



Eleven Amino Acids of HLA-DRB1 and Fifteen Amino Acids of HLA-DRB3, 4, and 5 Include Potentially Causal Residues Responsible for the Risk of Childhood Type 1 Diabetes

Lue Ping Zhao,¹ George K. Papadopoulos,^{2,5} William W. Kwok,³ Bryan Xu,⁴ Matthew Kong,⁵ Antonis K. Moustakas,⁶ George P. Bondinas,² Annelie Carlsson,⁷ Helena Elding-Larsson,⁸ Johnny Ludvigsson,⁹ Claude Marcus,¹⁰ Martina Persson,¹⁰ Ulf Samuelsson,⁹ Ruihan Wang,¹¹ Chul-Woo Pyo,¹¹ Wyatt C. Nelson,¹¹ Daniel E. Geraghty,¹¹ and Åke Lernmark⁷

Diabetes 2019;68:1692–1704 | <https://doi.org/10.2337/db19-0273>

Next-generation targeted sequencing of HLA-DRB1 and HLA-DRB3, -DRB4, and -DRB5 (abbreviated as DRB345) provides high resolution of functional variant positions to investigate their associations with type 1 diabetes risk and with autoantibodies against insulin (IAA), GAD65 (GADA), IA-2 (IA-2A), and ZnT8 (ZnT8A). To overcome exceptional DR sequence complexity as a result of high polymorphisms and extended linkage disequilibrium among the DR loci, we applied a novel recursive organizer (ROR) to discover disease-associated amino acid residues. ROR distills disease-associated DR sequences and identifies 11 residues of DRB1, sequences of which retain all significant associations observed by DR genes. Furthermore, all 11 residues locate under/adjointing the peptide-binding groove of DRB1, suggesting a plausible functional mechanism through peptide binding. The 15 residues of DRB345, located respectively in the β 49–55 homodimerization patch and on the face of the molecule shown to interact with and bind to the accessory molecule CD4, retain their significant disease

associations. Further ROR analysis of DR associations with autoantibodies finds that DRB1 residues significantly associated with ZnT8A and DRB345 residues with GADA. The strongest association is between four residues (β 14, β 25, β 71, and β 73) and IA-2A, in which the sequence ERKA confers a risk association (odds ratio 2.15, $P = 10^{-18}$), and another sequence, ERKG, confers a protective association (odds ratio 0.59, $P = 10^{-11}$), despite a difference of only one amino acid. Because motifs of identified residues capture potentially causal DR associations with type 1 diabetes, this list of residuals is expected to include corresponding causal residues in this study population.

Genetic studies, from classical HLA studies or genome-wide association studies (1–3), have confirmed that the strongest genetic associations of type 1 diabetes (T1D), potentially causal, are with the HLA-DR and -DQ genes on chromosome 6, dwarfing ~50 non-HLA genetic associations (1,4–6). While

¹Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA

²Laboratory of Biophysics, Biochemistry, Biomaterials and Bioprocessing, Faculty of Agricultural Technology, Technological Educational Institute of Epirus, Arta, Greece, presently known as Department of Agriculture, University of Ioannina, Ioannina, Greece

³Benaroya Research Institute at Virginia Mason, Seattle, WA

⁴College of Letters and Sciences, University of California, Berkeley, CA

⁵Department of Computer Sciences, Carnegie Mellon University, Pittsburgh, PA

⁶Department of Food Science and Technology, Ionian University, Argostoli, Cephalonia, Greece

⁷Department of Clinical Sciences, Lund University/Clinical Research Centre, Skåne University Hospital, Malmö, Sweden

⁸Department of Pediatrics, Lund University, Lund, Sweden

⁹Crown Princess Victoria Children's Hospital, Region Östergötland, and Division of Pediatrics, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

¹⁰Department of Clinical Science and Education and Institution of Medicine, Clinical Epidemiology, Karolinska Institutet, Solna, Sweden

¹¹Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

Corresponding authors: Lue Ping Zhao, lzhao@fredhutch.org; George K. Papadopoulos, papadpg@gmail.com; and Åke Lernmark, ake.lernmark@med.lu.se
Received 13 March 2019 and accepted 21 May 2019

This article contains Supplementary Data online at <http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db19-0273/-/DC1>.

B.X. and M.K. were summer interns at Fred Hutchinson Cancer Research Center during the summer months of 2018.

G.P.B. is an adjunct member of the Laboratory.

[§]G.K.P. retired from the Technological Educational Institute of Epirus.

© 2019 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <http://www.diabetesjournals.org/content/license>.

many studies attempted to distinguish two closely located HLA-DR and -DQ associations (7–10), extending to the HLA-DP locus (11), a recent effort aims to investigate individual amino acid residues (or residues for short) and their associations with T1D, following the shared epitope hypothesis in rheumatoid arthritis research (12). One approach is to scan individual residues by testing differential frequencies of individual residues between patients and their control subjects (13), but this fails to account for linkage disequilibrium (LD) among residues. An alternative approach is conditional analysis (14), which repeats residue-specific association tests while adjusting for selected residues. Because of the extensive LD within the HLA genes, it is nearly impossible to deconvolute empirically observed associations into biological versus LD-induced associations; results from this recent study lead to an ongoing controversy. At issue is whether identified residues through conditional analysis are “causal” or just the best surrogate markers for defining HLA genotype because they contain the most discriminative ability (15). Additionally, HLA genotypes used in the conditional study (14) are imputed at an intermediate resolution with an average accuracy of 96.7%, but their accuracy for the majority of HLA alleles (that are less frequent) is substantially lower (16). Genotyping quality could negatively affect the precision of deducing causal residues.

Using next-generation targeted sequencing (NGTS) technology, we recently sequenced 962 pediatric patients with T1D and 636 control subjects on four DR genes (HLA-DRB1, -DRB3, -DRB4, and -DRB5), explored their associations with T1D, and reported their association results earlier with a subset of 448 control subjects (17). For clarity, we use DRB1 to denote the gene and DR β 1 to denote the corresponding protein sequence. A similar convention is applicable to HLA-DRB3, -DRB4, and -DRB5 genes. Unlike diploid DRB1, each chromosome carries at most one of the DRB3, DRB4, and DRB5 genes (18,19). For simplicity, we use DRB345 to denote the three separate genes. Additionally, we developed a novel method, known as recursive organizer (ROR), to achieve statistical parsimony by likelihood-based associations while retaining the sequence-specific LD among all residues and sequence motifs of multiple selected amino acids that retain DR-T1D associations (20). In essence, ROR shared the same analytic objective with the conditional analysis (14), but avoided the influence of local LD on eventually selected residues, assuring inclusion of all possible causal variants in the discovery process.

Here, we applied ROR to the case-control data of T1D and high-quality HLA gene sequences to identify potentially causal residues in DRB1 and DRB345. Among all patients with T1D, the study also measured a panel of autoantibodies against GAD65 (GADA), insulin (IAA), IA-2 (IA-2A), and the three variants (residue R, W, or Q on position 325) of ZnT8 autoantibodies (ZnT8A) (21–23) at the time of T1D diagnosis. This availability also allowed us to investigate HLA-DR associations with increased titer of these autoantibodies at the time of the disease diagnosis.

RESEARCH DESIGN AND METHODS

Study Design and Population

This case-control study involved 962 patients from the nationwide Swedish Better Diabetes Diagnosis (BDD) study (21,22,24) and included 636 nationwide and geographically representative control subjects (25), which has been detailed earlier (17). The Karolinska Institute Ethics Board approved the BDD study (2004/1:9).

DNA Extraction

The Plasmid Maxi Isolation Kit (QIAGEN, Bothell, WA) was used to isolate DNA according to the manufacturer's instructions from frozen whole-blood samples of patients and control subjects.

HLA NGTS Analysis

HLA typing was carried out using the ScisGo HLA v6 typing kit (Cisco Genetics, Seattle, WA). Briefly, the method uses an amplicon-based two-stage PCR, followed by sample pooling and sequencing using a MiSeq v2 PE 500 (Illumina, San Diego, CA). The protocol yields three-field coverage of all HLA loci, including exons 1–4 for DRB1 and DRB345, and genotypes of two-field coverage are used here. Phase is determined in part by overlapping sequences for HLA class I and database lookup for HLA class II (18). These types were converted into amino acid sequences corresponding to amino acids in the β -sheet of codon β 1–237 as well as those residues in the signal peptide from –29 to –1. The mature protein started from codon β 1, extending to codon β 237.

Islet Autoantibodies

GADA, IA-2A, IAA, and three variants of ZnT8A (ZnT8-RA, ZnT8-WA, and ZnT8-QA) were determined in quantitative radiobinding assays using in-house standards to determine levels as previously described in detail (22,26).

Statistical and Molecular Simulation Analysis

By the case-control design, we assessed allelic and haplotypic association analysis of DR genes (27) (using the haplo.stats package of R) and then applied the ROR (20) to discover associated amino acids. To examine potential roles of these amino acids, we performed molecular simulation analysis. Details are provided in the Supplementary Data.

RESULTS

HLA-DR Associations at the Sequence Resolution

The disease association analysis of high-resolution DRB1 and DRB345 in this case-control study (636 control subjects and 962 patients) uncovered detailed DR associations with T1D, most of which are consistent with reported associations on DR-T1D. It is worth noting that associations are impressive, for example, of DRB1*03:01 (odds ratio [OR] 2.09, $P = 2.71 \times 10^{-20}$) and DRB1*04:01 (3.64, $P = 3.19 \times 10^{-64}$) (see Supplementary Table 1A for details). DRB1 had a total of 25 alleles with more than five copies in this study. Five risk alleles (P values highlighted in green) were found to have an OR ranging from 1.58 to 5.95, in which some P values were as low as 3.19×10^{-64} .

