	UCL (SD)	*P*-value (SD)	*P*-value range
9	1.39 (0.09)	0.78 (0.05)	2.47 (0.16)	0.2792 (0.092)	0.0980–0.5790	1.65 (0.11)	0.93 (0.06)	2.95 (0.2)	0.0971 (0.037)	0.0157–0.1921
62	2.35 (0.12)	1.15 (0.06)	4.79 (0.26)	0.0201 (0.008)	0.0066–0.0540	—	—	—	—	—
90	2.46 (0.08)	1.39 (0.04)	4.36 (0.15)	0.0022 (0.001)	0.0011–0.0053	2.93 (0.12)	1.42 (0.06)	6.03 (0.22)	0.0038 (0.001)	0.0018–0.0094
95	—	—	—	—	—	2.51 (0.28)	1.22 (0.13)	5.17 (0.59)	0.0179 (0.021)	0.0006–0.1316
127	0.59 (0.02)	0.32 (0.01)	1.1 (0.03)	0.0989 (0.022)	0.0624–0.1870	0.51 (0.03)	0.24 (0.01)	1.07 (0.06)	0.0783 (0.028)	0.0275–0.1914
149	—	—	—	—	—	0.29 (0.03)	0.1 (0.01)	0.83 (0.06)	0.0225 (0.010)	0.0104–0.0857
184	0.56 (0.03)	0.32 (0.02)	0.99 (0.06)	0.0498 (0.023)	0.0183–0.1211	—	—	—	—	—
193	—	—	—	—	—	0.44 (0.03)	0.23 (0.02)	0.84 (0.05)	0.0137 (0.008)	0.0029–0.0432
246	0.53 (0.04)	0.3 (0.02)	0.94 (0.06)	0.0325 (0.018)	0.0081–0.0870	—	—	—	—	—

[1]Nine variable sites out of the original 22 sites were fitted in this final logistic regression model. These sites were selected based on a stepwise logistic regression model wherein amino acid candidate sites were identified as described in *Materials and Methods*. A dichotomous amino acid mismatches coding was used. Subjects are Caucasian recipients.

LCL, average lower 95% confidence limit over 100 imputations; OR, average odds ratio over 100 imputations; *P*-value, average *P*-value over 100 imputations; UCL, average upper 95% confidence limit over 100 imputations.

In the transplant-recipient group that included the PRA-positive recipients, increased risk of DGF was found to be associated with mismatches at one functionally important position at the antigen recognition site 62 (OR, 2.35; CI, 1.15–4.79; $P = 0.020$). Similarly, mismatch at site 95 was associated with increased risk of DGF in the PRA-negative patients (OR, 2.51; CI 1.22–5.17; $P = 0.017$). Notably, position 90 (in the α1 domain, not a peptide site) was also highly associated with DGF in the transplant recipient group that included the PRA-positive subjects and in the PRA-negative group (OR, 2.46; CI, 1.39–4.36; $P = 0.0022$ and OR, 2.93, CI 1.42–6.03; $P = 0.0038$, respectively). Variability at positions 62 and 90 was previously reported to be restricted to the HLA-A locus when compared to the HLA-B locus (16, 17). Because of a high correlation between mismatches at position 90 with mismatches at 163 ($r = 0.96$), a mismatch at position 163 was not selected in this logistic regression model. However, when we adjusted for all other mismatched sites except for position 90, the effect of a mismatch at position 163 on DGF was essentially the same as shown in Table 4 for a mismatch at 90 (data not shown). Of interest, a decreased risk of DGF was found to be associated with mismatches at HLA-A family-specific sites: position 149 (OR, 0.29; CI, 0.10–0.83; $P = 0.022$) and position 193 (OR, 0.44; CI, 0.23–0.84; $P = 0.013$) in the PRA-negative patients; position 184 (OR, 0.56; CI, 0.32–0.99; $P = 0.049$) and position 246 (OR, 0.53; CI, 0.30–0.94; $P = 0.032$) in the transplant group that included the PRA-positive subjects (see Table 4). The apparent disparity in the assortment of mismatched sites associated with DGF observed between the two groups of recipients could be because of a difference in the distribution of HLA-A alleles in the donor population in the two groups, as donor selection can be affected by the presence of donor-directed anti-HLA antibody during cross-matching. Very similar associations of amino acid mismatched sites with DGF were observed using another 100 imputed datasets, indicating the reproducibility of these findings (data not shown).

