



Motifs of Three HLA-DQ Amino Acid Residues (α 44, β 57, β 135) Capture Full Association With the Risk of Type 1 Diabetes in DQ2 and DQ8 Children

Lue Ping Zhao,¹ George K. Papadopoulos,^{2,§} William W. Kwok,³ Antonis K. Moustakas,⁴ George P. Bondinas,² Helena Elding Larsson,⁵ Johnny Ludvigsson,⁶ Claude Marcus,⁷ Ulf Samuelsson,⁶ Ruihan Wang,⁸ Chul-Woo Pyo,⁸ Wyatt C. Nelson,⁸ Daniel E. Geraghty,⁸ and Åke Lernmark⁵

Diabetes 2020;69:1573–1587 | <https://doi.org/10.2337/db20-0075>

HLA-DQA1 and -DQB1 are strongly associated with type 1 diabetes (T1D), and DQ8.1 and DQ2.5 are major risk haplotypes. Next-generation targeted sequencing of HLA-DQA1 and -DQB1 in Swedish newly diagnosed 1- to 18 year-old patients ($n = 962$) and control subjects ($n = 636$) was used to construct abbreviated DQ haplotypes, converted into amino acid (AA) residues, and assessed for their associations with T1D. A hierarchically organized haplotype (HOH) association analysis allowed 45 unique DQ haplotypes to be categorized into seven clusters. The DQ8/9 cluster included two DQ8.1 risk and the DQ9 resistant haplotypes, and the DQ2 cluster included the DQ2.5 risk and DQ2.2 resistant haplotypes. Within each cluster, HOH found residues α 44Q (odds ratio [OR] 3.29, $P = 2.38 \times 10^{-85}$) and β 57A (OR 3.44, $P = 3.80 \times 10^{-84}$) to be associated with T1D in the DQ8/9 cluster representing all ten residues (α 22, α 23, α 44, α 49, α 51, α 53, α 54, α 73, α 184, β 57) due to complete linkage disequilibrium (LD) of α 44 with eight such residues. Within the DQ2 cluster and due to LD, HOH analysis found α 44C and β 135D to share the risk for T1D (OR 2.10, $P = 1.96 \times 10^{-20}$). The motif “QAD” of α 44, β 57, and β 135 captured the T1D risk association of DQ8.1 (OR 3.44, $P = 3.80 \times 10^{-84}$), and the corresponding

motif “CAD” captured the risk association of DQ2.5 (OR 2.10, $P = 1.96 \times 10^{-20}$). Two risk associations were related to GAD65 autoantibody (GADA) and IA-2 autoantibody (IA-2A) but in opposite directions. CAD was positively associated with GADA (OR 1.56, $P = 6.35 \times 10^{-8}$) but negatively with IA-2A (OR 0.59, $P = 6.55 \times 10^{-11}$). QAD was negatively associated with GADA (OR 0.88; $P = 3.70 \times 10^{-3}$) but positively with IA-2A (OR 1.64; $P = 2.40 \times 10^{-14}$), despite a single difference at α 44. The residues are found in and around anchor pockets 1 and 9, as potential T-cell receptor contacts, in the areas for CD4 binding and putative homodimer formation. The identification of three HLA-DQ AAs (α 44, β 57, β 135) conferring T1D risk should sharpen functional and translational studies.

Type 1 diabetes (T1D) is one of the most common chronic diseases among children and young adults, and the disease incidence is steadily rising worldwide, posing challenges to affected families and increasing financial stress on health care systems (1–4). The lifetime risk is >1% in both North America and Europe (5,6). Decades of research have shown

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

³Benaroya Research Institute at Virginia Mason, Seattle, WA

⁴Department of Food Science and Technology, Faculty of Environmental Sciences, Ionian University, Argostoli, Cephalonia, Greece

⁵Department of Clinical Sciences, Lund University CRC, Skåne University Hospital, Malmö, Sweden

⁶Crown Princess Victoria Children's Hospital, Region Östergötland, and Division of Pediatrics, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

⁷Division of Pediatrics, Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

⁸Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA

Corresponding authors: Lue Ping Zhao, lzhao@fredhutch.org, George K. Papadopoulos, papadogg@gmail.com, and Åke Lernmark, ake.lernmark@med.lu.se

Received 22 January 2020 and accepted 30 March 2020

This article contains supplementary material online at <https://doi.org/10.2337/db20-4567/suppl.12053640>.

[§]G.K.P. has been retired from Technological Educational Institute (TEI) of Epirus, Arta, Greece since 1 September 2018. The affiliation is given for identification purposes only. As of 1 October 2018, the TEI of Epirus has been absorbed by the University of Ioannina. The respective department is now called Department of Agriculture.

© 2020 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 <https://www.diabetesjournals.org/content/license>.

that T1D is an autoimmune disease, in which the host immune system wrongfully attacks pancreatic islet β -cells (7). The initial immune cell activation is reflected in the generation of autoantibodies against specific autoantigens present in the β -cells. These autoantibody biomarkers appear first to either insulin (IAA) or GAD65 (GADA), representing two different endotypes (8–11). Approximately 60% of children with a first autoantibody within 1 year developed a second autoantibody against insulin, GAD65, IA-2 (IA-2A), or zinc transporter 8 (ZnT8A). Genetic studies, from classical HLA studies and genome-wide association studies (12–14), have confirmed that the strongest genetic associations with T1D are with the class II HLA genes on chromosome 6, dwarfing ~ 50 non-HLA genetic associations (12,15–17). While several recent studies have attempted to distinguish between two adjacent MHC class II loci (HLA-DR and -DQ) (18–21), even extending to the HLA-DP locus (22), we have recently identified 11 residues in HLA-DRB1 and 15 residues in HLA-DRB3, -DRB4, and -DRB5 that may include critical residues by applying the recursive organizer (ROR) (23).

Early T1D investigations showed that HLA-DQ may have greater roles in disease susceptibility than HLA-DR genes (24,25). HLA-DQ includes two genes (DQA1 and DQB1), located, respectively, at (32,637,357–32,643,652) and (32,659,467–32,666,684) on human genome hg38 [UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly]. Due to their proximity (only 15,815 bases apart), they are in high linkage disequilibrium (LD). Functionally, proteins coded by *DQA1* and *DQB1*, respectively, form a heterodimer, as a typical MHC II molecule, with the DQA1 and DQB1 polypeptides. Their biosynthesis in-vivo occurs in concert with that of the invariant chain (26), which associates with the MHC II $\alpha\beta$ heterodimer having the CLIP peptide portion bound into the groove of this heterodimer (27).

Mature DQ α 1 and DQ β 1 protein molecules consist of, respectively, 232 and up to 237 (because of polymorphism in the length of the intracytoplasmic tail) amino acids (AAs) (for the intricacies of the HLA-DQA and -DQB AA numbering system see Research Design and Methods). Note that AA and residue are used interchangeably hereafter. Within the predominantly Caucasoid population, many residues are monomorphic, along with some other residues that are polymorphic but not identified with any specific function. The critical residues for T1D risk are expected to exhibit strong associations with T1D. Further, critical residues are in high LD among themselves and with other polymorphic residues but without defined function. Further complicating their identification, critical residues become functional only in the context of a particular “DQ α and β heterodimer structure” specified by a set of residues in the molecule, i.e., statistically, representing a form of high-order interaction.

Considering these DQ gene complexities, the goal here is to investigate whether state of the art analytic approaches should make it possible to identify polymorphic residues

that significantly associate with T1D susceptibility by using sequence-based HLA genotypes. One recent approach is the conditional analysis, i.e., systematically searching through all residues for most significantly associated residues and then conditioning on selected residues and searching for the next significantly associated residues (28). After its application to T1D-associated residues, the single residue β 57 was found to explain $>15\%$ of class II association. However, this finding should be interpreted cautiously because the local LD could condition out “other critical residues” (29). To ease this concern, we recently applied an ROR to investigate HLA-DRB genes with T1D risk and have been able to find multiple residues and their motifs that capture all known DR associations (23). Further scrutiny of the ROR approach, when considering both high LD and high-order interactions, has suggested that ROR is capable of finding multiple associated residues in high LD but may miss others in the presence of high-order interactions. Consequently, residues identified by ROR include critical residues but may still leave some critical variants undiscovered.

In this study, the aim is to identify potential T1D critical residues in HLA-DQA1 and -DQB1 that are mediated through their interactions with multiple residues corresponding to protein structures. To conquer this aged challenge, we revise the ROR approach to a hybrid approach. Clusters of AA sequences that potentially are indicative of different protein structures were first identified. This was followed by searching for specific T1D AAs in a haplotype association analysis using their specific AAs. This approach is hereafter referred to as a hierarchically organized haplotype (HOH) association analysis. Briefly, HOH carries out phylogenetic analysis of AA sequences of unique DQA1-DQB1 haplotypes and hierarchically organizes all DQ haplotypes; i.e., the shorter the distance between two DQ haplotypes, the closer they will be placed (see the flowchart in Supplementary Fig. 1 for details). Two clusters of DQ haplotypes emerge: one cluster includes two high-risk DQ8.1 haplotypes, in addition to a resistant DQ9 haplotype, and the other cluster includes a high-risk DQ2.5 haplotype. The first cluster was referred to thus as DQ8/9 to indicate the inclusion of both DQ8 and DQ9 haplotypes, while the second cluster was referred to as the DQ2 cluster. Within each cluster, HOH carries out T1D association analysis with individual residues within their respective cluster, leading to the identification of residues associated with T1D. The association analysis is facilitated through a haplotype association analysis with known phases. Among 962 patients and 636 control subjects collected in the Better Diabetes Diagnosis (BDD) study, we have genotyped for HLA-DQA1 and -DQB1 by the next-generation targeted sequencing technology and have also constructed their haplotypes through their LD as basic analytic units. After filtering out uninformative AAs, we have hierarchically organized all observed 45 unique DQ haplotypes, with two clusters that include three T1D-associated DQ haplotypes,

DQ2.5 and DQ8.1 within the DQ2 and DQ8/9 clusters. Additionally, both clusters also include resistant haplotypes, despite highly similar AA sequences, and offer an opportunity to discover specific T1D-associated residues within clusters. Further, among all T1D patients, we investigate how some of these critical residues may associate with GADA, IAA, IA-2A, and the three variants (residue R, W, or Q on position 325) of ZnT8A (30–32) at the time of T1D diagnosis.

RESEARCH DESIGN AND METHODS

Much of the materials and methods used for this study has previously been described (23) and is included as Supplementary Material for readers' convenience. We thus include sections specific to this study.