Table 1—A total of 22 alleles, formed by 11 amino acids, resulted from merging association-equivalent and highly similar alleles

ID	DRB1	Control	Patient	OR	HS	P-value	Allele Group	14	25	47	57	60	71	73	74	78	85	86	IA2A
1	ERFDYAAAYVG	9(0.71)	3(0.16)	0.22	-2.50	1.24E-02	DRB1*15:02	.	.	F	G
2	ERFDYAAAYVV	204(16.04)	10(0.52)	0.03	-16.47	6.35E-61	DRB1*15:01~*15:03	E	R	F	D	Y	A	A	A	Y	V	V	
3	ERFDYEAAYVG	56(4.4)	86(4.47)	1.02	0.09	9.28E-01	DRB1*13:02	.	.	F	.	.	E	G
4	ERFDYEAAYVV	102(8.02)	41(2.13)	0.27	-7.95	1.85E-15	DRB1*11:02~*11:03~*13:01	.	.	F	.	.	E	EREA
5	ERFDYKGRYVV	153(12.03)	484(25.16)	2.09	9.23	2.71E-20	DRB1*03:01	.	.	F	.	.	K	G	R	.	.	.	ERKG
6	ERFDYRAAYVG	68(5.35)	19(0.99)	0.18	-7.43	1.11E-13	DRB1*11:01~*13:05~*16:09	.	.	F	.	.	R	G
7	ERFDYRAAYVV	19(1.49)	2(0.1)	0.07	-4.56	5.10E-06	DRB1*11:04~*11:84	.	.	F	.	.	R	
8	ERFVSRAYAV	31(2.44)	14(0.73)	0.30	-3.95	7.67E-05	DRB1*12:01~*12:02	.	.	F	V	S	R	A
9	ERYAHRAEYVV	28(2.2)	5(0.26)	0.12	-5.18	2.17E-07	DRB1*14:01~*14:04~*14:54	.	.	Y	A	H	R	.	E	.	.	.	
10	ERYDYEAAYVG	11(0.86)	2(0.1)	0.12	-3.31	9.18E-04	DRB1*01:03	.	.	Y	.	.	E	G
11	ERYDYEAAYVV	7(0.55)	27(1.4)	2.55	2.25	2.47E-02	DRB1*04:02	.	.	Y	.	.	E	EREA
12	ERYDYKAAYVG	126(9.91)	694(36.07)	3.64	16.92	3.19E-64	DRB1*04:01	.	.	Y	.	.	K	G
13	ERYDYRAAYAV	13(1.02)	11(0.57)	0.56	-1.45	1.47E-01	DRB1*01:02	.	.	Y	.	.	R	A
14	ERYDYRAAYVG	152(11.95)	190(9.88)	0.83	-1.88	6.05E-02	DRB1*01:01~*04:08~*10:01~*14:02~*16:01~*16:02	.	.	Y	.	.	R	G
15	ERYDYRAAYVV	61(4.8)	146(7.59)	1.58	3.10	1.91E-03	DRB1*04:04	.	.	Y	.	.	R	ERRA
16	ERYDYRAEYVG	12(0.94)	2(0.1)	0.11	-3.53	4.23E-04	DRB1*04:07	.	.	Y	.	.	R	.	E	.	.	.	G
17	ERYDYRAEYVV	12(0.94)	2(0.1)	0.11	-3.53	4.23E-04	DRB1*04:03~*04:06	.	.	Y	.	.	R	.	E	.	.	.	
18	ERYSYKAAYVG	10(0.79)	5(0.26)	0.33	-2.14	3.27E-02	DRB1*13:03	.	.	Y	S	.	K	G
19	ERYSYRAAYVG	4(0.31)	27(1.4)	4.46	3.09	2.00E-03	DRB1*04:05	.	.	Y	S	.	R	G
20	ERYSYRALYVG	60(4.72)	73(3.79)	0.80	-1.29	1.98E-01	DRB1*08:01~*08:03	.	.	Y	S	.	R	.	L	.	.	.	G
21	ERYVSRAEVVG	12(0.94)	30(1.56)	1.65	1.47	1.41E-01	DRB1*09:01	.	.	Y	V	S	R	.	E	V	.	.	G
22	KQYVSRGQVVG	114(8.96)	46(2.39)	0.27	-8.23	1.85E-16	DRB1*07:01	K	Q	Y	V	S	R	G	Q	V	.	.	G

The allelic frequencies are *n* (%) in 636 control subjects and 936 patients along with their estimated ORs, haplotype scores (HSs), and *P* values. These amino acid positions in the Protein Data Bank are listed above the individual amino acids. The dots are used to denote the same amino acid as the referenced one highlighted in red. Highlighted amino acids, ERKA, in blue, are involved in associations with IA-2A.

Meanwhile, 12 protective alleles (*P* values highlighted in red) had an OR ranging from 0.03 to 0.33 (*P* value was as low as 3.99×10^{-60}). Note that if $P < 0.05$ (without correcting for multiple comparisons), it is declared to be significant and is color-coded for risk/positive (green) or protective/negative (red) associations. Furthermore, some exceptionally small *P* values are approximated from Gaussian normal distribution and could best be interpreted qualitatively as extremely significant.

DRB345 had nine alleles (Supplementary Table 1B). Two risk alleles (DRB3*01:01 and DRB4*01:03) had significant associations (ORs 1.36 and 2.00, $P = 3.26 \times 10^{-5}$ and 9.94×10^{-49} , respectively). Five protective alleles associated with T1D had an OR ranging from 0.04 to 0.80.

Haplotyping DRB1 and DRB345 genes resulted in 28 haplotypes (Supplementary Table 1C). Six haplotypes, colored green, associated with an increased T1D risk, while 14 haplotypes were protective. Most haplotypes are uniquely formed by one DRB1 and DRB345 allele, and hence, corresponding haplotypic associations are attributed to both genes. There are three haplotype pairs (4, 5), (13, 14), and (21, 22) (highlighted in light blue font), in which one DRB1 allele shares haplotypes with two DRB345 alleles. Such pairs provide an opportunity to investigate the independent and interactive role of DRB345 on the corresponding DRB1 allele. For example, for individuals carrying DRB1*03:01, the DRB3*01:01 conferred an OR of 1.95, and DRB3*02:02 had an OR of 2.94, implying that DRB3*02:02 adds to the risk. Yet, by itself, DRB3*02:02 was supposedly protective (OR 0.48).

The current study included seven DR4 alleles (DRB1*04:01, *04:02, etc., in Supplementary Table 1A). Despite

that they had the same antigen (DR4), their associations with T1D were influenced by polymorphic residues; three DR4 alleles were associated with T1D risk, three were protective, and one had a neutral association (Supplementary Table 1A). Hence, dissecting responsible residues would provide insights into how genetic polymorphisms influence T1D susceptibility.

Observed associations with individual alleles/haplotypes of DRB1 and/or DRB345 are largely consistent with reported associations in the literature. With the better precision of NGTS and large sample size, association results presented here (Supplementary Table 1) constitute a synthesis of DR-T1D associations in a European population.

Parsimonious Sequences of 11 Residues Capture Disease Associations With DRB1

Recognizing DRB1 polymorphisms, extensive LD, and complex associations with T1D, ROR sought to achieve a parsimony by recursively eliminating residues without altering the observed disease associations with full DRB1 sequence alleles (Supplementary Table 1A). Starting with 266 residues, in the HLA-DRB1 and DRB345 signaling peptides and mature protein structures, ROR recursively eliminated residues and eventually led to a subset of residues (see ROR methodology in the Supplementary Data). Through the ROR process, we computed an overall *P* value and incremental *P* values that measure significances of all and incrementally deleted residues (Fig. 1A). Precisely which residues were deleted step by step, eventually leading to a group of 11 residues selected at ROR step 12, is