Discussion

In this prospective cohort study of recipients of deceased-donor renal transplants, we showed that incremental HLA-A mismatches detected at the allele level were significantly associated with DGF after adjusting for matching at the antigen level and multiple covariates. In contrast, mismatches because of HLA-B, -DRB1 antigens and alleles were not significantly associated with DGF (see Table 1). This finding prompted us to evaluate the association of amino acid mismatches with DGF at individual variable sites of HLA-A molecules.

In Caucasian recipients, we showed in a multivariate logistic regression model, adjusted for nonimmunologic risk factors, that DGF was significantly associated with mismatches at position 62, 95, and 163 (see Table 4). These amino acid sites are known to be functionally important sites at the antigen recognition site of the HLA-A molecule. Our finding is consistent with other studies, indicating that immunologic factors may increase the risk of DGF (10). Possible interactions that might exist between HLA-A amino acid mismatches and the nonimmunological risk factors (donor age, recipient race, longer CIT, and donor cause of death) were not evaluated in our study.

Position 62 is particularly interesting. It is a peptide and TCR contact site, which maps onto the α1 domain on the surface of the binding site of the HLA molecule. It is the most variable site in the HLA-A molecule and its variability is predominantly restricted to the HLA-A locus when compared to the same site in HLA-B (11, 16, 17). Position 95 maps onto the β-strand that forms the base of the antigen-presenting domain; consequently it would affect the conformation of presented peptides. Position 163 is located in the α2 region and, in addition to being a TCR contact site, is a peptide contact site (11). The association of DGF with mismatches at these peptide and TCR contact sites may be a reflection of their importance in organ alloreactivity influencing the development of DGF. The major source of alloreactivity is thought to be a TCR cross-reactivity between distinct HLA allelic variants from the same HLA class I (or class II) gene (11, 18). Peptide-driven recognition of allogeneic HLA molecules is thought to be the most frequent cause of alloreactivity. In addition, alloreactive T cells can also recognize allelic polymorphisms at TCR contact sites of the HLA molecule (11, 19). Particularly remarkable was that position 90, which is located in the α1 domain but outside the binding groove, was also significantly associated with DGF (see Table 4). Like position 62, variability at position 90 is restricted to the HLA-A locus (11, 16). However, the association of mismatches at position 90 with DGF could be secondary to the very high correlation that exists between mismatches at positions 90 and 163 ($r = 0.96$) in this population. Alternatively, variability at position 90 may also affect alloreactivity indirectly by inducing conformational changes in the sites of interactions (17, 20).

A decreased risk of DGF was associated with mismatches at HLA-A family-specific sites in the α2 (position 149, a TCR contact) and α3 domains (positions 149, 193, and 246) (Table S3). HLA-A families are closely related to HLA-A lineages that separate MHC class I A locus in the chimpanzee and gorilla. Importantly, differences in nucleotide substitutions in pairs of

HLA-A alleles that belong to separate families are larger than those within families (21, 22). Furthermore, many of the polymorphisms at HLA-A family-specific sites including extended amino acid sequence motifs appear to have been fixed during evolution, as they are preserved in the MHC class I A alleles of the chimpanzee or gorilla (22). Our findings indicate that evolutionary features of HLA-A polymorphism separating HLA-A families and lineages among donor-recipient pairs may correlate with the magnitude of alloreactivity influencing the development of DGF.

No significant differences in the assortment of mismatched candidate sites were observed between the dichotomous and trichotmous coding of amino acid mismatches. However, a potential dose-dependent effect of specific amino acid residues at individual amino acid sites was not examined in this study. A few additional candidate sites (positions 12, 17, 43, and 109) were identified by feature selection methods, but were not tested in the logistic regression analysis because only a small proportion (3 to 7%) of donor-recipient pairs were mismatched at these sites (see Fig. S2), limiting our statistical power. Furthermore, examination of mismatches in AA recipients revealed several sites (74, 77, 107, 161, and 184) that were associated with DGF and were frequently mismatched (25 to 30% of recipients, data not shown). However, these sites were not further evaluated in this study because of the small number of recipients.

The statistical approaches used in this study allowed an evaluation of the association between amino acid mismatches at variable sites and DGF, regardless of the combination of HLA-A antigen or allele mismatch. This differs from previous studies, which have examined only amino acid mismatches among HLA alleles from the same antigen groups (23). However, our study had some limitations. The sensitivity of the logistic regression analysis may be restricted by some degree of uncertainty, introduced by the partial imputation of the HLA-A alleles. For example, mismatches at functionally important sites, such as position 9 and 156, were significantly associated with DGF in the univariate analysis but did not remain in the multivariate logistic regression model. The lack of association of mismatches at position 9 and 156 with DGF could be because of (*i*) an uncertainty introduced by the partial imputation of the HLA-A alleles, (*ii*) a constraint imposed by the strong correlations that were observed between a few mismatches sites (see Table S2), and (*iii*) a relatively small number of subjects in our cohort.