AA Sequences of DQA1 and DQB1 Alleles

As NGTS sequenced DNA nucleotides of selected exons and used sequences to infer DQA1 and DQB1 alleles, we determined HLA-DQA1 and -DQB1 genotypes at the high resolution of six digits or higher, based on the HLA nomenclature of IMGT, July 2019 release (https://raw.githubusercontent.com/ANHIG/IMGTHLA/Latest/alignments/DQA1_prot.txt and https://raw.githubusercontent.com/ANHIG/IMGTHLA/Latest/alignments/DQB1_prot.txt).

Further, we determined physical positions of individual AAs in the α -chain from codon $\alpha 1$ to $\alpha 232$ and in the β -chain from codon $\beta 1$ to $\beta 237$. We have adopted the numbering system first suggested in 1998 (33) and improved in 2007 (34) that essentially allows for structural equivalence among residues of various MHC II alleles, regardless of gene locus or species, and based on the structure of the first published MHC II allele, HLA-DR1 (35,36). There were also 23 and 32 residues in the signaling peptides of DQA1 and DQB1, respectively, and their positions were labeled as position (–23) and (–32) to position (–1). By convention, the first AA of the mature functional chain is numbered +1.

Molecular Depiction of HLA-DQ Structures

All depictions of determined structures of various HLA-DQ alleles have been carried out using WebLabViewer 3.5 and the DS ViewerPro 6.0 and three-dimensional molecular rendering software Accelrys (BIOVIA, San Diego, CA), based on coordinates freely available in the Protein Data Bank (www.rcsb.org). The respective codes are as follows: 1jk8.pdb for the HLA-DQ8–insulin B11–23/24Gly complex (37) and 6mfg.pdb for the HLA-DQ2–gliadin- $\alpha 1$ complex (38).

Data and Resource Availability

The data set generated in the current study is not publicly available due to Swedish law on protecting human subjects but is available from Å.L. upon application and approval by BDD Steering Committee.

RESULTS

Construction of DQ Haplotypes

The Swedish population is relatively homogeneous, except for ~7% immigrants identified through origins of parents or grandparents, and individual ethnicity is privacy protected by law (31). Relying on the high LD between HLA-DQA1 and -DQB1 within the Swedish population, we constructed their individual pairs of DQ haplotypes. A total of 1,524 DQ haplotype pairs from 636 control subjects and 962 patients were constructed with 100% posterior probabilities because they were doubly homozygous, semi-homozygous, or doubly heterozygous, with alternative DQ haplotypes being absent in this study population. The remaining 4.6% of the subjects (25 control subjects and 49 patients) had their constructed DQ haplotypes inferred with Bayesian posterior probabilities <100% posterior probability. All of them were double heterozygotes; thirty-three subjects had genotypes DQA1*03:01/*05:05" and "DQB1*03:02/*03:01, and their posterior probabilities associated with DQA1*03:01-B1*03:02/A1*05:05-B1*03:01 or DQA1*03:01-B1*03:01/A1*05:05-B1*03:02 were 0.99985 or 0.00015. For simplicity, A1 and B1 are omitted hereafter. By highly differentiated posterior probabilities between two possible diplotypes, the constructed pair of haplotypes was taken as DQ*03:01-*03:02/*05:05-*03:01. Similar haplotype/diplotype inferences were made for 26 pairs, DQ*03:01-*03:02/*03:03-*03:01, with the posterior probability of 0.99562; 9 pairs, DQ*03:01-*03:02/*03:02-*03:03, with 0.99949; 3 pairs, DQ*03:02-*03:02/*05:05-*03:01, with 0.98767; 1 pair, DQ*03:01-*03:02/*03:02-*03:04, with 0.99467; 1 pair, DQ*03:02-*03:03/*04:01-*04:02, with 0.99985; and 1 pair, DQ*05:01-*02:01/*05:05-*03:02, with 0.99959. Thus, all constructed DQ haplotype pairs were of high quality with limited ambiguity.

HOH Clusters Correspond to Known Serotypes

Our earlier report showed in detail that multiple haplotypes/diplotypes of HLA-DQ alleles had strong resistance or risk associations with the risk of T1D, largely consistent with known results published to date (39–41). Briefly, using the HLA nomenclature, the genetic association analysis showed that two DQ8.1 haplotypes (h20, DQA1*03:01-DQB1*03:02, and h26, DQ*03:02-*03:02) and one DQ2.5 haplotype (h35, DQ*05:01-*02:01) had significant risk associations with T1D (corresponding *P* values shown in Table 1). Additionally, there were 9 resistant haplotypes (h2, h6, h7, h11–13, h17, h18, and h39) (corresponding *P* values shown in Table 1) and 10 neutral haplotypes (h1, h4, h5, h8, h9, h27, h28, h30, h32, and h33). Further, there were 18 rare haplotypes with three or fewer observed members in the current study (h3, h10, h14–16, h21–24, h29, h31, h34, 38, and h40–44). Finally, there were three haplotypes with more than 4 observed copies, but two (h19 and h36) were observed among control subjects and two (h25 and h45) only among patients. In total, there were 1272 and 1924 observed haplotypes

Table 1—A total of 45 unique DQA1-DQB1 haplotypes are observed among control subjects and patients, and their association results are shown

ID	DQA1-DQB1	Serotype	Control 1,272 (%)	Patient 1,924 (%)	OR	z	P
h1	*01:01-*05:01	DQ5.1	146 (11.48)	180 (9.36)	0.82	-1.94	5.18E-02
h2	*01:01-*05:03		18 (1.42)	3 (0.16)	0.11	-4.33	1.51E-05
h3	*01:02-*05:01		3 (0.24)				
h4	*01:02-*05:02	DQ5.2	16 (1.26)	15 (0.78)	0.62	-1.36	1.75E-01
h5	*01:02-*05:04	DQ5.4	3 (0.24)	2 (0.1)	0.44	-0.92	3.55E-01
h6	*01:02-*06:02	DQ6.2	201 (15.8)	8 (0.42)	0.03	-16.39	2.25E-60
h7	*01:02-*06:03	DQ6.3	5 (0.39)	1 (0.05)	0.13	-2.18	2.91E-02
h8	*01:02-*06:04	DQ6.4	52 (4.09)	80 (4.16)	1.02	0.10	9.22E-01
h9	*01:02-*06:09	DQ6.9	2 (0.16)	6 (0.31)	1.98	0.86	3.91E-01
h10	*01:02-*06:14		1 (0.08)				
h11	*01:03-*06:01	DQ6.1	9 (0.71)	1 (0.05)	0.07	-3.25	1.14E-03
h12	*01:03-*06:03	DQ6.3	91 (7.15)	39 (2.03)	0.28	-7.28	3.37E-13
h13	*01:04-*05:03		8 (0.63)	1 (0.05)	0.08	-2.73	6.38E-03
h14	*01:05-*05:01		1 (0.08)				
h15	*01:07-*05:03		2 (0.16)	1 (0.05)			
h16	*01:10-*06:03		1 (0.08)				
h17	*02:01-*02:02	DQ2.2	78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09
h18	*02:01-*03:03	DQ9.2	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10
h19	*03:01-*03:01	DQ7.3	7 (0.55)				
h20	*03:01-*03:02	DQ8.1	149 (11.71)	769 (39.97)	3.41	18.13	1.97E-73
h21	*03:01-*03:03		2 (0.16)				
h22	*03:01-*03:04	DQ7.3	1 (0.08)				
h23	*03:01-*03:05		2 (0.16)				
h24	*03:01-*04:02		1 (0.08)				
h25	*03:02-*02:02	DQ2.3		6 (0.31)			
h26	*03:02-*03:02	DQ8.1	10 (0.79)	69 (3.59)	4.56	5.05	4.30E-07
h27	*03:02-*03:03	DQ9.3	11 (0.86)	29 (1.51)	1.74	1.61	1.08E-01
h28	*03:02-*03:04	DQ7.3	3 (0.24)	4 (0.21)	0.88	-0.17	8.68E-01
h29	*03:02-*04:02	DQ4.3	2 (0.16)				
h30	*03:03-*03:01	DQ7.3	56 (4.4)	59 (3.07)	0.70	-1.92	5.51E-02
h31	*04:01-*03:03		1 (0.08)				
h32	*04:01-*04:02	DQ4.4	51 (4.01)	72 (3.74)	0.93	-0.39	7.00E-01
h33	*04:02-*04:02		3 (0.24)	4 (0.21)	0.88	-0.17	8.68E-01
h34	*04:04-*04:02		1 (0.08)	1 (0.05)			
h35	*05:01-*02:01	DQ2.5	153 (12.03)	484 (25.16)	2.09	9.23	2.71E-20
h36	*05:01-*03:02			1 (0.05)			
h37	*05:03-*03:01		7 (0.55)				
h38	*05:05-*02:01			1 (0.05)			
h39	*05:05-*03:01	DQ7.5	126 (9.91)	39 (2.03)	0.20	-9.63	6.24E-22
h40	*05:05-*03:02		1 (0.08)	2 (0.1)			
h41	*05:05-*03:09		1 (0.08)				
h42	*05:05-*03:19		3 (0.24)	1 (0.05)			
h43	*05:09-*03:01		1 (0.08)				
h44	*05:10-*03:01		1 (0.08)				
h45	*06:01-*03:01	DQ7.6	6 (0.47)				

Data are *n* (%) unless otherwise indicated. Association results are shown as haplotypic frequency and percentage (%) among control subjects, those among patients, estimated ORs, *z* scores, and *P* values. The 45 unique haplotypes are observed in 636 control subjects (resulting in 1,272 haplotypes) and 962 patients (1,924 haplotypes). Risk and protective haplotypes are highlighted red and green, respectively, at the significance level of 5%.