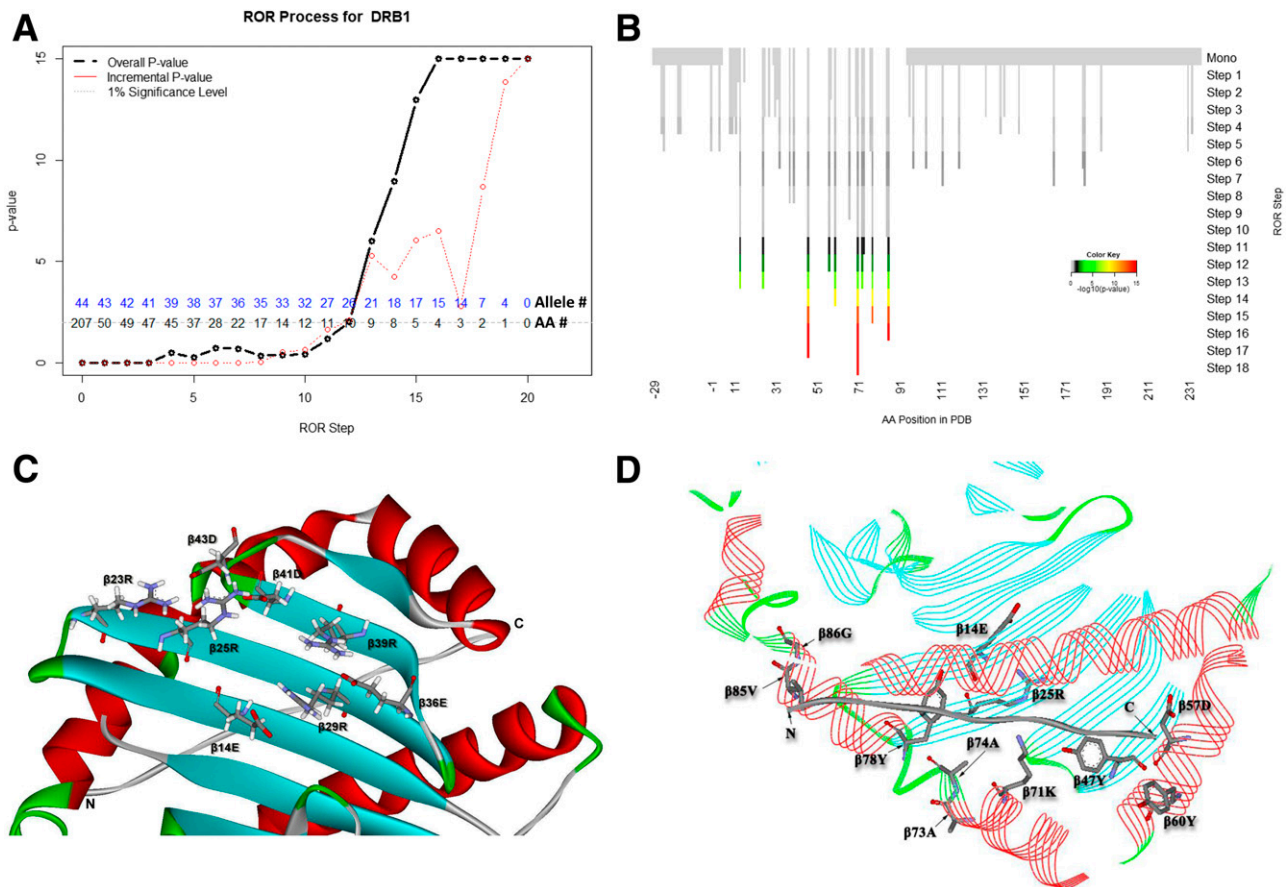


Figure 1—Results from analyzing T1D association with HLA-DRB1. **A:** Results from analyzing DRB1 association with T1D. The logarithmic *P* values (black line) were computed from the likelihood χ^2 test over successive ROR steps, at each of which the numbers of amino acids (AAs) were gradually deleted and the number of unique alleles reduced; comparable *P* values for incremental change (red line) were also computed. **B:** Results from analyzing DRB1 association with T1D, visually displaying how specific AAs were deleted throughout the ROR procedure. Color intensity corresponds to the logarithmic *P* values. **C:** A typical HLA-DR molecule (DRB1*01:01, bound to native peptide from HLA-A, 1AQD.pdb) showing the interactions of key residues in the bottom side of the β -sheet floor of the α 1 β 1 domain (below the antigen-binding groove). In the majority of alleles, β 14 is a Glu, while β 25 is an Arg. The two interact extensively with each other as well as with neighboring residues β 31Phe, β 36Glu, β 39Arg, β 41Asp, and β 43Asp, mostly through electrostatic interactions and weaker cation- π interactions involving β 31Phe and β 39Arg. This set of interactions is seen in essentially all crystal structures of HLA-DR molecules known to date. Water molecules documented in the crystal structure and bridging interactions between polar side chains are not shown for clarity. The antigenic peptide is in the gray ribbon above the β -sheet floor: N indicates its amino terminus, while C indicates its carboxy terminus. **D:** The crystal structure of the heterodimer HLA-DR α /HLA-DRB1:04:01 in complex with an antigenic peptide showing mostly the antigenic peptide-binding domains α 1 β 1 (2seb.pdb) as well as the 11 AAs identified by ROR and located within the binding groove (nine residues) or right below (two residues). The MHC II molecule is shown in line ribbons colored according to secondary structure (red, α -helix; turquoise, β -sheet; green, random coil), with the β 1-helix at the bottom of the picture and the α 1-helix at the top. The antigenic peptide backbone is shown in gray, with its amino terminus labeled as N and its carboxy terminus labeled as C. The 11 residues revealed by ROR to play a role in the pathogenesis of T1D within HLA-DRB1 alleles are shown in stick form (red, oxygen; blue, nitrogen; gray, carbon; no hydrogens shown). Residues β 14Glu and β 25Arg are located below the β -sheet floor, and the entire network of interactions is shown in greater detail in panel C. The entire HLA-DRB1*04:01 complex has been tilted with respect to the conventional TCR view (shown in Supplementary Video for a simulated DRB1*04:01 complex tilted toward the reader so that all 11 residues, so-revealed by the ROR process, could be exposed to view). PDB, Protein Data Bank.

shown in Fig. 1B. Identified residues were β 14, β 25, β 47, β 57, β 60, β 71, β 73, β 74, β 78, β 85, and β 86 (Table 1). These 11 residues formed 22 unique sequence alleles (Table 1). Each unique sequence allele corresponded to one or more classical alleles and is referred to as an allele group (see Allele Group column in Table 1). Allelic association analysis of the 11 amino acid sequences revealed 5 risk alleles and 12 protective alleles (Table 1), capturing the full allelic associations listed in Supplementary Table 1A, while several uncommon alleles were merged

into their respective common alleles (e.g., DRB1*15:03 with DRB1*15:01 and DRB1*11:02 and *11:03 with DRB1*13:01).

Most of the 11 Residues Fall in the Peptide-Binding Groove of DRB1, While the Remainder Are Nearby

As shown in Fig. 1C, all 11 residues were located between positions 14 and 86. When we examined the crystal-resolved structure of DRB1*04:01 (Fig. 1D), we found that all 11 residues were nearby or in the peptide-binding

grove of DR. Notably, β 86 was in pocket 1, β 71 was at the border separating pockets 4 and 7 while being part of both pockets, β 74 and β 78 were in pocket 4, β 47 was in pocket 7, and β 57 was in pocket 9. Of the remaining residues, β 14 and β 25 were underneath the groove below pocket 4, β 60 was just over pocket 9, and β 73 was adjacent to pocket 4.

Four Residues of DRB1 Capture the Risk, Protective, or Neutral Associations With DR4 Subtypes

Upon eliminating seven monomorphic residues across all DR4 subtypes, associated with seven common DR4 subtypes (Supplementary Table 1A) with an uncommon subtype (DRB1*04:06) merged with DRB1*04:03 (Supplementary Table 2), one had seven alleles with four residues (β 57, β 71, β 74, and β 86) corresponding to DR4 subtypes. Evidently, the presence of amino acid E at β 74 is associated with the strong protection against T1D, an unusual protection exhibited by the otherwise high-risk set of DR4 alleles. We conducted a molecular simulation of the structure of the DRB1*04:03 molecule in complex with two human insulin A-chain peptides that could bind with high affinity to this molecule (27,28) (Fig. 2A and B, Supplementary Fig. 2, details of the method in the Supplementary Data). Figure 2A and B show the tight fitting of the insulin A8–19 insulin A-chain peptide with p4Leu, p6Gln, and p9Asn anchors, where the latter two anchors are not common for DRB1*04 alleles (29–31) (Fig. 2A and B). Indeed, it was evident that the combination of β 74Glu with β 13His, β 26Phe, β 70Gln, β 71Arg, and β 78Tyr of pocket 4 in the DRB1*04:03 protein excludes acidic anchors at this pocket, accommodating Leu. The proteins DRB1*04:04 and B1*04:05, variably conferring susceptibility to T1D, also accommodate p4Leu in the epitope of GAD65 555–567 (32).

Fifteen Residues Account for Disease Association With DRB345

Processing alleles associated with HLA-DRB345, the ROR procedure removed 195 monomorphic amino acids and recursively organized the remaining 72 residues with 14 unique sequence alleles. After the first step, ROR eliminated 49 amino acids, leaving 23 residues forming 14 unique alleles. Following four additional iterative steps, ROR achieved a parsimony at ROR step 5, where the *P* value exceeded 5% (Fig. 3A). Visually, the ROR procedure selected 15 residues at positions β 18, β 25, β 40, β 41, β 44, β 48, β 51, β 70, β 76, β 81, β 104, β 135, β 180, β 181, and β 187 (Fig. 3B). The allelic association analysis of DRB345 showed a total of two risk amino acid sequences and four protective amino acid sequences (Table 2), which captured the full DRB345 allelic associations (Supplementary Table 1B).

These 15 Residues Have Different Functional Roles in Their Causal Association With Disease

Of these residues, the ones from β 18–81 are in the α 1 β 1 antigenic peptide-binding domain, while the

remainder are in the α 2 β 2 accessory domain of the DR β 345 molecules. We have placed all 15 residues onto the crystal structure of HLA-DRB3*03:01, which has a nearly identical secondary structure to the now-established MHC class II (MHC II) fold (33) (Fig. 3C). Yet, the great majority of the residues identified with HLA-DRB345 linkage to susceptibility to or protection from T1D project from the β 1 β 2 domain surfaces. According to one hypothesis, these surfaces form the homodimerization surface of the so-called ($\alpha\beta$)₂ homodimer of $\alpha\beta$ heterodimers, loaded with identical antigenic peptides and, consequently, the two neighboring surfaces onto which may fit two accessory CD4 molecules from the T cell bearing the cognate T-cell receptor (TCR) (33). Only two of the first 10 residues that belong to the α 1 β 1 domain (β 70 and β 81) point into the groove and participate in the formation of pockets 4/7 and 1, while the rest are either pointing directly away from the groove, being on the β -sheet floor, or to the side of the groove. It turns out, nonetheless, that, of the remaining thirteen residues, one is β 25, also seen in linkages to DRB1 alleles. β 104 is at the presumed interphase of the two $\alpha\beta$ heterodimers forming the ($\alpha\beta$)₂ homodimer, while the other remaining 11 residues are all located on the face of the β 1 β 2 surface that is supposed to interact productively with the accessory molecule CD4 (Fig. 3C). Of these residues, β 135 is within the β 134–148 binding site of the CD4 molecule, while most of the others are on the surface that presumably interacts with CD4 (33,34). Residue β 51 is located near the antigenic peptide C-terminus opening of the peptide-binding groove right next to β 52Glu, a residue considered pivotal with β 55Arg, in the formation of cognate TCR-mediated homodimerization of HLA-DR $\alpha\beta$ heterodimers in the so-called β 49–55 homodimerization patch (33,35).

Haplotype-Based Additive Associations and Interactions Between HLA-DRB1 and -DRB345

Local LD between two genes in proximity dictated that DRB1 alleles shared haplotypes with none or one of the DRB345 genes (Table 3). Twenty-seven haplotypes were formed between, respectively, 22 and 6 alleles of DRB1 and DRB345. Full allelic associations of individual genes were captured by five risk haplotypes and nine protective haplotypes. To gain an insight into the potential modifying effects of DRB345 on DRB1 effects, haplotype sets were sorted according to DRB1 sequences, and sorted haplotype sets (4, 5), (6, 7), (16, 17, 18), and (26, 27) were colored in purple or red. Consistent disease associations of haplotypes between two haplotypes suggest that DRB345 may have limited modifying effects in these pairs. In Supplementary Table 3, the same 27 haplotypes and their association statistics are listed by sorting DRB345 first. Examination of DRB1 sequence variations across groups of DRB345 sequences suggested that DRB1 probably drove the haplotype associations with T1D.