In summary, we observed that amino acid polymorphism at functionally important positions at the antigen-recognition site of HLA-A molecule has a significant influence on DGF. In addition, our findings suggest that evolutionary features of HLA-A polymorphism separating HLA-A families and lineages among donor-recipient pairs may correlate with the magnitude of alloreactivity influencing the development of DGF. Our findings support the hypothesis that HLA-A amino acid polymorphism can be used to assess the influence of the degree of HLA-A matching on DGF. Further studies with larger cohorts are needed to evaluate the effect of the interactions of HLA-A amino acid mismatches with covariates, such as CIT and pretransplant anti-HLA antibodies, as well as the influence of HLA amino acid mismatches on other outcomes, such as acute rejection and allograft loss. Based on these findings, a predictor model for allograft outcomes could then be developed using a statistical approach similar to the one used in this study.

Materials and Methods

Study Design and Procedures. We enrolled 705 recipients of deceased-donor renal transplants into a prospective cohort study during the period April 1998 through November 2001, as previously described (14). Patients were recruited at the time of transplantation from eight centers served by the Gift of Life Donor Program in eastern Pennsylvania, as previously reported (15). Patients who were more than 18 years of age and had received a deceased-donor renal transplant were eligible for this study. Informed consent was obtained from each study participant. During the initial transplant hospitalization, study participants were interviewed regarding their demographic and medical characteristics. Race was determined by self-report. The baseline demographic, medical, and surgical characteristics at the time of transplantation were previously reported (15). The study protocol was approved by the Institutional Review Board at the University of Pennsylvania.

HLA Typing. Whole blood that remained from routine clinical testing was used as a source of DNA for the transplant recipients. Whole blood or lymph node tissue was provided as a source of DNA for the donors by Gift of Life Organ Procurement Organization.

Confirmatory HLA-A and -B low- to intermediate-level resolution typing and high resolution (allele-level) DNA-based typing of HLA-A, -B, and -DRB1 for all study subjects (and their respective donors) were performed in a single center. Low-resolution typing was performed by DNA-based methods for HLA-A and -B on all donors and recipients to better define antigen splits and minimize errors resulting from antigens that are difficult to type serologically (24). High-resolution HLA-A, -B, and -DRB1 typing was performed only in donor-recipient pairs that had fewer than two mismatches at their respective HLA loci by low resolution typing. Of the 705 recipients, 364 (51.6%) DNA-based methods were used for HLA-DRB1 low-resolution typing; for the remaining 341 (48.4%), low-resolution typing was performed by serological methods in multiple centers using at least two commercial HLA typing trays. We resolved low-resolution (serological or antigen level) disparities by converting the DNA-based typing to its lower-level serologic equivalent. We described all of these procedures in a previous article (14, 24).

HLA typing was performed prospectively during the period of 1997 to 2003. Because of limitations of the sequence specific primer (SSP) typing methodology, allele-level assignment was not always possible because of the presence of ambiguous allele combinations. Therefore, when SSP typing indicated one of several possible alleles, the allele assignment was made by ruling out extremely rare alleles (frequency <0.001) for each subjects' racial population and by taking into account established HLA haplotype frequencies (25)

Imputation of Missing HLA-A Alleles. Because of a lack of sufficient archived samples, it was not feasible to perform allele-level typing on all donor-recipient pairs. HLA-A allele data were obtained through high-resolution typing only when recipients and their donors were matched at the antigen level. For recipients and donors who were mismatched at the low-resolution (antigen) level, we imputed donor and recipient alleles using the following procedure. We used the allele frequency data from the nonimputed recipient subjects (Caucasians and African Americans, separately) in our cohort as the source of data for the imputation of recipient alleles in this cohort, as previously described (26). We imputed the HLA-A alleles of donors using population-based frequencies according to race (ref. 25 and www.allelefrequencies.net) and because these individuals were more likely to be similar to the general population. The rationale for this approach was that recipients are different from the general population because of the well-known associations of HLA alleles with autoimmune diseases usually seen in kidney transplant recipients. With this multiple imputation, we created a total of 100 separate HLA-A allele datasets, each representing a separate imputation (27). The mean allele frequencies from the 100 imputed datasets (Fig. S3) were found to be very similar to those of the nonimputed published data according to race (25, 28, 29).