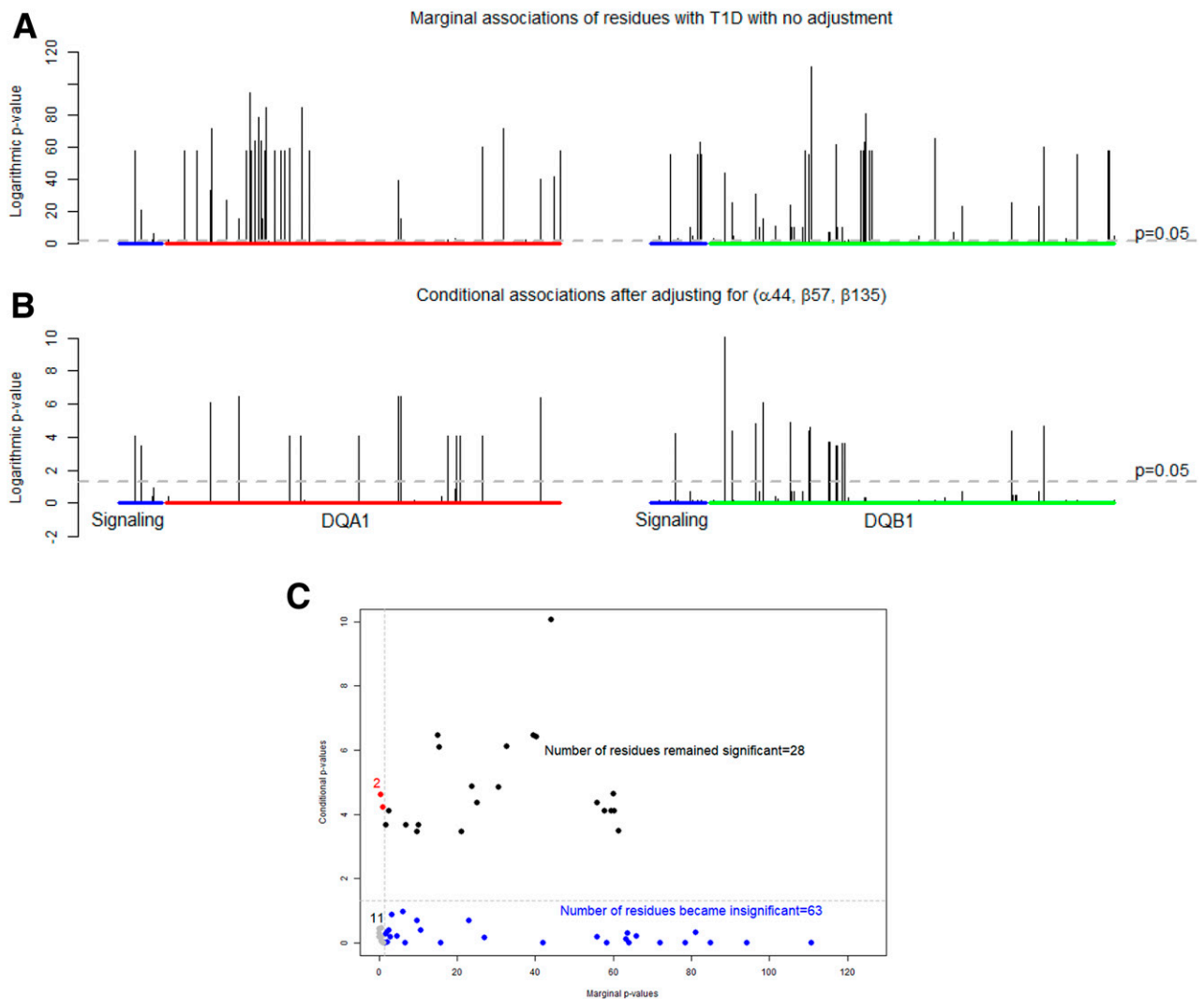


Figure 1—Association scan of individual residues with T1D with the logarithmically transformed P value (y -axis) across all AAs from DQA1 to DQB1. **A:** The scan is from the marginal associations of individual residues without any adjustment (y -axis is scaled from 0 to 100, corresponding to P value of 1 down to 10^{-120}). **B:** The scan is from the adjusted associations for three residues (α 44, β 57, β 135) (y -axis is scaled from 0 to 10, corresponding to P value of 1 down to 10^{-10}). **C:** XY display of computed logarithmic P values from marginal and conditional association analyses. Comparisons of P values from adjusted and unadjusted analyses on all 104 residues.

among control subjects and patients (twice respective sample sizes). Corresponding serotypes are listed in the second column.

By the established DQ library, we converted DQ haplotypes to AA sequences with 524 residues in total (including signal sequences). In this study population, 420 residues were monomorphic among current patients and control subjects and were thus eliminated from further consideration, leaving 104 AAs for further analysis. Directly correlating each residue of the remaining AAs with T1D, we assessed the marginal associations of individual AAs with T1D status and measured such associations via their respective P values (Fig. 1A). Note that the y -axis had a base 10 logarithmic scale, with the 5% threshold shown by the gray dashed line in the bottom of the figure. By the chosen 5% threshold, many P values were much

less than the threshold, approaching 10^{-120} . Sequence motifs of these selected residues (48 on the α -chain and 56 on the β -chain) uniquely corresponded to all 45 DQ haplotypes. While motif and haplotype were equivalent terminologies, the former was used here to refer to an allele specified by selected AAs, while a haplotype referred to an allele specified by DQ alleles in classical nomenclature.

Focusing on the 45 unique DQ haplotypes and their AA sequences, we computed their pairwise distances, i.e., numbers of differing residues and their sums of squared differences. Then, we hierarchically organized them by their distances, placing those haplotype pairs closer together if their distances were shorter. For the results from hierarchical clustering analysis, all 45 unique haplotypes were organized and displayed as a “fan-shaped” phylogenetic tree (Fig. 2). Visually, this phylogenetic tree had two major

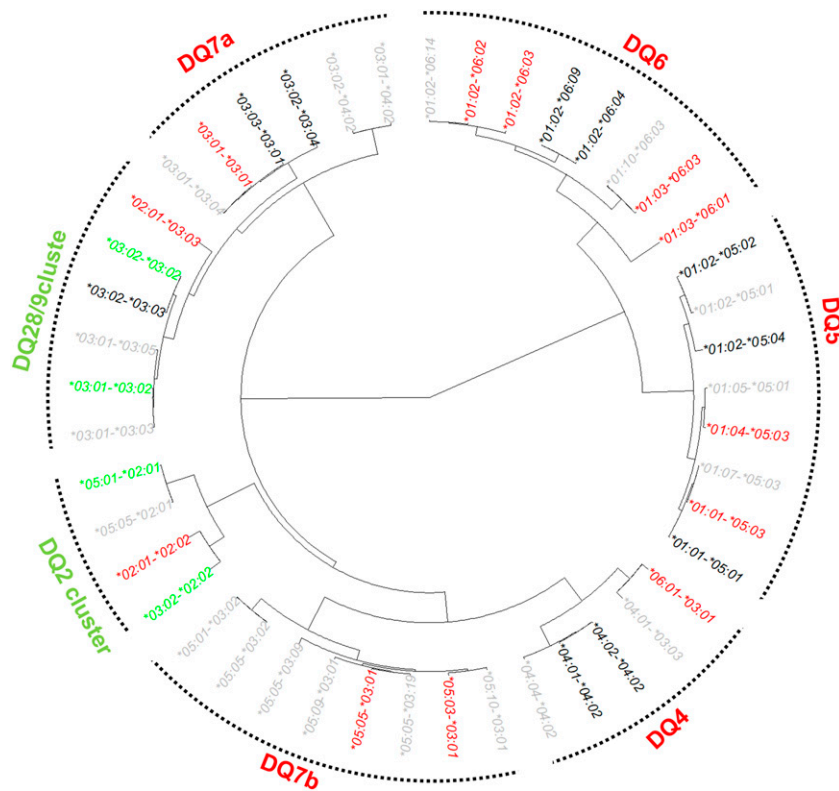


Figure 2—A fan representation of hierarchically organized DQ haplotypes following the phylogenetic analysis of AA sequences of all unique DQ haplotypes. Among seven clusters, DQ2 and DQ8/9 clusters correspond with two DQ haplotype clusters that include both risk (font colored green), resistant (font colored red), and neutral (font colored black) haplotypes, in addition to uncommon DQ haplotypes (with <3 copies of alleles in the study).

branches: the left branch included a subbranch, labeled as the DQ8/9 cluster, which included two risk DQ8.1 haplotypes (font colored green), one resistant haplotype (red font), and one neutral haplotype (black font), in addition to two rare haplotypes (font colored gray). An adjacent cluster, labeled as the DQ2 cluster, in green font includes the risk haplotypes DQ2.5 and DQ*03:02:*02:02 (observed only among 6 patients) and one resistant haplotype (red font), plus a rare haplotype. Additionally, there were five other clusters that included their respective DQ haplotypes, and they included only resistant or neutral DQ haplotypes, aligned with their respective serotypes. The concordance of haplotype clusters with serotypes was not surprising because the phylogenetic tree was constructed by distances of AAs and similarities of AA sequences specified the protein structures and hence serotypes.

Both DQ8/9 and DQ2 clusters included risk, resistant, and neutral haplotypes, and variable associations within each cluster offered a unique opportunity to identify T1D-associated residues in the context of, respectively, DQ8/9 and DQ2 protein structures.

Ten AAs Accounted for Risk/Resistance to T1D Among Carriers of DQ8/9

Most AAs in the haplotype cluster DQ8/9 were monomorphic and thus were not associated with T1D risks, even

though they were part of the specific protein structures (or a profile of AAs from statistical perspective). Eliminating 91 monomorphic AAs led to 13 AAs with polymorphisms (Table 2), in which one AA (β 26) was invariant, except for a variant associated with the rare haplotype, and was excluded from the further analysis. They formed six unique motifs, two of which corresponded to DQ8.1 haplotypes (odds ratio [OR] 3.41 and 4.56, $P = 1.97 \times 10^{-73}$ and 4.30×10^{-7} , respectively) and one resistant haplotype DQ*02:01-*03:03 (DQ9.2, OR 0.09, $P = 1.99 \times 10^{-10}$). Additionally, the cluster included one neutral haplotype (DQ*03:02-*03:03, OR 1.74, $P = 0.108$) and two rare haplotypes (DQ*03:01-*03:03 and *03:01-*03:05).