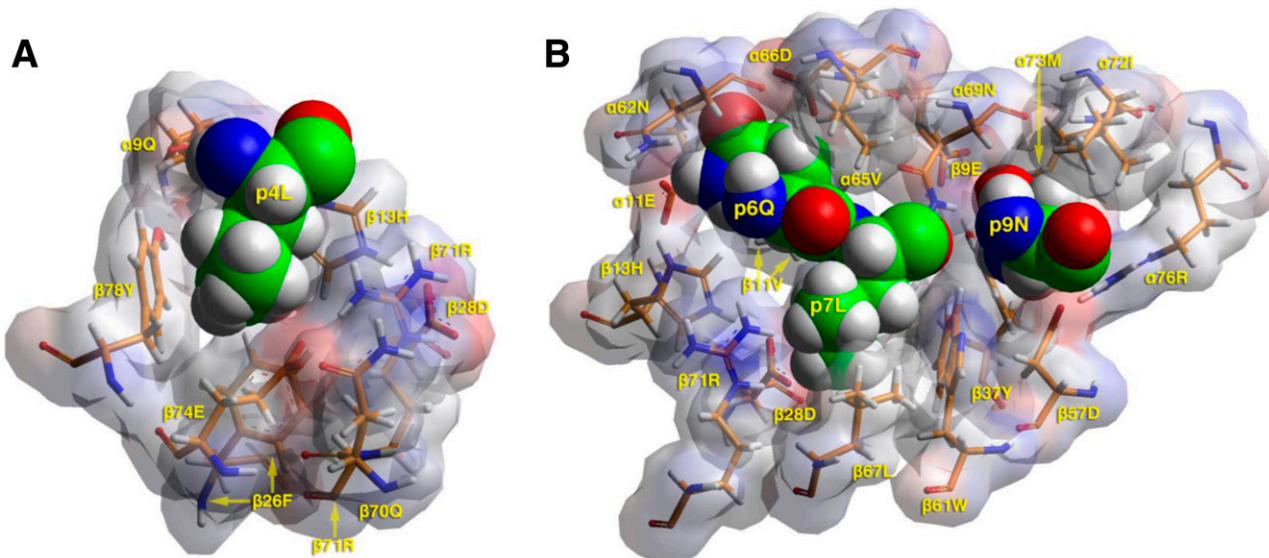


Figure 2—Structure modeling and simulation on HLA-DRB1*04:03, an allele conferring resistance to T1D in several populations. **A:** TCR view of pocket 4 with p4Leu from peptide human insulin A8–19 after energy minimization as detailed in RESEARCH DESIGN AND METHODS. The images from the conventional MHC II depiction are rotated by 10° around the y-axis (right-hand side into the plane of the paper, left-hand side out of the plane of the paper) for better clarity. The orientation of several residues constituting this pocket (e.g., β28Asp, β70Gln, β71Arg, β74Glu) is different from one image to the other because of different residues at p5–p7, which interact with residues constituting this pocket (not shown). Colors and conventions are as in Fig. 1D. In addition, the anchor residue from the antigenic peptide (p4Leu in this case) is shown in space-filling mode with identical atomic color conventions except that carbon is in green. The depicted MHC II residues also have transparent surfaces colored according to atomic charge (blue, positive; red, negative; shades of gray, in-between values). **B:** TCR view of pockets 6, 7, and 9 of the antigenic peptide human insulin A8–19 (Thr-Ser-Ile-Cys-Ser-Leu-Tyr-Gln-Leu-Glu-Asn-Tyr, with boldface indicating anchors) in the groove of DRB1*04:03 (DR*04:03) after energy minimization as detailed in RESEARCH DESIGN AND METHODS. The p6Gln fits snugly into pocket 6, and p7Leu sits well in pocket 7, while p9Asn floats at the opening of pocket 9. Because of the presence of p10Tyr (not shown; there is a terminal peptide bond between p9 and p10 instead of a negatively charged carboxylate at p9), the interaction of α76Arg-β57Asp is uninterrupted. Molecular colors and conventions as in A.

Specific Residues of DR Associate With Islet Autoantibodies Among Patients

Elevated titers of islet autoantibodies against various autoantigens were long suspected to associate with polymorphisms of DR genes. Table 4 lists 20 HLA-DR haplotypes and their associations with IAA, GADA, IA-2A, and ZnT8RA, ZnT8WA, and ZnT8QA among all 962 patients. Evidently, DR haplotypes had strong associations with alterations of IA-2A (minimum $P = 5.28 * 10^{-17}$) and of GADA (minimum $P = 6.58 * 10^{-5}$), while their associations with IAA or with ZnT8A were relatively weaker.

Four Residues of DRB1 Are Responsible for IA-2A Positivity

By applying the ROR procedure to DRB1 association with six islet autoantibodies, we established that associations of DRB1 with IAA, GADA, ZnT8RA, or ZnT8WA were not significant by the threshold $P < 0.01$ (Supplementary Fig. 1A). DRB1 was found to be strongly associated with IA-2A positivity among patients, with four detected residues at positions β14, β25, β71, and β73 with the successive log likelihood $P < 10^{-10}$ (Supplementary Fig. 1A and Table 5). The respective amino acid sequence (ERKA) was found to

Table 2—A total of six alleles, formed by 15 amino acids, resulted from merging association-equivalent and highly similar alleles

ID	DRB345	Control	Patient	OR	HS	P-value	Allelic Group	18	25	40	41	44	48	51	70	76	81	104	135	180	181	187
1	*****	214(16.82)	260(13.51)	0.80	-2.60	9.20E-03	DRBX*null	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
2	FRFDVRRQDHAGVTE	206(16.19)	149(7.74)	0.48	-7.17	7.60E-13	DRB3*02:02~DRB3*02:24	F	R	F	D	V	R	R	Q	D	H	A	G	V	T	E
3	FRFDVRTDDHASVTE	212(16.67)	13(0.68)	0.04	-16.52	2.65E-61	DRB5*01:01~DRB5*01:02	F	R	F	D	V	R	T	D	D	H	A	S	V	T	E
4	FRFDVRTQDHAGVTE	281(22.09)	523(27.18)	1.23	3.37	7.58E-04	DRB3*01:01~DRB3*03:01~DRB5*02:02	F	R	F	D	V	R	T	Q	D	H	A	G	V	T	E
5	LWYNLQTRDYSGMMQ	317(24.92)	949(49.32)	1.98	14.74	3.71E-49	DRB4*01:03	L	W	Y	N	L	Q	T	R	D	Y	S	G	M	M	Q
6	LWYNLQTRDYSSMMQ	38(2.99)	30(1.56)	0.52	-2.58	9.94E-03	DRB4*01:01	L	W	Y	N	L	Q	T	R	D	Y	S	S	M	M	Q

The allelic frequencies are n (%) in 636 control subjects and 936 patients along with their estimated ORs, haplotype scores (HSs), and P values. These amino acid positions in the Protein Data Bank are listed above the individual amino acids.