Once the imputed datasets were created, we merged them with HLA-A amino acid sequence data using the HLA Sequence Database Alignment Tool (IMGT/HLA) database (www.ebi.ac.uk/imgt/hla) to create the full amino acid sequence for each donor and recipient. Based on this array of amino acid sequence data, we compared 66 polymorphic amino acid sites to identify site-specific donor-recipient mismatches. Donor-recipient mismatches were characterized at each residue across the length of the HLA-A sequence as 0 (no mismatch), 1 (mismatch on one of the allele pairs between donor and recipient), or 2 (mismatch on both of the allele pairs). A second set of 100 imputed datasets was created using a dichotomous coding scheme (0, no mismatch; 1, any mismatch) applied to each of the 66 aa positions. The final imputed datasets contained these 66 mismatch variables in addition to the following demographic and clinical variables: recipient and donor age, race and sex, cause of donor death, and presence or absence of prior reactive antibodies in recipients. The outcome variable was DGF, coded dichotomously as Yes or No. Of the original 705 recipients, 8 had missing data for DGF; thus, each imputed dataset contained 697 records, each representing a unique donor-recipient pair.

Statistical Analysis. HLA antigen (serologic) and allele (high resolution) mismatches were evaluated using the entire cohort including non-AA and AA

recipients. Mismatches were defined separately for HLA-A, -B, and -DRB1 loci. The mismatch at antigen level was defined as a three-level ordinal variable (0, 1, or 2 mismatches). Additional mismatches that were identified at allele level were defined as a binary quantity. To explore the association between HLA mismatches and DGF, we first fit a logistic regression model using the six HLA mismatches as independent variables. The model was then refitted by adjusting for the following covariates: recipient race, dialysis pretransplant, previous transplant, donor age, and CIT.

Analyzes of HLA-A amino acid mismatches were restricted to Caucasian recipients only. All analyses were performed separately on each of 100 imputed datasets and the results pooled across them using the usual rules for multiple imputation (27). We describe the analysis of the separate imputed datasets. For each imputed HLA dataset, we identified the mismatched sites associated with DGF by means of several machine-learning and statistical feature selection methods, including correlation-based feature selection (30), and χ^2 and univariate logistic regression, respectively. The feature selection process used a supervised learning approach with delayed graft function as the class variable, focusing on position-specific mismatches, and was applied to each dataset separately.

In addition to the univariate analysis (where one site was considered at a time, conditioned on DGF), we also used a rule discovery method, JRip, to identify mismatches and matches that appeared to work in concert with others in classifying DGF as present or absent. JRip is a variant of RIPPER (Repeated Incremental Pruning to Produce Error Reduction) (31). We then pooled these selected mismatched sites into one single group of candidate variables for further analysis using multivariate logistic regression modeling.

To evaluate the joint effects of the selected amino acid sites on DGF, we performed a stepwise multivariate logistic regression analysis separately on each imputed data set using the following covariates: donor age, recipient age, donor cause of death, and cold ischemia time, with $P < 0.25$ as the inclusion criterion. Finally, we selected variable mismatched sites, which were observed with a frequency of 20% or higher over the 100 imputations, with $P < 0.05$, and fitted a final logistic regression model for each permuted dataset to assess the joint effects of the selected mismatched sites on DGF. We repeated these analyses separately for recipients without pretransplant anti-HLA antibodies (PRA-negative subjects), as the distribution of HLA-A alleles in the donor population may be influenced by the HLA antibody cross-match in the PRA-positive subjects. Statistical analyses were not done on African American recipients because of the small number of subjects combined with a higher degree of allelic variability in these individuals. All machine learning-based feature selection analyses were performed using Weka (32) and the statistical analyzes were performed using SAS (SAS Institute).

ACKNOWLEDGMENTS. We thank the staff and patients at the eight participating transplant centers in eastern Pennsylvania, Dr. Paula Romano and Robert Scott for providing archived donor material, and The Gift of Life Donor Program. This study was supported in part by Grants R01-AI043295 and K24-DK02651 (to H.I.F.), Grant K23-DK062829 (to A.K.I.), and the Norman S. Coplon Grant from Satellite Healthcare (to A.K.I.).