To formally investigate their empirical associations of individual AAs with T1D risk among carriers of DQ8/9, we performed an AA association analysis, the second step of HOH analysis. Following the initial step of hierarchically organizing all unique haplotypes and focused on the DQ8/9 cluster, HOH created an indicator for each DQ haplotype to be in the DQ8/9 cluster or not. Treating paired indicators as a “pseudo genotype,” HOH carried out a “haplotype association analysis” of the DQ8/9 indicator and each of 12 AAs with T1D risk (β 26 was excluded), with known phases. Association statistics measured T1D associations of individual AAs within the DQ8/9 cluster (Table 3) and estimated frequencies and percentages of specific AAs on

Table 2—Direct comparisons of allelic sequences within mixture 1 and 2 results in identification of 12 and 25 AAs that differentiate between risk and protective alleles in, respectively, DQ8/9 and DQ2 mixtures of haplotypes

A) DQ8/9 mixture	Control	Patient	OR	z	P	α(-6)	α22	α23	α44	α49	α51	α53	α54	α73	α157	α184	β26	β57	
*03:01-*03:03	2 (0.16)					M	Y	S	Q	R	F	R	F	V	A	T	L	D	
*03:01-*03:02	149 (11.71)	769 (39.97)	3.41	18.13	1.97E-73	M	Y	S	Q	R	F	R	F	V	A	T	L	A	
*03:01-*03:05	2 (0.16)					M	Y	S	Q	R	F	R	F	V	A	T	G	A	
*03:02-*03:03	11 (0.86)	29 (1.51)	1.74	1.61	1.08E-01	T	Y	S	Q	R	F	R	F	V	D	T	L	D	
*03:02-*03:02	10 (0.79)	69 (3.59)	4.56	5.05	4.30E-07	T	Y	S	Q	R	F	R	F	V	D	T	L	A	
*02:01-*03:03	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	M	F	T	K	H	L	F	-	L	A	A	L	D	

B) DQ2 mixture	Control	Patient	OR	z	P	α(-13)	α(-6)	α22	α23	α31	α37	α44	α47	α48	α49	α50	α51	α53	α54	α72	α73	α104	α153	α157	α158	α160	α172	α184	α212	β135
*03:02-*02:02		6 (0.31)	>1	2.00	4.60E-02	A	T	Y	S	E	E	Q	L	F	R	R	R	F	F	I	V	T	F	D	D	I	E	T	L	G
*02:01-*02:02	78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09	A	M	F	T	E	E	K	L	F	H	R	L	F	-	I	L	T	F	A	D	I	E	A	L	G
*05:05-*02:01		1 (0.05)				T	M	Y	T	Q	G	C	V	L	R	Q	F	F	-	S	L	I	L	A	E	S	K	A	F	D
*05:01-*02:01	153 (12.03)	484 (25.16)	2.09	9.23	2.71E-20	A	M	Y	T	Q	G	C	V	L	R	Q	F	F	-	S	L	I	L	A	E	S	K	A	F	D

Estimated haplotype frequencies (%) among control subjects, those among patients, estimated ORs, z scores, and P values are listed. Risk and protective haplotypes are highlighted green and red, respectively, at the significance level of 5%. Neutrally associated haplotypes with more than three subjects are in black font, while neutrally associated haplotypes with fewer than three subjects are colored gray. The AA β26 has a variant G only on the rare DQ haplotype and is thus excluded from further analysis. Similarly, the AA α(-13) has a variant T-associated haplotype and is also excluded.

haplotypes of DQ8/9 cluster among control subjects and patients, ORs, z scores, and P values. To assess whether polymorphisms of two or more AAs in the DQ8/9 cluster had comparable T1D associations, we used observed frequencies of AAs among carriers of DQ8/9 and computed the Fisher exact P value for each residue. For example, residue α(-6) in the signal peptide could be either M (methionine) or T (threonine), and both were positively associated with T1D risk (respectively, OR 2.71 and 3.09, P = 2.05 * 10⁻⁵⁸ and 4.63 * 10⁻⁷ [Table 3]). However, their T1D associations were not significantly different between two AAs (Fisher exact P = 0.712). Hence, both α(-6) and α157 were not responsible for variable T1D associations within the DQ8/9 cluster and were thus excluded from further consideration. Similarly, the AA α157 had either A (alanine) or D (aspartic acid), and both had similar risk association patterns as α(-6), due to complete LD in the DQ8/9 cluster. On the other hand, the AA α44 took either K (lysine) or Q (glutamine), which had significantly resistant and risk associations (OR 0.09 and 3.29, P = 1.99 * 10⁻¹⁰ and 2.38 * 10⁻⁸⁵). The difference between these two opposite associations was statistically significant (Fisher exact P = 5.38 * 10⁻²²). Highly similar association patterns were observed for the other eight AAs (α22, α23, α49, α51, α53, α54, α73, α184) because these eight AAs are in complete LD with α44 among carriers of DQ8/9, with minor variations in the general population. Residue β57 is either Ala (A) or Asp (D), which has a risk/resistant relationship with T1D (OR 3.44 and 0.47, P = 3.80 * 10⁻⁸⁴ and 3.93 * 10⁻⁴). The associations of β57A/D with T1D in the DQ8/9 cluster were significantly different (Fisher P = 6.34 * 10⁻¹⁷).

The motifs of residues (α44, β57), coupled with the monomorphic residue β135 (see below for the reason

for this inclusion), represented full polymorphisms of DQ haplotypes within the DQ8/9 cluster. Table 4 shows three unique motifs—KDD (a1), QAD (a2), and QDD (a3)—among carriers of DQ8/9, in which the β135 was monomorphic as expected. In particular, the motif a2 represented both risk haplotypes of DQ8.1 (OR 3.44, P = 3.80 * 10⁻⁸⁴), and the motif a1 corresponded to DQ9.2, a resistant haplotype (OR 0.09, P = 1.99 * 10⁻¹⁰). The motif a3 differed from a2 by a single β57, and the AA D (aspartic acid) at this position was shared with the resistant motif a1. Consequently, the risk association of motif a3 was compromised (OR 1.47, P = 0.235).

Seventeen Residues Account for Risk/Resistance to T1D Among Carriers of DQ2

Across four haplotypes among carriers of DQ2, there were 25 polymorphic AAs (Table 2), one of which α(-13) was associated only with a rare DQ haplotype and was excluded from further analysis. Also, it was noted that 12 AAs (α31, α37, α47, α48, α50, α72, α104, α153, α158, α160, α172, α212) were in complete LD with β135 and thus were represented by β135 in further calculations. Evaluating T1D associations among carriers of DQ2, HOH analysis showed their associations AA by AA (Table 3). The AA α(-6) was either M (methionine) or T (threonine). The methionine variant was positively associated with the T1D risk (OR 1.51, P = 6.54 * 10⁻¹⁰), and threonine was found only among patients. Their associations appeared to be comparable among carriers of DQ2 (Fisher exact P = 0.186). Similar association patterns were observed for AAs at α23, α53, α54, α73, α157, and α184 because they were in complete LD with each other. The AA at α44 took C (cysteine), K (lysine), or Q (glutamine), the first two of which had variable T1D associations with T1D (OR 2.10

Table 3—T1D associations with individual AAs within clusters of, respectively, DQ8/9 and DQ2 haplotypes

ID	Amino Acids	Carriers of DQ8/9						Carriers of DQ2					
		Control	Patient	OR	z	P	Fisher	Control	Patient	OR	z	P	Fisher
1	α(-6) M	189 (14.86)	774 (40.23)	2.71	16.11	2.05E-58	7.12E-01	231 (18.16)	526 (27.34)	1.51	6.18	6.54E-10	1.86E-01
	T	21 (1.65)	98 (5.09)	3.09	5.04	4.63E-07			6 (0.31)	>1			
2	α22 F	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	5.38E-22	78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09	3.11E-18
	Y	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85		153 (12.03)	491 (25.52)	2.12	9.45	3.28E-21	
3	α23 S	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85	5.38E-22		6 (0.31)	>1		1.86E-01	
	T	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10		231 (18.16)	526 (27.34)	1.51	6.18	6.54E-10	
4	α44 C							153 (12.03)	485 (25.21)	2.10	9.26	1.96E-20	3.48E-18
	K	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	5.38E-22	78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09	
	Q	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85			6 (0.31)	>1			
5	α49 H	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	5.38E-22	78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09	3.11E-18
	R	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85		153 (12.03)	491 (25.52)	2.12	9.45	3.28E-21	
6	α51 F	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85	5.38E-22	153 (12.03)	491 (25.52)	2.12	9.45	3.28E-21	3.11E-18
	L	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10		78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09	
7	α53 F	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	5.38E-22	231 (18.16)	526 (27.34)	1.51	6.18	6.54E-10	1.86E-01
	R	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85			6 (0.31)	>1			
8	α54 -	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	5.38E-22	231 (18.16)	526 (27.34)	1.51	6.18	6.54E-10	1.86E-01
	F	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85			6 (0.31)	>1			
9	α73 L	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	5.38E-22	231 (18.16)	526 (27.34)	1.51	6.18	6.54E-10	1.86E-01
	V	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85			6 (0.31)	>1			
10	α157 A	189 (14.86)	774 (40.23)	2.71	16.11	2.05E-58	7.12E-01	231 (18.16)	526 (27.34)	1.51	6.18	6.54E-10	1.86E-01
	D	21 (1.65)	98 (5.09)	3.09	5.04	4.63E-07			6 (0.31)	>1			
11	α184 A	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10	5.38E-22	231 (18.16)	526 (27.34)	1.51	6.18	6.54E-10	1.86E-01
	T	174 (13.68)	867 (45.06)	3.29	19.58	2.38E-85			6 (0.31)	>1			
12	β57 A	161 (12.66)	838 (43.56)	3.44	19.44	3.80E-84	6.34E-17						
	D	49 (3.85)	34 (1.77)	0.46	-3.54	3.93E-04							
13*	β135 D							153 (12.03)	485 (25.21)	2.10	9.26	1.96E-20	4.24E-16
	G							78 (6.13)	47 (2.44)	0.40	-5.20	2.00E-07	

Among polymorphic AAs in respective clusters, 2 out of 11 AAs in the DQ8/9 haplotype cluster show no association (Fisher exact *P* value >0.05, highlighted in yellow), and 7 out of 12 AAs in the DQ2 haplotype cluster show no association (Fisher exact *P* value >0.05) with T1D. Within respective clusters, the computed Fisher exact calculations indicate whether the residue has differential associations between AAs (Fisher exact *P* value <0.01, highlighted in yellow). Ten AAs, represented by two AAs, are found to have significant differential associations in the DQ8/9 cluster, in which (α22, α23, α44, α49, α51, α53, α54, α73) are in complete LD in the DQ8/9 cluster. Seventeen AAs in the DQ2 cluster, represented by two AAs (α44, β135), have differential T1D associations, in which (α22, α44, α49, α51) are in complete LD and so are (α31, α37, α47, α48, α50, α72, α158, α160, α172, α212, β135). Four shared AAs (α22, α44, α49, α51) are indicative of associations of AAs among carriers of DQ8/9 or those of DQ2. Frequencies (%) of AAs among control subjects are listed within two DQ haplotype clusters, and so are estimated ORs, z scores, and *P* values. Fisher exact *P* values measure the significance of AA association with T1D within each DQ cluster. *There are thirteen residues on the α-chain together with β135 here that are in perfect LD with each other and are listed in Supplementary Table 1. Risk and protective haplotypes are highlighted green and red, respectively, at the significance level of 5%.

and 0.35, $P = 1.96 \times 10^{-20}$ and 8.47×10^{-9} , respectively), and glutamine was observed among patients only. Residue α44 was in complete LD with three other AAs (α22, α49, α51), except that six patients of α44C were separated from the corresponding patient group within the DQ2 cluster. Further, their association patterns were identical. Finally, the AA at β135, representing twelve other AAs in LD, took D (aspartic acid) or G (glycine), which had differential T1D associations (OR 2.10 and 0.40, $P = 1.96 \times 10^{-20}$ and 2.00×10^{-7}), and the difference of their ORs is statistically significant (Fisher $P = 4.24 \times 10^{-16}$). Detailed associations with these β135-linked AAs are shown in Supplementary Table 1. Note that β135D shared the same haplotype with α44C, i.e., they were in high LD, but they were not in complete LD because β135G corresponded to either α44K or α44Q.