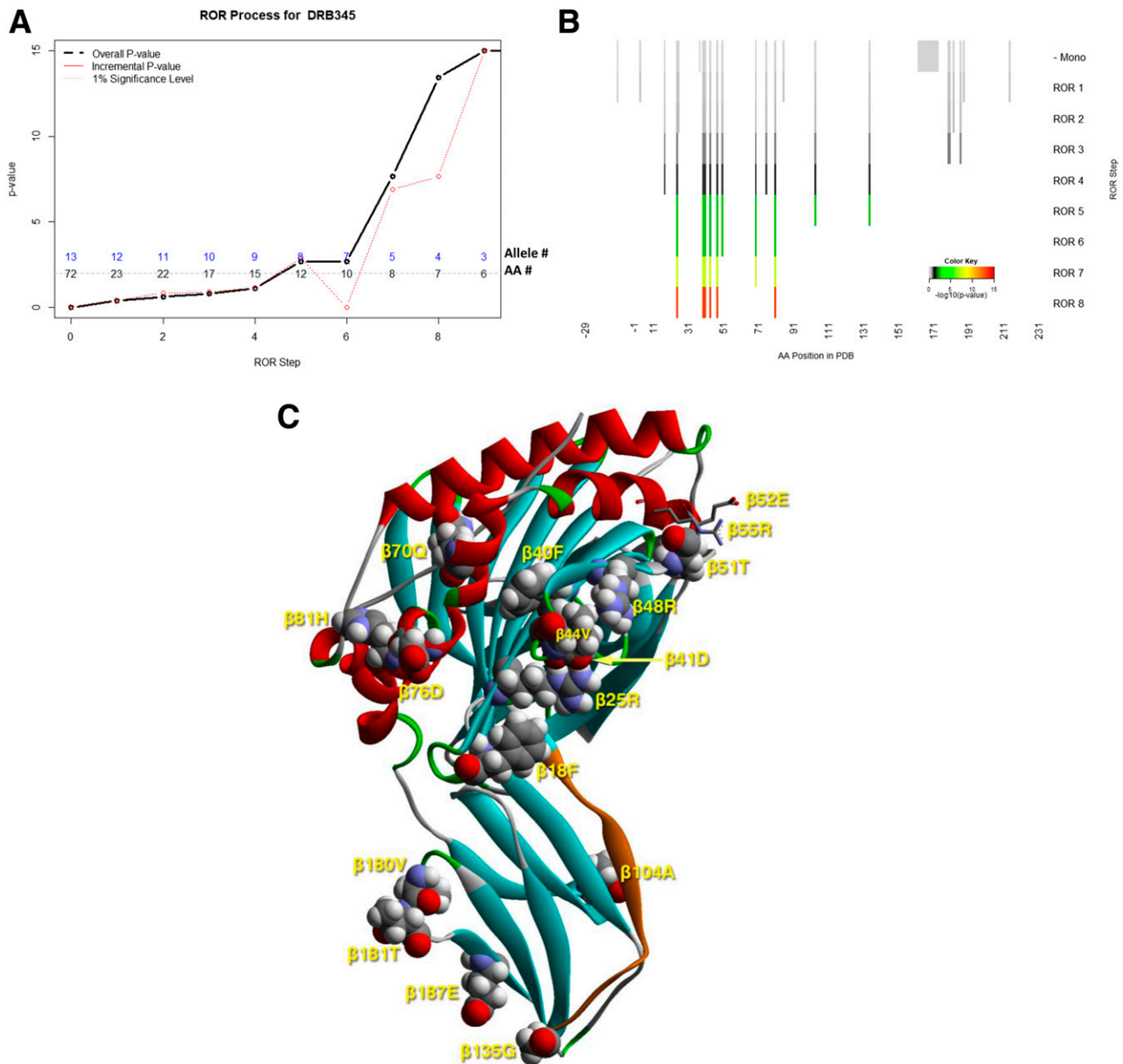


Figure 3—Results from analyzing T1D association with HLA-DRB345. **A:** Results from analyzing DRB345 association with T1D. The logarithmic P values (black line) were computed from the likelihood χ^2 test over successive ROR steps, at each of which the numbers of amino acids were gradually deleted and the number of unique alleles reduced; comparable P values for incremental change (red line) were also computed. **B:** Results from analyzing DRB345 association with T1D, visually displaying how specific AAs were deleted throughout the ROR procedure. Color intensity corresponds to the logarithmic P values. **C:** The crystal structure of the heterodimer HLA-DRA/HLA-DRB3:03:01 (3c5j.pdb) shown in a manner presumed to participate in the formation of a homodimer of two HLA-DR $\alpha\beta$ heterodimers, as first recorded in the crystal structure of HLA-DR1 and subsequently several other DR alleles (33,35,40). The second HLA-DR $\alpha\beta$ heterodimer in the homodimer complex is presumed to be rotated by $\sim 180^\circ$ around a near-vertical axis so that its two regions $\beta 50$ – 55 make contact through salt bridges between $\beta 52$ Glu and $\beta 55$ Arg (33). This is the same depiction mode as in Fig. 1D, with the exception of the secondary structure of the MHC II molecule being shown in solid, instead of line, ribbons. The 15 AAs identified by ROR are shown in space-filling form with the same atomic color conventions as in Fig. 1C. Notice that all but one are on the same side of the surface of the molecule as the CD4-binding surface. The $\beta 134$ – 148 stretch that binds to CD4 is shown in orange. The two key residues presumed to participate in the homodimer formation, $\beta 52$ Glu and $\beta 55$ Arg, are shown in stick form, without their hydrogen atoms for clarity. It is of interest that the two DRB4 alleles linked to T1D protection and susceptibility in this study (Table 2) differ only in position $\beta 135$ (Ser for the protective allele DRB4*01:01:01:01 [OR 0.5] and Gly for the susceptible allele DRB4*01:03:01 [OR 2.0]). PDB, Protein Data Bank.

be highly significant and positively associated with IA-2A positivity (OR 2.15, $P = 8.83 \times 10^{-17}$). Meanwhile, two other amino acid sequences (EREA and ERKG) associated with protection from developing IA-2A (OR 0.70 and 0.59,

$P = 0.0461$ and 7.84×10^{-11} , respectively). The motif (ERKA) captures three DRB1 alleles (DRB1*04:01, *04:13, and *13:03). In contrast, the motif ERKG captures two alleles (DRB1*03:01 and *03:05), while EREA captures

Table 3—A total of 27 haplotypes are formed, respectively, by 11 and 15 amino acids of DRB1 and DRB345

ID	DRB1-DRB345	Control (%)	Patient (%)	OR	HS	P-value
1	ERFDYAAAYVG FRFDVRTDDHASVTE	8(0.63)	3(0.16)	0.25	-2.24	2.52E-02
2	ERFDYAAAYVV FRFDVRTDDHASVTE	203(15.96)	10(0.52)	0.03	-16.41	1.59E-60
3	ERFDYEAAYVG FRFDVRTQDHAGVTE	55(4.32)	86(4.47)	1.03	0.20	8.44E-01
4	ERFDYEAAYVV FRFDVRRQDHAGVTE	44(3.48)	13(0.69)	0.20	-5.88	4.04E-09
5	ERFDYEAAYVV FRFDVRTQDHAGVTE	58(4.54)	28(1.44)	0.32	-5.55	2.80E-08
6	ERFDYKGRYVV FRFDVRRQDHAGVTE	21(1.66)	96(4.98)	3.00	4.62	3.89E-06
7	ERFDYKGRYVV FRFDVRTQDHAGVTE	132(10.37)	388(20.18)	1.95	7.52	5.36E-14
8	ERFDYRAAYVG FRFDVRRQDHAGVTE	64(5.01)	19(0.99)	0.20	-7.04	1.93E-12
9	ERFDYRAAYVV FRFDVRRQDHAGVTE	19(1.49)	2(0.1)	0.07	-4.56	5.10E-06
10	ERFVSRAYAV FRFDVRRQDHAGVTE	29(2.27)	14(0.73)	0.32	-3.66	2.52E-04
11	ERYAHRAEYVV FRFDVRRQDHAGVTE	28(2.2)	4(0.2)	0.09	-5.40	6.63E-08
12	ERYDYEAYVG *****	11(0.86)	2(0.1)	0.12	-3.31	9.18E-04
13	ERYDYEAYVV LWYNLQTRDYSGMMQ	7(0.55)	27(1.4)	2.55	2.25	2.47E-02
14	ERYDYKAAAYVG LWYNLQTRDYSGMMQ	123(9.66)	690(35.86)	3.71	16.97	1.44E-64
15	ERYDYRAAYAV *****	13(1.02)	11(0.57)	0.56	-1.45	1.47E-01
16	ERYDYRAAYVG *****	125(9.82)	168(8.73)	0.89	-1.04	3.00E-01
17	ERYDYRAAYVG FRFDVRTQDHAGVTE	20(1.57)	14(0.73)	0.46	-2.29	2.20E-02
18	ERYDYRAAYVG LWYNLQTRDYSGMMQ	7(0.56)	8(0.42)	0.74	-0.61	5.39E-01
19	ERYDYRAAYVV LWYNLQTRDYSGMMQ	60(4.72)	143(7.43)	1.58	3.04	2.38E-03
20	ERYDYRAEYVG LWYNLQTRDYSGMMQ	12(0.94)	2(0.1)	0.11	-3.53	4.23E-04
21	ERYDYRAEYVV LWYNLQTRDYSGMMQ	12(0.94)	2(0.1)	0.11	-3.53	4.23E-04
22	ERYSYKAAAYVG FRFDVRTQDHAGVTE	10(0.79)	5(0.26)	0.33	-2.14	3.27E-02
23	ERYSYRAAYVG LWYNLQTRDYSGMMQ	4(0.31)	25(1.3)	4.13	2.93	3.43E-03
24	ERYSYRALYVG *****	59(4.64)	73(3.79)	0.82	-1.18	2.38E-01
25	ERYVSRAYVVG LWYNLQTRDYSGMMQ	12(0.94)	28(1.46)	1.54	1.30	1.93E-01
26	KQYVSRGQVVG LWYNLQTRDYSGMMQ	77(6.06)	24(1.25)	0.21	-7.52	5.35E-14
27	KQYVSRGQVVG LWYNLQTRDYSMMQ	37(2.9)	22(1.14)	0.39	-3.53	4.14E-04

The haplotypic frequencies are *n* (%) in 636 control subjects and 936 patients along with their estimated ORs, haplotype scores (HSs), and *P* values. The total number of haplotype counts is twice the number of subjects.

six alleles (DRB1*01:03, *04:02, *11:02, *11:03, *13:01, *13:02).

DRB1*ERKA Is More Associated With IA-2A-Positive Onset, and DRB1*ERKG Is More Associated With IA-2A-Negative Onset

DRB1*ERKA was found to be positively associated with IA-2A positivity (OR 2.15, *P* = 8.83 * 10⁻¹⁸), while DRB1*ERKGY was negatively associated with IA-2A positivity (OR 0.59,

P = 7.84 * 10⁻¹¹) (Table 5). When contrasting with population-based controls, both DRB1*ERKAY and *ERKGY were associated with IA-2A-positive and -negative patients, except that their association magnitudes were reversed (OR 3.77 vs. 1.75 for ERKAY and 1.85 vs. 3.14 for ERKGY) (Table 6). DRB1*ERKGY uniquely corresponded with DRB1*03:01:01, while DRB1*ERKAY corresponded with DRB1*04:01:01 and *13:03:01. In Table 1, ERKA is highlighted in light blue. It is also clear that sequences of these

Table 4—Estimated association statistics (ORs, z scores, and P values) on assessing genetic associations with positivity of each islet autoantibody at the time of diagnosing T1D among 962 patients