1. Snell G, Smith P, Gabrielson F (1953) Analysis of the histocompatibility-2 locus in the mouse. *J Natl Cancer Inst* 14:457–480.
2. Dausset J, Rapaport F, Colombani J, Feingold N (1965) A leucocyte group and its relationship to tissue histocompatibility in man. *Transplantation* 3:701–705.
3. Takemoto S, Terasaki P, Gjertson D, Cecka J (2000) Twelve years' experience with national sharing of HLA-matched cadaveric kidneys for transplantation. *N Engl J Med* 343:1078–1084.
4. Cecka JM (2004) The OPTN/UNOS Renal Transplant Registry 2003. *Clinical Transpl* 2003:1–12.
5. Kamar N, *et al.* (2006) Impact of early or delayed cyclosporine on delayed graft function in renal transplant recipients: A randomized, multicenter study. *Amer J Transplant* 6:1042–1048.
6. Quiroga I, *et al.* (2006) Major effects of delayed graft function and cold ischaemia time on renal allograft survival. *Nephrol Dial Transplant* 21:1689–1696.
7. Mikhalski D, *et al.* (2008) Cold ischemia is a major determinant of acute rejection and renal graft survival in the modern era of immunosuppression. *Transplantation* 85:S3–S9.
8. Feldman H, *et al.* (1998) Race and delayed kidney allograft function. *Nephrol Dial Transplant* 13:704–710.
9. Feldman HI, *et al.* (1996) Delayed function reduces renal allograft survival independent of acute rejection. *Nephrol Dial Transplant* 11:1306–1313.
10. Irish WD, *et al.* (2003) Nomogram for predicting the likelihood of delayed graft function in adult cadaveric renal transplant recipients. *J Am Soc Nephrol* 14:2967–2974.
11. Reche PA, Reinherz EL (2003) Sequence variability analysis of human class I and class II MHC molecules: Functional and structural correlates of amino acid polymorphisms. *J Mol Biol* 331:623–641.
12. Bjorkman P, *et al.* (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329:512–518.
13. Bjorkman P, *et al.* (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 329:506–512.
14. Kamoun M, *et al.* (2007) Assessment of differences in HLA-A, -B, and -DRB1 allele mismatches among African-American and non-African-American recipients of deceased kidney transplants. *Transplant Proc* 2007;39:55–63.
15. Weng FL, *et al.* (2005) Race and electronically measured adherence to immunosuppressive medications after deceased donor renal transplantation. *J Am Soc Nephrol* 16:1839–1848.
16. Hedrick PW, Whittam TS, Parham P (1991) Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. *Proc Natl Acad Sci USA* 88:5897–5901.
17. Parham P, *et al.* (1988) Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc Natl Acad Sci USA* 85:4005–4009.
18. Archbold JK, *et al.* (2006) Alloreactivity between disparate cognate and allogeneic pMHC-I complexes is the result of highly focused, peptide-dependent structural mimicry. *J Biol Chem* 281:34324–34332.
19. Lombardi G, Barber L, Sidhu S, Batchelor J, Lechler R (1991) The specificity of alloreactive T cells is determined by MHC polymorphisms which contact the T cell receptor and which influence peptide binding. *Int Immunol* 3:769–775.
20. Teng JMC, Hogan KT (1994) Residues outside of the HLA-A2 peptide-binding groove can abrogate or enhance recognition of influenza virus matrix peptide pulsed cells by cytotoxic T lymphocytes. *Mol Immunol* 31:445–457.
21. Lawlor D, Ward F, Ennis P, Jackson A, Parham P (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 335:268–271.
22. Lawlor DA, Warren E, Taylor P, Parham P (1991) Gorilla class I Major histocompatibility complex alleles: comparison to human and chimpanzee Class I. *J Exp Med* 174:1491–1509.
23. Kawase T, *et al.* (2007) High-risk HLA allele mismatch combinations responsible for severe acute graft-versus-host disease and implication for its molecular mechanism. *Blood* 110:2235–2241.
24. Gourley IS, *et al.* (2002) HLA class I typing of volunteers for a bone marrow registry: QC analysis by DNA-based methodology identifies serological typing discrepancies in the assignment of HLA-A and B antigens. *Tissue Antigens* 59:211–215.
25. Cao K, *et al.* (2001) Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum Immunol* 62:1009–1030.
26. Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys* (John Wiley & Sons, New York).
27. Schafer JL (1997) *Analysis of Incomplete Multivariate Data* (Chapman & Hall, London).
28. Schipper RF, *et al.* (1997) HLA gene and haplotype frequencies in bone marrow donors worldwide registries. *Hum Immunol* 52:54–71.
29. Zachary AA, Steinberg AG, Bias WB, Leffell MS (1996) The frequencies of HLA alleles and haplotypes and their distribution among donors and renal patients in the UNOS registry. *Transplantation* 62:272–283.
30. Frank E, Hall M, Trigg L, Holmes G, Witten I (2004) Data mining in bioinformatics using Weka. *Bioinformatics* 12:2479–2481.
31. Cohen W (1995) in *Fast Effective Rule Induction*, eds Prieditis A, Russell S (Morgan Kaufmann), pp 115–123.
32. Witten I, Frank E (2005) *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, San Francisco).