With elimination of the uninformative AAs α(-6), α23, α53, α54, α73, α157, and α184, it was clear that two

tagging AAs (α44, β135) were fully representing their T1D associations among carriers of DQ2. For consistency with those used in the DQ8/9 analysis, we included the monomorphic (for the DQ2 cluster) residue β57 and computed their motifs as follows: CAD (b1), KAG (b2), and QAG (b3) (Table 4). The motif CAD conferred risk (OR 2.10, $P = 1.96 \times 10^{-20}$), while the motif KAG conferred resistance (OR 0.35, $P = 8.47 \times 10^{-9}$). Despite the relatively low frequency (0.31%) of the QAG motif, observed among patients only, its absence among control subjects suggested that this motif may predispose carriers to an exceptionally high risk for T1D onset, which would be worthy of independent validation.

The AA β57, a well-established T1D-associated AA, was monomorphic among carriers of DQ2 with the high-risk AA A (alanine). However, the empirical result would suggest that monomorphic β57 did not drive the variable T1D association among DQ2 carriers.

Table 4—Haplotypic association results with selected AAs from both α -chain and β -chain

Motif	T1D				GADA				IA2A				Haplotype Groups	α 44	β 57	β 135	
	Control	Patient	OR	z	P	Pos %	OR	z	P	Pos %	OR	z					P
A) DQ8/9 Carriers																	
a1 KDD	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10									*02:01- *03:03	K	D	.
a2 QAD	161 (12.66)	838 (43.56)	3.44	19.44	3.80E-84									*03:01-*03:02 ~ *03:01-*03:05	.	.	.
a3 QDD	13 (1.02)	29 (1.51)	1.47	1.19	2.35E-01									*03:01- *03:03 ~ *03:02-*03:03	.	D	.
B) DQ2 Carriers																	
b1 CAD	153 (12.03)	485 (25.21)	2.10	9.26	1.96E-20									*05:01-*02:01 ~ *05:05-*02:01	C	.	.
b2 KAG	78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09									*02:01- *02:02	K	.	G
a1 KDD	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10									*02:01- *03:03	K	D	.
a2 QAD	161 (12.66)	838 (43.56)	3.44	19.44	3.80E-84									*03:01-*03:02 ~ *03:01-*03:05	Q	A	D
b3 QAG		6 (0.31)												*03:02- *03:02	.	.	G
C) Carriers of either DQ8/9, DQ2 or both																	
b1 CAD	153 (12.03)	485 (25.21)	2.10	9.26	1.96E-20	345 (29.29)	1.56	5.41	6.35E-08	343 (22.16)	0.59	-6.53	6.55E-11	*05:01-*02:01 ~ *05:05-*02:01	C	.	.
b2 KAG	78 (6.13)	41 (2.13)	0.35	-5.76	8.47E-09	31 (2.63)	1.96	1.93	5.34E-02	35 (2.26)	1.42	0.81	4.18E-01	*02:01- *02:02	K	.	G
a1 KDD	36 (2.83)	5 (0.26)	0.09	-6.36	1.99E-10									*02:01- *03:03	K	D	.
a2 QAD	161 (12.66)	838 (43.56)	3.44	19.44	3.80E-84	487 (41.34)	0.88	-2.90	3.70E-03	730 (47.16)	1.64	7.63	2.40E-14	*03:01-*03:02 ~ *03:01-*03:05	Q	A	D
b3 QAG		6 (0.31)												*03:02- *02:02	.	.	G
a3 QDD	13 (1.02)	29 (1.51)	1.47	1.19	2.35E-01	20 (1.7)	1.41	0.87	3.85E-01	24 (1.55)	1.17	0.32	7.51E-01	*03:01- *03:03 ~ *03:02-*03:03	.	D	.

Estimated haplotype frequency (%) (out of 636 control subjects and 962 patients), estimated ORs, haplotype score, and P values from association analysis with T1D (in case-control study), with GADA level among patients, and with IA-2A level among patients. For given haplotypes, one or more DQA1-DQB1 haplotypes may be merged into haplotype groups. Estimated haplotype frequencies (%) among those with positive autoantibody, estimated ORs, z scores, and P value are listed: A) three motifs with effectively two AAs (α 44, β 57) among carriers of DQ8/9 haplotypes, B) three motifs with effectively two AAs (α 44, β 135) among carriers of DQ2, and C) six motifs with three AAs (α 44, β 57, β 135) among carriers of DQ8/9 or DQ2. AAs (α 44, α 23, α 49, α 51, α 53, α 54, α 73, α 73, α 73, α 73) in DQ8/9 mixture are in complete LD and can be represented by α 44, and AAs (α 22, α 49, α 51) are in complete LD in DQ2 mixture, and their polymorphisms are captured by α 44. Further, (β 135, α 31, α 37, α 48, α 50, α 72, α 104, α 153, α 158, α 160, α 172, α 212) are in complete LD with each other and thus represented by β 135. Hence, three AAs (α 44, β 57, β 135) capture all polymorphisms of their motifs in both DQ2 and DQ8/9 mixtures. Haplotype groups include haplotypes that share the same DQ motifs. DQ2.5 refers to DQA1*05:01-DQB1*02:01 (blue), and DQ8.1 includes either DQA1*03:01-DQB1*03:02 or DQA1*03:02-DQB1*03:02 (red). Risk and protective haplotypes are highlighted green and red, respectively, at the significance level of 5%. pos, positive.

Three AAs (α 44, β 57, β 135) Capture DQ Risk/Resistance Associations With T1D Among Carriers of DQ8/9 or DQ2

Due to extended LD of residues among carriers of DQ8/9 and those of DQ2, three AAs (α 44, β 57, β 135) appeared to adequately represent all T1D-associated polymorphisms of HLA-DQA1 and -DQB1 with six variable motifs (Table 4). We computed their associations among carriers of DQ8/9 and DQ2. Interestingly, six motifs among carriers of DQ8/9 or DQ2 consisted of those two sets of motifs observed above. Collectively, the motif CAD association captured that of DQ2.5 and QAD that of DQ8.1. Meanwhile, motifs KAG and KDD captured two resistant DQ haplotypes (*02:01-02:02 and *02:01-03:03, respectively). Tentatively, one would conclude that these six motifs captured all risk and resistant associations with T1D among carriers of DQ8/9 and DQ2. It was particularly worth noting that two high-risk DQ motifs shared the AAs at β 57 and β 135 but differed only at α 44 with AA C (cysteine) versus Q (glutamine).

Three Selected AAs Explain Substantial Portion of HLA-DQ Associations With T1D in the General Population

It was well known that individual AAs in HLA-DQ genes, due to LD, had substantial associations with T1D in the general population (Fig. 1A). At question was whether three identified AAs explained some of HLA-DQA1 and -DQB1 marginal associations. Figure 1B showed *P* values from the conditional association analysis, after adjustment for (α 44, β 57, β 135). Note that the *y*-axis is scaled from 0 to 10—much narrower than the scale from 0 to 120 in Fig. 1A. Clearly, these three AAs had greatly reduced *P* values for all HLA-DQ residues in the general population. Among 104 residues, conditional *P* values on 63 residues were $>5\%$, even though their marginal *P* values were $<5\%$ (Fig. 1C). Interestingly, there were two residues with insignificant marginal *P* values, but their conditional *P* values became significant. Finally, after the adjustment, 28 residues, with marginally significant *P* values, continued to have conditionally significant *P* values, even though conditional *P* values are much less significant than marginal ones.

Three Selected AAs Associate With Elevations of GADA and IA-2A Levels Among Their Carriers

To examine functional relevance of selected AAs, we then investigated whether their motifs associated with the panel of six islet autoantibodies measured at the time of diagnosing T1D. It turned out that their associations with GADA and IA-2A levels were substantial (Table 4). Associations with the other four autoantibody levels were minimum or modest, and results are presented in Supplementary Table 2. Among carriers of DQ8/9, the motif QAD appeared to be negatively associated with the GADA elevation (OR 0.88, $P = 3.70 \times 10^{-3}$) and yet to be positively

associated with IA-2A (OR 1.64, $P = 2.40 \times 10^{-14}$), listed in Table 4. On the other hand, the motif CAD among carriers of DQ2 appeared to be positively associated with GADA (OR 1.56, $P = 6.35 \times 10^{-8}$) and negatively associated with IA-2A (OR 0.59, $P = 6.55 \times 10^{-11}$) (Table 4). Among carriers of either DQ8/9 or DQ2, the same association patterns appear, with minor numerical variations due to changes of sample sizes.