DRB1-DRB345	IAA			GADA			IA2A			ZnT8RA			ZnT8WA			ZnT8QA			
	OR	Z	P-value	OR	Z	P-value	OR	Z	P-value	OR	Z	P-value	OR	Z	P-value	OR	Z	P-value	
1	ERFDYAAAYVV-FRFDVRTDDHASVTE						0.97	-0.04	9.71E-01	0.77	-0.41	6.84E-01	0.62	-0.76	4.50E-01				
2	ERFDYEAAYVG-FRFDVRTQDHAGVTE	0.87	-0.59	5.54E-01	0.58	-2.61	9.18E-03	0.63	-1.98	4.82E-02	1.89	2.76	5.87E-03	1.73	2.48	1.36E-02	1.67	2.42	1.53E-02
3	ERFDYEAAYVV-FRFDVRRQDHAGVTE	1.15	0.24	8.11E-01	0.85	-0.29	7.75E-01	0.45	-1.49	1.37E-01	0.78	-0.48	6.32E-01	0.50	-1.26	2.06E-01	1.10	0.15	8.84E-01
4	ERFDYEAAYVV-FRFDVRTQDHAGVTE	1.08	0.20	8.42E-01	0.79	-0.64	5.21E-01	1.02	0.03	9.72E-01	0.96	-0.10	9.20E-01	0.75	-0.79	4.32E-01	0.86	-0.35	7.28E-01
5	ERFDYKGRYVV-FRFDVRRQDHAGVTE	0.80	-0.92	3.59E-01	1.94	2.79	5.22E-03	0.34	-5.19	2.08E-07	0.69	-1.72	8.54E-02	0.60	-2.23	2.54E-02	0.69	-1.55	1.21E-01
6	ERFDYKGRYVV-FRFDVRTQDHAGVTE	0.97	-0.33	7.38E-01	1.46	3.99	6.57E-05	0.69	-3.83	1.26E-04	0.92	-0.96	3.36E-01	0.98	-0.33	7.41E-01	0.85	-1.77	7.60E-02
7	ERFDYRAAYVG-FRFDVRRQDHAGVTE	0.98	-0.04	9.67E-01	1.09	0.17	8.61E-01	2.06	1.00	3.17E-01	0.56	-1.26	2.06E-01	1.59	0.99	3.20E-01	0.72	-0.64	5.22E-01
8	ERFVSRAYAV-FRFDVRRQDHAGVTE	0.85	-0.28	7.81E-01	1.58	0.79	4.30E-01	1.46	0.50	6.17E-01	2.84	1.69	9.11E-02	0.70	-0.68	4.97E-01	2.69	1.92	5.48E-02
9	ERYDYEAYVV-LWYNLQTRDYSGMMQ	0.89	-0.26	7.95E-01	0.74	-0.59	5.54E-01	1.07	0.13	8.96E-01	0.97	-0.08	9.36E-01	0.93	-0.38	7.05E-01	1.26	0.42	6.76E-01
10	ERYDYKAAAYVG-LWYNLQTRDYSGMMQ	1.12	1.91	5.65E-02	0.82	-3.58	3.50E-04	2.12	8.38	5.28E-17	1.01	0.24	8.14E-01	1.14	2.30	2.13E-02	1.12	1.90	5.72E-02
11	ERYDYRAAYAV *****				0.76	-0.46	6.47E-01	1.09	0.11	9.09E-01	1.36	0.49	6.24E-01	0.53	-1.04	3.01E-01	1.68	0.87	3.84E-01
12	ERYDYRAAYVG *****	0.69	-2.28	2.27E-02	0.79	-1.68	9.25E-02	0.92	-0.46	6.49E-01	1.26	1.57	1.17E-01	0.95	-0.36	7.21E-01	0.95	-0.30	7.63E-01
13	ERYDYRAAYVG-FRFDVRTQDHAGVTE	1.18	0.30	7.65E-01	1.03	0.24	8.13E-01	0.32	-2.22	2.67E-02	0.18	-2.65	7.97E-03	0.00	-3.37	7.40E-04			
14	ERYDYRAAYVG-LWYNLQTRDYSGMMQ	2.12	1.09	2.74E-01															
15	ERYDYRAAYVV-LWYNLQTRDYSGMMQ	0.85	-0.87	3.82E-01	1.38	1.83	6.80E-02	0.72	-1.73	8.38E-02	1.10	0.59	5.57E-01	1.11	0.65	5.16E-01	0.96	-0.26	7.96E-01
16	ERYSYRAAYVG-LWYNLQTRDYSGMMQ	0.83	-0.44	6.63E-01	0.50	-1.79	7.33E-02	2.79	1.47	1.40E-01	0.44	-2.08	3.77E-02	0.44	-2.01	4.39E-02	0.18	-2.71	6.78E-03
17	ERYSYRALYVG *****	1.40	1.47	1.42E-01	0.58	-2.42	1.54E-02	0.80	-0.84	4.01E-01	0.71	-1.50	1.32E-01	0.77	-1.19	2.36E-01	0.76	-1.09	2.77E-01
18	ERYVSRAYVVG-LWYNLQTRDYSGMMQ	2.45	2.48	1.31E-02	1.41	0.73	4.65E-01	1.46	0.71	4.77E-01	1.20	0.47	6.36E-01	0.75	-0.58	5.59E-01	0.91	-0.12	9.08E-01
19	KQYVSRGQVVG-LWYNLQTRDYSGMMQ	1.27	0.58	5.60E-01	1.27	0.56	5.75E-01	1.70	0.88	3.80E-01	0.78	-0.63	5.30E-01	1.31	0.64	5.19E-01	1.45	0.90	3.69E-01
20	KQYVSRGQVVG-LWYNLQTRDYSMMQ	0.99	-0.02	9.84E-01	2.84	2.00	4.52E-02	1.09	0.17	8.67E-01	1.12	0.26	7.97E-01	0.64	-1.05	2.95E-01	1.15	0.32	7.47E-01

Table 5—Results from applying ROR to assess DRB1 with islet autoantibody IA-2A

AA	Neg (%) ¹	Pos (%) ¹	OR ²	Z ²	P-value ²	Equivalent Alleles
ERAA	2(0.53)	11(0.71)	1.34	0.38	7.03E-01	DRB1*15:01~*15:02~*15:03
EREA	40(10.64)	116(7.49)	0.70	-1.99	4.61E-02	DRB1*01:03~*04:02~*11:02~*11:03~*13:01~*13:02
ERKA	71(18.88)	628(40.57)	2.15	8.59	8.83E-18	DRB1*04:01~*04:13~*13:03
ERKG	142(37.77)	344(22.22)	0.59	-6.50	7.84E-11	DRB1*03:01~*03:05 DRB1*01:01~*01:02~*04:03~*04:04~*04:05~*04:06~*04:07~*04:08~*04:10 ~*08:01~*08:02~*08:03~*08:04~*09:01~*10:01~*11:01~*11:04~*11:84 ~*12:01~*12:02~*13:05~*14:01~*14:02~*14:04~*14:12~*14:54~*16:01 ~*16:02~*16:09
ERRA	114(30.32)	410(26.49)	0.87	-1.57	1.16E-01	
KQRG	7(1.86)	39(2.52)	1.35	0.76	0.44836	DRB1*07:01

The statistical significance at the 1% level is demonstrated for HLA-DRB1 with IA-2A, with four amino acids at β14, β25, β71, and β73. ¹Frequency counts of [neg]ative and [pos]itive IA-2A islet autoantibody and their percentages (%). ²OR, z score statistic, and P value computed under the null hypothesis.

five DRB1 residues caught major genetic associations with T1D risks, indicating that IA-2A elevation is an important seroconversion event leading to the disease onset.

DRB1*ERKA and *ERKG Genotypes Have Additive Associations With Type 1 Diabetes Risks

Homozygous carrier of DRB1*ERKA appeared to have a much stronger association than its allelic association (OR 8.71, P = 5.50 * 10⁻¹³), indicating an additive association with IA-2A positivity (Supplementary Table 4). On the other hand, the homozygous DRB1*ERKG had an additive association with patients with T1D whose IA-2A association tended to be with T1D with a normal IA-2A.

DISCUSSION

Recursively organizing complex amino acid sequences reveals that sequences of 11 residues capture all risk/protective associations of DRB1 full sequence alleles and, likewise, so do sequences of 15 DRB345 residues. All 11 residues of DRB1 are located either under (β14 and β25 and interacting with each other as well as neighboring amino acids) or within the peptide-binding groove (β47, β57, β60, β71, β73, β74, β78, β85, and β86), suggesting that specific patterns of these residues recognize specific peptides from autoantigens as the functional mechanism.

This list includes β71, which is one of two residues discovered in a previous study (14). On the other hand, residues β18, β25, β40, β41, β44, β48, β51, β70, β76, and β81 of DRB345 are all in the α1β1 domain, with β70 participating in the formation of pockets 4 and 7 and β81 as part of pocket 1 and hydrogen bond acceptor from the antigenic peptide backbone (33,35). The remaining residues are either under or to the side of the antigen-binding groove and all to the side that constitutes the presumed CD4 binding surface (33–36) (Fig. 3C). Likewise, residues β135, β180, β181, and β187 are in the α2β2 domain and are also in the visible side of Fig. 3C. Remarkably, β135 is part of the β134–148 stretch directly involved in CD4 binding (33,36), although one work a long time ago (albeit with no follow-up since) showed the polymorphisms in β179–191 of DRB345 to be responsible for poor CD4 binding and consequent activation (37). Only two amino acids, β51 and β104, could conceivably influence the presumed TCR-induced HLA-RA(αβ)₂ homodimerization: The first is actually within the HLA-DR homodimerization patch, next to near invariant β52Glu that participates in this process by salt bridges to β55Arg from the other homodimerized HLA-DR molecule (35,36); the other faces into the position of the other HLA-DRαβ–presumed heterodimer (33,35). The homodimerization and subsequent CD4 binding hypothesis have been contested, and evidence has been produced by several laboratories that is either

Table 6—Characterizing associations of four selected amino acids in DRB1 with IA-2A–positive and 1A-2A–negative patients and their differences

DRB1	IA2A-positive					IA2A-negative				Difference		Equivalent Alleles
	Control (%)	Patient (%)	OR	Z	P-value	Positive	OR	Z	P-value	Z	P-value	
ERAA	213(16.75)	11(0.71)	0.04	-15.09	1.97E-51	2(0.53)	0.03	-8.07	6.85E-16	0.38	7.03E-01	DRB1*15:01~*15:02~*15:03
EREA	176(13.84)	116(7.49)	0.54	-5.59	2.27E-08	40(10.64)	0.77	-1.65	9.87E-02	-1.99	4.61E-02	DRB1*01:03~*04:02~*11:02~*11:03~*13:01~*13:02
ERKA	137(10.77)	628(40.57)	3.77	18.06	6.40E-73	71(18.88)	1.75	4.16	3.12E-05	8.59	8.83E-18	DRB1*04:01~*04:13~*13:03
ERKG	153(12.03)	344(22.22)	1.85	7.21	5.67E-13	142(37.77)	3.14	11.27	1.82E-29	-6.50	7.84E-11	DRB1*03:01~*03:05 DRB1*01:01~*01:02~*04:03~*04:04~*04:05~*04:06 ~*04:07~*04:08~*04:10~*08:01~*08:02~*08:03 ~*08:04~*09:01~*10:01~*11:01~*11:04~*11:84 ~*12:01~*12:02~*13:05~*14:01~*14:02~*14:04 ~*14:12~*14:54~*16:01~*16:02~*16:09
ERRA	479(37.66)	410(26.49)	0.70	-6.40	1.58E-10	114(30.32)	0.81	-2.64	8.35E-03	-1.57	1.16E-01	
KQRG	114(8.96)	39(2.52)	0.28	-7.42	1.16E-13	7(1.86)	0.21	-4.57	4.77E-06	0.76	4.48E-01	DRB1*07:01

consistent with or against this hypothesis (33,36–43). One matter complicating all analyses is the considerable *in vitro* instability of the human CD4 molecule, necessitating in many cases the use of mutant forms that form stable complexes with HLA-DR molecules to obtain crystallization of CD4-HLA-DR complexes (41,42). It is certain, however, that F(ab')₂-fragment antibody-mediated homodimerization of HLA-DP or -DQ or -DR on monocytes leads to distinct downstream activation patterns (44).

Further examination of the role of given amino acids shows five residues of DRB1 (β 14, β 25, β 71, β 73, and β 78) to have strong associations with elevated IA-2A titers for patients with T1D at the time of diagnosis. In particular, the sequence ERKA, found in DRB1*04:01 and *13:03, associates with IA-2A-positive patients with T1D, while ERKG, uniquely found in DRB1*03:01, associates more with IA-2A-negative patients with T1D. Despite the difference of only one amino acid, A versus G at position β 73, this difference indicates a potential effect on peptide binding or subsequent cognate TCR binding mechanisms as well as possible effector outcomes (thymocyte deletion or T-helper or T-regulatory cell selection) with respect to IA-2A (45). Finally, effects of specific amino acid sequences ERKA or ERKG appear to be additive. A recent study underscored the importance of HLA-DRB4*01:01 allele for T1D pathogenesis over HLA-DR4 alleles in LD with the former (46).

Our list of 11 residues in DRB1 points to these residues' involvement in four different pockets. Both β 85 and β 86 are within pocket 1, with the Gly/Val dimorphism of the latter position determining the aromatic/aliphatic crucial p1 anchor for HLA-DR molecules (30,35). Furthermore, β 71 is part of both pockets 4 and 7, while β 74 and β 78 are within pocket 4, and β 47 is within pocket 7. Finally, β 57 is within pocket 9, with Asp/non-Asp occupancy at this position determining the type of anchor residue at this pocket (mostly small aliphatic vs. acidic, respectively) (30,35). This residue has also been shown to contribute immensely to the stability of the peptide-HLA-DR complex (40,47).

Also of interest are DRB1 residues β 14, β 25, β 60, and β 73, which do not locate in any one of known pockets. The given residues are either under (β 14 and β 25) or adjoining (β 60 and β 73) the groove or within it and as parts of pockets (β 47, β 57, β 71, β 74, β 78, β 85, and β 86). To our knowledge, the variations in positions 14, 25, 60, and 73 have never been tested in any alleles with regard to peptide binding, cognate TCR recognition, or accessory CD4 binding, and one could well speculate on what effect such substitutions might have (e.g., β 73Ala→Gly would make the MHC II molecule slightly more flexible and peptide binding weaker or stronger depending on the anchor residue at adjacent pocket 4; it could, of course, also affect TCR recognition because of the extra flexibility involved [45]). Because there may certainly be interactions of pMHC II complexes with other molecules in the course of immune tolerance

formation and immune response, one cannot easily speculate any further.

It has been shown that T1D-susceptible and -resistant HLA-DR alleles bind to different peptide epitopes within the same protein autoantigen and with different binding affinities and, even more importantly, with slower off-kinetics for the resistant alleles (28,31). This work also demonstrates how tight the binding can be for the resistant alleles, such as DRB1*04:03, because of just a few key-positioned amino acids, among them β 74Glu. The sum of knowledge generated regarding T1D pathogenesis through particular HLA-DR alleles binding to specific autoantigenic epitopes could lead to the design of antigen-specific interventions (perhaps coupled with nonspecific ones) either in the pre- or in the new-onset T1D phase. The targets would probably have to include effector as well as regulatory CD4/CD8 T cells (48–52).

One by-product from the current analysis is that many uncommon alleles of HLA genes are now naturally grouped with common alleles on the basis of ROR-selected amino acids. Following the selection of 11 residues by the ROR procedure, we have now merged several common alleles with several uncommon alleles, netting 22 common amino acid sequences and 4 uncommon ones (data not shown). Grouping common alleles with similar disease associations achieves analytic efficiency. More importantly, grouping uncommon alleles into common ones allows us to assess genetic risks for those individual carriers of uncommon alleles, risks for which would otherwise be unknown.

Another practical benefit of reducing complex HLA gene sequences and related nomenclature to a representation with a minimum set of residues is its stable representation, despite ever-evolving sequencing technologies and increasing numbers of uncommon alleles. When a minimum set of amino acids, referring to tagging residues (patterning after tagging single nucleotide polymorphisms), can be identified for each disease-associated HLA gene, it would be stable, regardless of how many novel nucleotide sequences and related nucleotide/amino acid sequences are added with better genotyping technologies or with more diverse populations.

In comparison with the early discovery of β 13 and β 71, our list includes one but not the other (14). The consistently discovered amino acid β 71, which is expected, has one of four possible residues (Lys, Arg, Ala, and Glu), with Lys being the most common, and is indicative of DRB1*03:01 and *04:01, two of the most common risk alleles (Table 1). Also, because of their high allelic frequencies, imputations of HLA genotypes from single nucleotide polymorphisms tend to be accurate. On the other hand, it is somewhat surprising that β 13, also a member of pocket 4, was not picked up by ROR. The selection process of ROR (Fig. 1B) indicates that β 13 is deselected quite early on, probably because of limited association or limited polymorphisms in the current HLA sequence data. Interestingly, however, there are two residues, β 14 and β 25, both underneath β 13, where the electrostatic interactions

between the first two, as well as among them and neighboring residues, add to the stability of the MHC II-peptide complex (Fig. 1C). Given the nature of conditional analysis (14), it is possible that the former analysis may not find $\beta 14$ or $\beta 25$, possibly because of local LD. However, if there was strong local LD between $\beta 13$ and $\beta 14/\beta 25$ and both were significantly associated with T1D, ROR would retain $\beta 13$ because its analysis is not conditional. One possible explanation is that imputations may introduce extra variations in calling uncommon HLA alleles tagged by $\beta 13$. Another possible scenario is that this occurs as a result of limited population variation, as our study population is restricted only to the Swedish population. Of course, such conjectures would not be addressable without obtaining HLA genotypes at the nucleotide sequence resolution, more so when we acquire knowledge of specific HLA-DR-peptide combinations being responsible for susceptibility or resistance to T1D.

Besides the difference in HLA genotyping technology, another important difference is the analytic method (i.e., the conditional analysis vs. ROR). While both share the same analytic goal of achieving a parsimony with respect to amino acids, ROR recursively eliminates amino acids if they are monomorphic or their deletion does not alter the association profile. ROR avoids extraction of any information from LD between residues in the selection process. In contrast, the conditional analysis uses the stepwise procedure to search for significantly associated amino acids with T1D while conditioning on those selected amino acids. It uses both disease associations with residues and LD associations between amino acids to achieve the parsimony. Consequently, the conditional analysis can achieve a greater reduction in the number of selected residues, but selected residues should be viewed as best biological surrogate markers that encapsulate both disease association and LD associations. Indeed, it is well suited for developing risk prediction models when parsimonious models achieve greater statistical stability in predictions.

Before concluding, it is important to recognize three limitations of this study. First, our analysis focuses on T1D association with HLA-DR genes without including HLA-DQ genes. If a particular HLA-DR allelic association with T1D is mediated by LD between DR and DQ genes, and hence is not causal, the corresponding motif of amino acids should be treated as a surrogate marker for T1D. For the same reason, discovered amino acids may not include causal residues. The remedy is to investigate HLA-DQ genes and their amino acids jointly with HLA-DR genes, which will be pursued and reported in a separate study. Second, the current investigation is restricted to genetic variants in protein-coding exons, ignoring genetic variants in noncoding regions. Given the available high-resolution HLA genotypes, we are positioned to investigate regulatory roles of noncoding genetic variants, results from which will be reported elsewhere. Finally, the Swedish population is relatively homogeneous, and hence, HLA-DR diversity is

limited. Thus, identified residues and related motifs are inevitably limited. To be comprehensive, it is important to conduct such studies in other ethnic populations or, even better, to carry out an international consortium study with multiple ethnic populations. Indeed, our discovery should be thought of as the first stepping stone toward fully understanding T1D associations with class II genes.

Acknowledgments. The authors would like to thank Dr. Gun Forsander (Department of Pediatrics, the Queen Silvia Children's Hospital, Sahlgrenska University Hospital, Gothenburg, Sweden) and Dr. Sten A. Ivarsson (Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden) for clinical expertise in the conduct of the BDD study. Equally, the authors thank Anita Ramelius, Linda Faxius, Karl Moritz, Petra Moritz, Ingrid Wigheden, Ida Jönsson, and Rasmus Håkansson (Department of Clinical Sciences, Lund University/Clinical Research Centre, Malmö, Sweden) for expert technical assistance.

Funding. The study was supported by National Institute of Diabetes and Digestive and Kidney Diseases grants R56-DK-117276, R01-DK-117276, DK-63861, and DK-26190 and by a grant from the European Foundation for the Study of Diabetes Clinical Research Grants Programme 2013. The study also was supported in part by the Swedish Child Diabetes Foundation (Barndiabetesfonden); a Swedish Research Council Linné grant to Lund University Diabetes Centre; the Skåne County Council for Research and Development; the Swedish Association of Local Authorities and Regions; and Forskningsrådet om Hälsa, Arbetsliv och Välfärd. The Silicon Graphics Fuel instrument and the accompanying software were obtained through a grant from the Epirus Regional Development Program to the Technological Educational Institute of Epirus (MIS 91949) through the 3rd Community Support Framework of the European Union (80% European Union funds, 20% Hellenic state funds).

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. L.P.Z., G.K.P., W.W.K., D.E.G., and Å.L. researched and analyzed the data and wrote the manuscript. B.X., M.K., R.W., C.-W.P., and W.C.N. contributed to the NGTS, researched data, and reviewed the manuscript. A.K.M. and G.P.B. carried out molecular simulations and graphical representations of select HLA-DR molecules. A.C., H.E.-L., J.L., C.M., M.P., and U.S. designed the BDD study, researched data, contributed to the discussion, and reviewed the manuscript. Å.L. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Data and Resource Availability. Interested researchers are encouraged to contact Å.L. about accessing the underlying data set, following institutional review of human subject research and institutional policies.