Structural Analysis

The structural ramifications of the HOH analysis for the DQ8/9 cluster are shown in Fig. 3A. We chose as an example the binding of the human insulin B11-23/24Gly peptide to the HLA-DQ8 (A1*0301/B1*0302) molecule as shown by X-ray crystallography (38,39). This was presumed to take place in the endosome of the antigen-presenting cells and later displayed on the surface of such cells for presentation to cognate CD4⁺ T cells. In the case of islet β -cells of mice and humans, it had been shown that islet macrophages were in contact with β -cells and received from the latter granules containing processed peptides from islet autoantigens such as proinsulin (42). All residues shown in Table 2 as linked to the pathogenesis of T1D are shown in Fig. 3A, except for $\alpha(-6)$ and α 184, as they were not part of the determined crystal structure (includes α -chain residues 2-181 and β -chain residues 3-192). The majority of the implicated residues (α 22, α 23, α 44, α 49, α 51, α 53, α 54) were around pocket 1 (p1E anchor shown for reference), while one (β 26) was part of pocket 4, two others (α 73, β 57) were part of pocket 9, and α 157 was one of the set of residues that participated in the presumed formation of the TCR-induced homodimer of HLA-DQ $\alpha\beta$ heterodimers bearing identical peptides (35,36,43). The residues around pocket 1 might influence peptide binding indirectly; yet, many of them appeared to have more influence on the TCR recognition by cognate CD4⁺ T cell clones that cross recognize the insulin B11-23 peptide bound to DQ8-*cis* or DQ8-*trans* (A1*0501/B1*0302) (44). We further analyzed the AA contributions in the DQ2 cluster of alleles predisposing to or protecting from T1D. The entire set of residues consisted of $\alpha(-13)$, $\alpha(-6)$, α 22, α 23, α 31, α 37, α 44, α 47, α 48, α 49, α 50, α 51, α 53, α 54, α 72, α 73, α 104, α 153, α 157, α 158, α 160, α 172, α 184, α 212, and β 135. Of these, we could not depict the first two or the last two residues from the α -chain because the relevant crystal structure only contained residues α 1- α 180. Of the depicted residues, two are in pocket 1 (α 31 and α 51), while nine others were near it (α 22, α 23, α 37, α 44, α 47, α 49, α 50, α 53, α 54), and two are part of pocket 9 (α 72, α 73), while α 157 and α 158 interacted with β 2 domain residues for the formation of the presumed homodimer of HLA-DQ $\alpha\beta$ heterodimers, with α 172 participating in the contact to the CD4 coreceptor molecule via such a presumed homodimer (Fig. 3B). The intramembrane residue α 212 was part of the cholesterol recognition motif of MHC II molecules within the lipid bilayer that HLA-DQ molecules cross in a single span (45,46). Last, β 135 was in

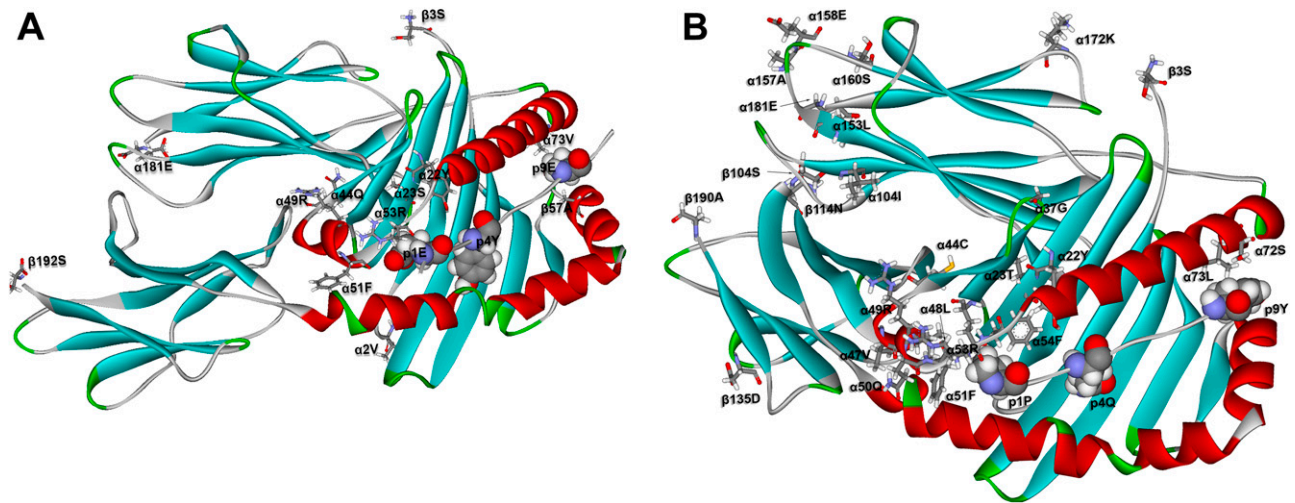


Figure 3—A: Depiction of the crystal structure of the HLA-DQ8–insulin B11–23/24Gly complex (from 1jk8.pdb) highlighting the AA residues shown by HOH analysis to be involved in T1D pathogenesis via this MHC II allele. The orientation of the molecule is with its long axis nearly parallel to the plane of the paper/screen. The cognate TCR would bind on the righthand side of the molecule, recognizing simultaneously peptide residues as well as HLA-DQ8 residues from the $\alpha 1\beta 1$ antigen-binding domain. The HLA-DQ8–insulin peptide complex is depicted in its secondary structure formation (α -helix in red, β -sheet in turquoise, β -turn in green, and random coil or any other form, such as polyproline II helix of the antigenic peptide backbone, in gray). The AA residues in question are depicted in stick form (atom color convention: carbon, gray; oxygen, red; nitrogen, blue; sulfur, yellow; hydrogen, white). Three anchor residues of the antigenic peptide (p1Glu, p4Tyr, and p9Glu) are shown in space-filling form and the same atomic color conventions, only for orientation purposes in order to appreciate the position and role of the identified residues from the HLA-DQ8–insulin peptide complex. The sequence $\beta 105$ – $\beta 112$ did not diffract satisfactorily and was not depicted in the originally reported structure. The gap is not clearly seen in this orientation. We also depict the N-terminal residues of the two DQ chains, $\alpha 2$ Val and $\beta 3$ Ser, as well as the two C-terminal residues of the two DQ chains: $\alpha 181$ Glu and $\beta 192$ Ser. B: Depiction of the crystal structure of the HLA-DQ2–gliadin $\alpha 1$ complex (from 6mfg.pdb) highlighting the AA residues shown by HOH analysis to be involved in T1D pathogenesis via this MHC II allele. A similar orientation and the same color and depiction conventions for secondary structures and individual AA renderings hold, as in A. Three anchor residues of the antigenic peptide (p1Pro, p4Gln, and p9Tyr) are again shown, only for orientation purposes. As in the structure of HLA-DQ8 (A), the sequence $\beta 105$ – $\beta 113$ did not diffract satisfactorily and was not depicted in the originally reported structure. We label here residues $\beta 104$ Ser and $\beta 114$ Asn in order to identify the break in the β -chain. We also depict the C-terminal residues of the two DQ chains: $\alpha 181$ Glu and $\beta 190$ Ala. The gliadin $\alpha 1$ peptide contains several prolines, a property unique to various gluten peptides and not encountered in any of the thus far identified DQ2-restricted antigenic peptides deemed to be involved in T1D pathogenesis.

the β -sheet stretch $\beta 134$ – 148 that is involved in CD4 binding (45–48).

DISCUSSION

Building upon profound literature on T1D associations with HLA-DQA1 and -DQB1 (21,39,41,49), we used the state of the art sequencing technology to genotype two DQ genes for all 636 control subjects and 962 patients in this large population-based case-control study of T1D. The high genotyping quality allowed us to reconstruct their respective AA sequences for each haplotype of DQA1-DQB1 and to drill into haplotype-specific AA sequences, searching for critical residues through applying the HOH approach. HOH identified seven phylogenetic clusters of DQ haplotypes based on distances of AA sequences between all unique DQ haplotypes, which coincided with classical serotypes. Here, HOH focused the initial discovery analysis on two risk DQ clusters: 1) the DQ8/9 cluster included carriers of two risk DQ8.1 and a resistant serotype DQ9.2, in addition to one neutral serotype DQ9.3 and two rare haplotypes, and 2) the DQ2 cluster included the well-known risk haplotype DQ2.5, relatively rare but risk haplotype DQ2.3, and one resistant one, DQ2.2, in addition

to a rare haplotype. Despite the high similarity of corresponding AA sequences between those haplotypes within each cluster, their associations with T1D risk ranged from resistance to risk, and the variable association allowed us to identify specific associated AAs. Indeed, the result from HOH analysis within carriers of DQ8/9 led to identification of 10 AAs ($\alpha 22$, $\alpha 23$, $\alpha 44$, $\alpha 49$, $\alpha 51$, $\alpha 53$, $\alpha 54$, $\alpha 73$, $\alpha 184$, $\beta 57$), which could be represented by ($\alpha 44$, $\beta 57$, $\beta 135$). These AAs formed three motifs, KDD, QAD, and QDD, uniquely corresponding to resistant, risk, and neutral haplotype groups (OR 0.09, 3.44, and 1.47, $P = 1.99 \times 10^{-10}$, 3.80×10^{-84} , and 0.235, respectively). In contrast, among carriers of DQ2, HOH analysis uncovered a set of AAs, represented by three AAs ($\alpha 44$, $\beta 57$, $\beta 135$) that formed motifs KAG and CAD, capturing resistant and risk associations of DQ haplotypes (OR 0.35 and 2.10, $P = 8.47 \times 10^{-9}$ and 1.96×10^{-20}). Further, the motif QAG was relatively uncommon but observed only among patients, as DQA1*03:02-B1*02:02, implying that it was more likely to be a risk motif. Complete LD with $\beta 135$ also implicated potential associations with 12 other AAs on the α -chain ($\alpha 31$, $\alpha 37$, $\alpha 47$, $\alpha 48$, $\alpha 50$, $\alpha 72$, $\alpha 104$, $\alpha 153$, $\alpha 158$, $\alpha 160$, $\alpha 172$, $\alpha 212$). Furthermore, integrating both sets of identified AAs led to

identification of three AAs ($\alpha 44$, $\beta 57$, $\beta 135$) and six motifs among carriers of either DQ8/9 or DQ2, or both, and their associations with T1D were consistent with those separately observed above. The same association patterns with GADA and IA-2A were reproduced with three identified AAs. Collectively, these three AAs, together with those LD-linked AAs, likely harbored critical residues for T1D risks, mediated by GADA and IA-2A, among carriers of DQ2, DQ8, and DQ9.