References

- Rich SS, Concannon P, Erlich H, et al. The Type 1 Diabetes Genetics Consortium. *Ann N Y Acad Sci* 2006;1079:1–8
- Cooper JD, Howson JM, Smyth D, et al.; Type 1 Diabetes Genetics Consortium. Confirmation of novel type 1 diabetes risk loci in families. *Diabetologia* 2012;55:996–1000
- Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013;8:e64683
- Concannon P, Rich SS, Nepom GT. Genetics of type 1A diabetes. *N Engl J Med* 2009;360:1646–1654
- Concannon P, Erlich HA, Julier C, et al.; Type 1 Diabetes Genetics Consortium. Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes* 2005;54:2995–3001
- Barrett JC, Clayton DG, Concannon P, et al.; Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009;41:703–707

7. Thomson G, Valdes AM, Noble JA, et al. Relative predispositional effects of HLA class II DRB1-DQB1 haplotypes and genotypes on type 1 diabetes: a meta-analysis. *Tissue Antigens* 2007;70:110–127
8. Kockum I, Sanjeevi CB, Eastman S, Landin-Olsson M, Dahlquist G, Lernmark A. Complex interaction between HLA DR and DQ in conferring risk for childhood type 1 diabetes. *Eur J Immunogenet* 1999;26: 361–372
9. Graham J, Kockum I, Sanjeevi CB, et al. Negative association between type 1 diabetes and HLA DQB1*0602-DQA1*0102 is attenuated with age at onset. Swedish Childhood Diabetes Study Group. *Eur J Immunogenet* 1999; 26: 117–127
10. Erlich H, Valdes AM, Noble J, et al.; Type 1 Diabetes Genetics Consortium. HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. *Diabetes* 2008; 57:1084–1092
11. Valdes AM, Thomson G, Erlich HA, Noble JA. Association between type 1 diabetes age of onset and HLA among sibling pairs. *Diabetes* 1999;48:1658–1661
12. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;30:1205–1213
13. Roark CL, Anderson KM, Simon LJ, Schuyler RP, Aubrey MT, Freed BM. Multiple HLA epitopes contribute to type 1 diabetes susceptibility. *Diabetes* 2014; 63:323–331
14. Hu X, Deutsch AJ, Lenz TL, et al. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nat Genet* 2015;47:898–905
15. van Heemst J, Huizinga TJ, van der Woude D, Toes RE. Fine-mapping the human leukocyte antigen locus in rheumatoid arthritis and other rheumatic diseases: identifying causal amino acid variants? *Curr Opin Rheumatol* 2015;27: 256–261
16. Zhang XC, Li SS, Wang H, Hansen JA, Zhao LP. Empirical evaluations of analytical issues arising from predicting HLA alleles using multiple SNPs. *BMC Genet* 2011;12:39
17. Zhao LP, Alshiekh S, Zhao M, et al.; Better Diabetes Diagnosis (BDD) Study Group. Next-generation sequencing reveals that HLA-DRB3, -DRB4, and -DRB5 may be associated with islet autoantibodies and risk for childhood type 1 diabetes. *Diabetes* 2016;65:710–718
18. Nelson WC, Pyo CW, Vogan D, et al. An integrated genotyping approach for HLA and other complex genetic systems. *Hum Immunol* 2015;76:928–938
19. Smith AG, Pyo CW, Nelson W, et al. Next generation sequencing to determine HLA class II genotypes in a cohort of hematopoietic cell transplant patients and donors. *Hum Immunol* 2014;75:1040–1046
20. Zhao LP, Huang X. Recursive organizer (ROR): an analytic framework for sequence-based association analysis. *Hum Genet* 2013;132:745–759
21. Carlsson A, Kockum I, Lindblad B, et al.; Swedish Better Diabetes Diagnosis Study Group. Low risk HLA-DQ and increased body mass index in newly diagnosed type 1 diabetes children in the Better Diabetes Diagnosis study in Sweden. *Int J Obes* 2012;36:718–724
22. Delli AJ, Vaziri-Sani F, Lindblad B, et al.; Better Diabetes Diagnosis Study Group. Zinc transporter 8 autoantibodies and their association with SLC30A8 and HLA-DQ genes differ between immigrant and Swedish patients with newly diagnosed type 1 diabetes in the Better Diabetes Diagnosis study. *Diabetes* 2012;61: 2556–2564
23. Andersson C, Larsson K, Vaziri-Sani F, et al. The three ZNT8 autoantibody variants together improve the diagnostic sensitivity of childhood and adolescent type 1 diabetes. *Autoimmunity* 2011;44:394–405
24. Delli AJ, Lindblad B, Carlsson A, et al.; Better Diabetes Diagnosis (BDD) Study Group. Type 1 diabetes patients born to immigrants to Sweden increase their native diabetes risk and differ from Swedish patients in HLA types and islet autoantibodies. *Pediatr Diabetes* 2010;11:513–520
25. Gyllenberg A, Asad S, Piehl F, et al.; Swedish Childhood Diabetes Study Group; Diabetes Incidence in Sweden Study Group; Better Diabetes Diagnosis Study group. Age-dependent variation of genotypes in MHC II transactivator gene (CIITA) in controls and association to type 1 diabetes. *Genes Immun* 2012;13:632–640
26. Vaziri-Sani F, Delli AJ, Elding-Larsson H, et al. A novel triple mix radiobinding assay for the three ZnT8 (ZnT8-RWQ) autoantibody variants in children with newly diagnosed diabetes. *J Immunol Methods* 2011;371:25–37
27. Reichstetter S, Papadopoulos GK, Moustakas AK, et al. Mutational analysis of critical residues determining antigen presentation and activation of HLA-DQ0602 restricted T-cell clones. *Hum Immunol* 2002;63:185–193
28. Ge X, James EA, Reijonen H, Kwok WW. Differences in self-peptide binding between T1D-related susceptible and protective DR4 subtypes. *J Autoimmun* 2011;36:155–160
29. Scholz E, Mestre-Ferrer A, Daura X, et al. A comparative analysis of the peptide repertoires of HLA-DR molecules differentially associated with rheumatoid arthritis. *Arthritis Rheumatol* 2016;68:2412–2421
30. Friede T, Gnau V, Jung G, Keilholz W, Stevanović S, Rammensee HG. Natural ligand motifs of closely related HLA-DR4 molecules predict features of rheumatoid arthritis associated peptides. *Biochim Biophys Acta* 1996;1316:85–101
31. Davenport MP, Godkin A, Friede T, et al. A distinctive peptide binding motif for HLA-DRB1*0407, an HLA-DR4 subtype not associated with rheumatoid arthritis. *Immunogenetics* 1997;45:229–232
32. Masiewicz SA, Papadopoulos GK, Swanson E, Moriarity L, Moustakas AK, Nepom GT. Modulation of T cell response to hGAD65 peptide epitopes. *Tissue Antigens* 2002;59:101–112
33. Brown JH, Jardetzky TS, Gorga JC, et al. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature* 1993;364:33–39
34. Cammarota G, Scheirle A, Tacacs B, et al. Identification of a CD4 binding site on the beta 2 domain of HLA-DR molecules. *Nature* 1992;356:799–801
35. Stern LJ, Brown JH, Jardetzky, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 1994;368:215–221
36. König R, Shen XL, Germain RN. Involvement of both major histocompatibility complex class II alpha chains and beta chains in CD4 function indicates a role for ordered oligomerization in T cell activation. *J Exp Med* 1995;182:779–787
37. Fleury S, Thibodeau J, Croteau G, et al. HLA-DR polymorphism affects the interaction with CD4. *J Exp Med.* 1995;182:733–741
38. Jönsson P, Southcombe JH, Santos AM, et al. Remarkably low affinity of CD4/peptide-major histocompatibility complex class II protein interactions. *Proc Natl Acad Sci U S A* 2016;113:5682–5687
39. Cochran JR, Cameron TO, Stern LJ. The relationship of MHC-peptide binding and T cell activation probed using chemically defined MHC class II oligomers. *Immunity* 2000;12:241–250
40. Bondinas GP, Moustakas AK, Papadopoulos GK. The spectrum of HLA-DQ and HLA-DR alleles, 2006: a listing correlating sequence and structure with function. *Immunogenetics* 2007;59:539–553
41. Reinherz EL, Wang JH. Codification of bidentate pMHC interaction with TCR and its co-receptor. *Trends Immunol* 2015;36:300–306
42. Yin Y, Wang XX, Mariuzza RA. Crystal structure of a complete ternary complex of T-cell receptor, peptide-MHC, and CD4. *Proc Natl Acad Sci U S A* 2012;109: 5405–5410
43. Brameshuber M, Kellner F, Rosboth BK, et al. Monomeric TCRs drive T cell antigen recognition. *Nature Immunology* 2018;19:487–496
44. Matsuoka T, Tabata H, Matsushita S. Monocytes are differentially activated through HLA-DR, -DQ, and -DP molecules via mitogen-activated protein kinases. *J Immunol* 2001;166:2202–2208
45. Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol* 2015;33:169–200
46. James EA, Gillette L, Durinovic-Bello I, et al. DRB4*01:01 has a distinct motif and presents a proinsulin epitope that is recognized in subjects with type 1 diabetes. *J Immunol* 2018;201:3524–3533
47. Verreck FA, Vermeulen C, Poel AV, et al. The generation of SDS-stable HLA DR dimers is independent of efficient peptide binding. *Int Immunol* 1996;8:397–404

48. Moustakas AK, Papadopoulos GK. Use of MHC II structural features in the design of vaccines for organ-specific autoimmune diseases. *Curr Pharm Des* 2009;15:3262–3273
49. Atkinson MA, Roep BO, Posgai A, Wheeler DCS, Peakman M. The challenge of modulating β -cell autoimmunity in type 1 diabetes. *Lancet Diabetes Endocrinol* 2019;7:52–64
50. Sebastiani G, Ventriglia G, Stabilini A, et al. Regulatory T-cells from pancreatic lymphnodes of patients with type-1 diabetes express increased levels of microRNA miR-125a-5p that limits CCR2 expression. *Sci Rep* 2017;7: 6897
51. Schwedhelm K, Thorpe J, Murray SA, et al. Attenuated IL-2R signaling in CD4 memory T cells of T1D subjects is intrinsic and dependent on activation state. *Clin Immunol* 2017;181:67–74
52. Trotta E, Besette PH, Silveria SL, et al. A human anti-IL-2 antibody that potentiates regulatory T cells by a structure-based mechanism. *Nat Med* 2018;24: 1005–1014