The HLA-DQ8/9 cluster was remarkable in that it consisted of one resistant haplotype DQA1*0201-DQB1*0303 and several susceptible as well as neutral alleles (Table 2). The resistant haplotype was different in every 1 of the 10 AAs that distinguished members of the DQ8/9 cluster (Table 2). Of the 10 AAs, the great majority ($\alpha 22$, $\alpha 23$, $\alpha 44$, $\alpha 49$, $\alpha 51$, $\alpha 53$) were around pocket 1, with many of them being putative cognate TCR contacts. Residue $\alpha 73$ was part of pocket 5, while residue $\alpha 184$ had no known function, and $\beta 57$ was the well-known pivotal residue, when an Asp (mostly in resistant alleles among the T1D-related ones) formed a most important salt bridge with $\alpha 76$ Arg, in addition to hydrogen bonds to p9C = O and p10NH (35,36). In the case of $\beta 57$ Ala, acidic p9 anchors were preferred, with their side-chain carboxylate forming the said salt bridge (37). Considering the alleles under the DQ2 cluster, we noted first with interest the two alleles sharing the same β -chain: HLA-DQA1*02:01-DQB1*02:02 and -DQA1*03:02-DQB1*02:02. The first allelic combination was truly a resistant one (OR 0.36), while the second one was found only among six patients and not a single control subject, eliminating the possibility of being a resistant DQ haplotype. The DQB1*02:02 chain was remarkable among resistant alleles: it was a $\beta 57$ Asp⁻ allele (with an Ala instead), a feature reported for the first time regarding HLA-DQ alleles. The 17 uncovered AAs of the DQ2 cluster, represented by ($\alpha 44$, $\beta 135$), were ($\alpha 22$, $\alpha 44$, $\alpha 49$, $\alpha 51$) and ($\alpha 31$, $\alpha 37$, $\alpha 47$, $\alpha 48$, $\alpha 50$, $\alpha 72$, $\alpha 104$, $\alpha 153$, $\alpha 158$, $\alpha 160$, $\alpha 172$, $\alpha 212$, $\beta 135$) and $\alpha 31$ within or several others near pocket 1 ($\alpha 22$, $\alpha 31$, $\alpha 43$, $\alpha 44$, $\alpha 47$, $\alpha 48$, $\alpha 49$, $\alpha 50$, $\alpha 51$), in pocket 9 ($\alpha 72$), in the $\alpha 2$ domain as belonging to the set of residues stabilizing the presumed homodimer of heterodimers ($\alpha 158$), or interacting with CD4 ($\alpha 172$); one within the single intramembranous stretch responsible for cholesterol binding ($\alpha 212$ F but in these alleles it is L) (47,48); and one in the β -sheet from the $\beta 2$ domain ($\beta 135$) that binds to CD4 (35,45,46). Note that there are five clusters with resistant or neutral DQ haplotypes (Fig. 2), and identification of corresponding resistant residues is of interest and being pursued. Interestingly, the just published crystal structure of HLA-DQ2.2 (A1*02:01/B1*02:02), a nonsusceptible allele for celiac disease, shows that residue $\alpha 22$ F induces a different residue selectivity for the p3 shelf of the antigen-binding groove, such that cognate TCRs would dock at different angles to the pDQ2.2 complexes (as compared with complexes of the disease-susceptible DQA1*05:01/B1*02:01), resulting

in altogether different biochemical/biological results of such an engagement (50,51).

The four AAs ($\alpha 22$, $\alpha 44$, $\alpha 49$, $\alpha 51$), that recaptured collectively the risk of as well the resistance from T1D and the possibility or lack thereof for GADA and IA-2A production, were all near pocket 1 with $\alpha 22$ Tyr, certain alleles that participate in hydrogen bonding to the antigenic peptide backbone via a water bridge—in contrast to $\alpha 22$ Phe alleles that cannot do so (50,52). The importance of residues $\alpha 31$, $\alpha 44$ – $\alpha 50$, and $\alpha 53$ in the recognition of DQ8-cis- and DQ8-trans-insulin 12–23 complexes by the same cognate T-cell receptor (TCR) has recently been demonstrated (44).

In the discussion of the putative role(s) of several identified residues by the HOH process, we were hampered by the state of knowledge on the properties of the HLA-DR/DQ $\alpha 2\beta 2$ domain. This was a region sandwiched between the $\alpha 1\beta 1$ domain, which bound the antigenic peptide, and the thus-formed complex was presented to the cognate TCR for mutual activation of both the antigen-presenting cell (APC) and the CD4⁺ T cell, as well as the short single-span intramembranous and the intracellular domains. Therefore, any conformational changes in the $\alpha 1\beta 1$ domain upon TCR engagement were transmitted to the intramembranous and intracytoplasmic domains via the $\alpha 2\beta 2$ domain (53,54). Besides the interaction of this domain with CD4, subsequent to cognate TCR engagement, precious little was known about any other accessory proteins to the MHC II molecules (mouse, human, or other), especially in this domain (55). It was not known, for example, whether the MHC II molecules remained continuously “upright,” with respect to the APC’s cell membrane, for ease of TCR sampling. The simple architecture of the immunological synapse that formed upon phosphorylated MHC II–cognate TCR multiengagement seemed to be composed of cognate ligand and receptor molecules from both APC and T cells that were circa 75 Å in extracellular length, such as pMHC II and TCR (56,57). The only known and established interaction in this domain was that with the coreceptor CD4, and conflicting interpretations exist to this day regarding its mode of interaction, i.e., whether the MHC II molecule was in an $\alpha\beta$ monomeric form or an $(\alpha\beta)_2$ homodimeric form (35,45,58,59). The data in the mouse I-A molecule showed evidence in favor of a homodimer of heterodimers for CD4 coreceptor binding and subsequent T-cell activation (58,60,61), and the CD4-MHC II interaction in the mouse was of far lower affinity than in the human system; thus, one critical experiment concerning this issue might be mutagenesis of all the putative CD4 contact residues in the respective two $\alpha\beta$ surfaces of the HLA-DR/DQ $(\alpha\beta)_2$ homodimer that may act as the CD4 receptors (58,59,62,63). One of course could not eliminate a priori the possibility of a single charged residue, among the residues presuming to participate in the homodimer formation, mediating the most crucial of several residue-residue interactions, as in the case of the CD2-CD48 interaction (64). Intriguingly, the T1D susceptibility alleles

HLA-DQ2, HLA-DQ8, HLA-DQ2-*trans* (A1*03:01/B1*02:01), and HLA-DQ8-*trans* (A1*05:01/B1*03:02) are resistant (the first much more than the rest) to HLA-DM-mediated removal of the CLIP peptide in the endosome and parallel loading of processed antigenic peptides found there, a property not seen in T1D neutral (DQ1) or resistant (DQ6) alleles (65). CLIP and extended CLIP peptides are found in the pool of peptides eluted from the groove of these very HLA-DQ alleles, and it is possible that in the susceptible alleles this might influence the formation of the CD4⁺ T-cell repertoire in various ways (e.g., inability to form a sufficient amount of islet antigen-specific DQ2/8-*cis*- and -*trans*-restricted regulatory T cells).

One surprising and potentially important observation was that T1D associations with AAs of HLA-DQA1 and -DQB1 were specific to protein structures. A conventional approach in statistical genetic analysis was to identify a genetic polymorphism of either AA or nucleotide that was statistically associated with disease phenotype, under the assumption that the variant association is universal. Within DQ molecules, multiple crucial AAs had different distributions among carriers of DQ8/9 and those of DQ2 or had different association patterns. For example, residue α 44 had three different AAs (C, K, and Q), in which the AA C had noticeable frequencies (12% and 25% among, respectively, control subjects and patients carrying DQ2) and was completely absent in carriers of DQ8/9. Consequently, the conventional analysis, without acknowledging such complexities, would yield variable and seemingly inconsistent associations across residues within genes or across studies of the same residues. Statistically, this phenomenon could be referred to as high-order interactions, i.e., individual AAs associated with disease phenotype, within a specific "protein structure," which was specified by multiple AAs in this study. Indeed, recognizing the presence of such high-order interactions motivated the HOH strategy to uncover pertinent AAs from DQA1 and DQB1.

The current HOH analysis was based on haplotypes of DQA1 and DQB1, i.e., *cis* effects, without considering *trans* effects, which may be viewed as a limitation. *DQA1* coded an α -chain and *DQB1* coded a β -chain, with each chain containing α -helical and β -pleated sheet structural elements in their extracellular portions (>75% of the mature sequences). In the presence of double heterozygotes, e.g., DQ8.1/DQ2.5 = DQA1*03:01-B1*03:02/DQA1*05:01-B1*02:01, there were four possible heterodimers formed by DQA1*03:01-B1*03:02 (*cis*), DQA1*05:01-B1*02:01 (*cis*), DQA1*03:01-B1*02:01 (*trans*), and DQA1*05:01-B1*03:02 (*trans*). Such a heterodimeric effect was not explicitly incorporated in the current HOH analysis. This ignorance may have some impact on HOH focusing on carriers of DQ8/9 or carriers of DQ2, since some carriers of, say, DQ8/9, may be doubly heterozygous with DQ2 as the second haplotype. If so, estimated associations may be attenuated toward the null. However, their influence had not negated the discovery presented here. It was recognized that investigating heterodimeric effects

was important and required a new analytic HOH strategy to fully account for both *cis* and *trans* effects of doubly heterozygous DQA1 and DQB1 (to be pursued and presented in the future).

One potential limitation to the current HOH analysis is that HLA-DRB1 may confound DQ associations, due to high LD between DR and DQ genes. Specifically, DR4-DQ8.1 and DR3-DQ2.5 are known to be high-risk haplotypes for T1D. Hence, it is of importance to differentiate T1D associations with DR and DQ through haplotype association analysis or through family studies in the future. Similarly, the future study should also include the assessment of diplotypic associations and the joint DR-DQ association with autoantibodies. Another future direction is to assess binding affinities of discovered motif CAD and QAD as a functional validation, results from which will be reported in the future.

Taken together, these studies using next-generation sequencing of HLA DQA1-DQB1 in T1D patients and control subjects made it possible to phase haplotypes without knowing inheritance by descent. With application of an HOH association analysis, 45 unique DQ haplotypes were distributed among seven clusters. The DQ8/9 cluster included two DQ8.1 risk haplotypes and the DQ9 resistant haplotype, and the DQ2 cluster included the DQ2.5 risk and DQ2.2 resistant haplotypes, respectively. Within each cluster, HOH found residues α 44Q and β 57A to be significantly associated with T1D. Similarly, within the DQ2 cluster, HOH found α 44C and β 135D to share the same risk for T1D. The motif QAD of α 44, β 57, and β 135 captured the T1D risk association of DQ8.1, and the corresponding motif CAD captured the risk association of DQ2.5. Two risk associations were mediated through GADA and IA-2A but in opposite directions. CAD was positively associated with GADA but negatively with IA-2A, and QAD was negatively associated with GADA but positively with IA-2A despite the single difference at α 44. The AA residues that we now have identified are to be found in and around anchor pockets 1 and 9, as potential TCR contacts, in the areas for CD4 binding and putative homodimer formation. The identification of the three HLA-DQ AA residues (α 44, β 57, β 135), which capture, due to LD, the conferred T1D risk, should prove useful in future studies of autoantigen-specific immunomodulation. For example, CD4⁺ T cells generated from T1D patients can be tested for specificity to particular peptides of the known autoantigens (preproinsulin, GAD65, IA2, and ZnT8) and restriction of TCR recognition via particular HLA-DQ alleles (44). The importance of certain residues may then be revealed by the use of antigen-presenting cells bearing closely related HLA-DQ alleles that differ in one or more of the AA residues (preferably outside the antigen-binding groove) shown in this study to encapture all the risk for T1D.

Acknowledgments. The authors thank Drs. Annelie Carlsson, Department of Pediatrics, Lund University, Lund, Sweden; Gun Forsander, Department of

Pediatrics, Queen Silvia Children's Hospital, Sahlgrenska University Hospital, Gothenburg, Sweden; and Sten A. Ivarsson, Department of Clinical Sciences, Lund University/CRC, Skåne University Hospital, and Martina Persson, Clinical Epidemiology, Institution of Medicine, and Department of Clinical Science and Education, Karolinska Institutet, for clinical expertise in the BDD study. The authors also thank Anita Ramelius, Linda Faxius, Karl Moritz, Petra Moritz, Ingrid Wigheden, Ida Jönsson, and Rasmus Bennet (all at the Department of Clinical Sciences, Lund University/CRC) for expert technical assistance.

Funding. The study was supported by a National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health, grant (1R01DK117276); a grant from the European Foundation for the Study of Diabetes Clinical Research Grants Programme 2013 and in part the Swedish Child Diabetes Foundation (Barndiabetesfonden); the NIDDK, National Institutes of Health (DK63861 and DK26190); the Swedish Research Council including a Linné grant to Lund University Diabetes Centre; the Skåne County Council for Research and Development; and the Swedish Association of Local Authorities and Regions (SKL).

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

Author Contributions. L.P.Z., Å.L., and G.K.P. researched and analyzed the data and wrote the manuscript. W.W.K. reviewed and commented on the manuscript with respect to structural implications and discussions. D.E.G. led the team of R.W., C.-W.P., and W.C.N. in the next-generation sequencing and researched data. A.K.M. and G.P.B. carried out graphical representations of select HLA-DQ molecules, contributed to discussion, and reviewed the manuscript. J.L. critically reviewed statistical results and improved presentation of results. H.E.L., J.L., U.S., C.M., and Å.L. designed the BDD study, researched data, contributed to 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.

References

- Khunti K, Alsifri S, Aronson R, et al.; HAT Investigator Group. Impact of hypoglycaemia on patient-reported outcomes from a global, 24-country study of 27,585 people with type 1 and insulin-treated type 2 diabetes. *Diabetes Res Clin Pract* 2017;130:121–129
- Regnell SE, Lernmark Å. Early prediction of autoimmune (type 1) diabetes. *Diabetologia* 2017;60:1370–1381
- DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *Lancet* 2018;391:2449–2462
- Lind M, Svensson AM, Kosiborod M, et al. Glycemic control and excess mortality in type 1 diabetes. *N Engl J Med* 2014;371:1972–1982
- Katsarou A, Gudbjörnsdóttir S, Rawshani A, et al. Type 1 diabetes mellitus. *Nat Rev Dis Primers* 2017;3:17016
- Rawshani A, Landin-Olsson M, Svensson AM, et al. The incidence of diabetes among 0–34 year olds in Sweden: new data and better methods. *Diabetologia* 2014;57:1375–1381
- Herold KC, Usmani-Brown S, Ghazi T, et al.; Type 1 Diabetes TrialNet Study Group. β Cell death and dysfunction during type 1 diabetes development in at-risk individuals. *J Clin Invest* 2015;125:1163–1173
- Krischer JP, Lynch KF, Schatz DA, et al.; TEDDY Study Group. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia* 2015;58:980–987
- Krischer JP, Liu X, Lernmark Å, et al.; TEDDY Study Group. The influence of type 1 diabetes genetic susceptibility regions, age, sex, and family history on the progression from multiple autoantibodies to type 1 diabetes: a TEDDY study report. *Diabetes* 2017;66:3122–3129
- Krischer JP, Lynch KF, Lernmark Å, et al.; TEDDY Study Group. Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY Study. *Diabetes Care* 2017;40:1194–1202
- Vehik K, Lynch KF, Schatz DA, et al.; TEDDY Study Group. Reversion of β -cell autoimmunity changes risk of type 1 diabetes: TEDDY study. *Diabetes Care* 2016;39:1535–1542
- 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
- 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
- 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
- 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
- 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
- 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
- Zhao LP, Papadopoulos GK, Kwok WW, et al. 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. *Diabetes* 2019;68:1692–1704
- She JX. Susceptibility to type 1 diabetes: HLA-DQ and DR revisited. *Immunol Today* 1996;17:323–329
- Papadopoulos GK. Genetics of type 1 diabetes. *Immunol Today* 1998;19:288–289
- Li QR, Zang WQ, Gao F, Du Y, Zhang QY. Effects of CD74 gene on IFN- γ gene expression in MG TEC. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* 2011;27:730–732 [in Chinese]
- Murphy K, Weaver C. *Janeway's Immunobiology*. New York, NY, Garland Science/Taylor & Francis Group, LLC, 2016
- 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
- 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
- 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
- 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
32. 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
33. Fremont DH, Monnaie D, Nelson CA, Hendrickson WA, Unanue ER. Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. *Immunity* 1998; 8:305–317
34. 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
35. 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
36. Stern LJ, Brown JH, Jardetzky TS, et al. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 1994; 368:215–221
37. Lee KH, Wucherpfennig KW, Wiley DC. Structure of a human insulin peptide-HLA-DQ8 complex and susceptibility to type 1 diabetes. *Nat Immunol* 2001;2:501–507
38. Dahal-Koirala S, Ciacchi L, Petersen J, et al. Discriminative T-cell receptor recognition of highly homologous HLA-DQ2-bound gluten epitopes. *J Biol Chem* 2019;294:941–952
39. Hippich M, Beyerlein A, Hagopian WA, et al.; TEDDY Study Group; Teddy Study Group. Genetic contribution to the divergence in type 1 diabetes risk between children from the general population and children from affected families. *Diabetes* 2019;68:847–857
40. Fourlanos S, Harrison LC, Colman PG. The accelerator hypothesis and increasing incidence of type 1 diabetes. *Curr Opin Endocrinol Diabetes Obes* 2008; 15:321–325
41. Pociot F, Lernmark Å. Genetic risk factors for type 1 diabetes. *Lancet* 2016; 387:2331–2339
42. Unanue ER, Wan X. The immunoreactive platform of the pancreatic islets influences the development of autoreactivity. *Diabetes* 2019;68:1544–1551
43. 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
44. Chow IT, Gates TJ, Papadopoulos GK, et al. (2019). Discriminative T cell recognition of cross-reactive islet-antigens is associated with HLA-DQ8 transdimer-mediated autoimmune diabetes. *Sci Adv* 2019;5: eaaw9336
45. Cammarota G, Scheirle A, Takacs B, et al. Identification of a CD4 binding site on the beta 2 domain of HLA-DR molecules. *Nature* 1992;356: 799–801
46. 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
47. Dixon AM, Drake L, Hughes KT, et al. Differential transmembrane domain GXXG motif pairing impacts major histocompatibility complex (MHC) class II structure. *J Biol Chem* 2014;289:11695–11703
48. Dixon AM, Roy S. Role of membrane environment and membrane-spanning protein regions in assembly and function of the class II major histocompatibility complex. *Hum Immunol* 2019;80:5–14
49. Fourlanos S, Varney MD, Tait BD, et al. The rising incidence of type 1 diabetes is accounted for by cases with lower-risk human leukocyte antigen genotypes. *Diabetes Care* 2008;31:1546–1549
50. Ting YT, Dahal-Koirala S, Kim HSK, et al. A molecular basis for the T cell response in HLA-DQ2.2 mediated celiac disease. *Proc Natl Acad Sci U S A* 2020; 117:3063–3073
51. van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Papadopoulos GK, Koning F. Unique peptide binding characteristics of the disease-associated DQ(alpha 1*0501, beta 1*0201) vs the non-disease-associated DQ(alpha 1*0201, beta 1*0202) molecule. *Immunogenetics* 1997;46:484–492
52. Bodd M, Kim CY, Lundin KE, Sollid LM. T-cell response to gluten in patients with HLA-DQ2.2 reveals requirement of peptide-MHC stability in celiac disease. *Gastroenterology* 2012;142:552–561
53. Painter CA, Stern LJ. Conformational variation in structures of classical and non-classical MHCII proteins and functional implications. *Immunol Rev* 2012;250: 144–157
54. 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
55. Tambur AR. Human leukocyte antigen matching in organ transplantation: what we know and how can we make it better (revisiting the past, improving the future). *Curr Opin Organ Transplant* 2018;23:470–476
56. Junghans V, Santos AM, Lui Y, Davis SJ, Jönsson P. Dimensions and interactions of large T-cell surface proteins. *Front Immunol* 2018;9:2215
57. Fernandes RA, Ganzinger KA, Tzou JC, et al. A cell topography-based mechanism for ligand discrimination by the T cell receptor. *Proc Natl Acad Sci U S A* 2019;116:14002–14010
58. König R, Shen X, Germain RN. Involvement of both major histocompatibility complex class II alpha and beta chains in CD4 function indicates a role for ordered oligomerization in T cell activation. *J Exp Med* 1995;182:779–787
59. 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
60. Denning TL, Qi H, König R, Scott KG, Naganuma M, Ernst PB. CD4+ Th cells resembling regulatory T cells that inhibit chronic colitis differentiate in the absence of interactions between CD4 and class II MHC. *J Immunol* 2003;171:2279–2286
61. Shen X, Niu C, König R. Increased numbers and suppressive activity of regulatory CD25(+)CD4(+) T lymphocytes in the absence of CD4 engagement by MHC class II molecules. *Cell Immunol* 2013;282:117–128
62. 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
63. Brameshuber M, Kellner F, Rossboth BK, et al. Monomeric TCRs drive T cell antigen recognition. *Nat Immunol* 2018;19:487–496
64. Davis SJ, Davies EA, Tucknott MG, Jones EY, van der Merwe PA. The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proc Natl Acad Sci U S A* 1998;95:5490–5494
65. Jurewicz MM, Stern LJ. Class II MHC antigen processing in immune tolerance and inflammation. *Immunogenetics* 2019;71:171–187