The Pathogenesis and Natural History of Type 1 Diabetes

Mark A. Atkinson

College of Medicine, Departments of Pathology and Pediatrics, The University of Florida, Gainesville, Florida 32610-0275

Correspondence: atkinson@ufl.edu

The purpose of this article is to provide an overview that summarizes much in the way of our current state of knowledge regarding the pathogenesis and natural history of type 1 diabetes in humans. This information is presented to the reader as a series of seminal historical discoveries that, when advanced through research, transformed our understanding of the roles for the immune system, genes, and environment in the formation of this disease. In addition, where longitudinal investigations of these three facets occurred, their roles within the development of type 1 diabetes, from birth to symptomatic onset and beyond, are discussed, including their most controversial elements. Having an understanding of this disorder’s pathogenesis and natural history is key for attempts seeking to understand the issues of what causes type 1 diabetes, as well as to develop a means to prevent and cure the disorder.

Type 1 diabetes (T1D) is a disorder that arises following the autoimmune destruction of insulin-producing pancreatic β cells (Atkinson 2001; Bluestone et al. 2010). The disease is most often diagnosed in children and adolescents, usually presenting with a classic trio of symptoms (i.e., polydypsia, polyphagia, polyuria) alongside of overt hyperglycemia, poising the immediate need for exogenous insulin replacement—a medicinal introduction to the disorder whose therapeutic practice lasts a lifetime.

These introductory facets having been said, many other etiological and typology-based aspects for this disease remain either unclear or subject to significant debate within the medical research community. Among these are questions related to the percentage of T1D cases that are diagnosed in adults, a figure whose estimates range from a low of 25% to as much as 50% (Thunandera et al. 2008). Indeed, multiple factors contribute to this knowledge void, one being a failure in understanding the percentage of T1D cases that are errantly misclassified as type 2 diabetes (T2D). Specifically, it has been proposed that ~5%–15% of adults diagnosed with T2D may, in actuality, have T1D (for review, see Palmer et al. 2005). Were this true, the notion that 90%–95% of all diabetes cases are diagnosed as T2D would mean that the number of T1D cases is likely far underestimated. Attempts to distinguish T1D cases from those with T2D have also resulted in a proposed new disease classification, Latent Autoimmune Disease of Adults (LADA) (for review, see Leslie et al.)
2008). However, over this past decade, the lack of firm diagnostic criteria for LADA, taken together with other notions (e.g., genetic similarity between those with T1D and the so-called LADA patients), has dramatically decreased, but not eliminated, enthusiasm for adopting this presumed “new” disease entity as a novel category for diabetes (Rolandsson 2010). To be clear, such confusion over disease classification in settings of diabetes is not new because many terms (e.g., insulin-requiring diabetes, juvenile diabetes, insulin-dependent diabetes mellitus, etc.) have been used over the years to describe what we now refer to as T1D; each term eventually being replaced, to a large extent, as improvements occurred in our understanding of the pathogenesis and natural history of this disease. For T1D, expert panels formed under the auspices of the American Diabetes Association (ADA), as well as the World Health Organization (WHO), were instrumental in defining criteria for the diagnosis of and selecting the terminology for what we now consider T1D and T2D (The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus 1997; Alberti 1998; American Diabetes Association 2010).

Beyond a lack in understanding the boundaries for age on T1D, another facet subject to considerable debate relates to T1D heterogeneity, both in terms of patient disease course as well as pathogenic mechanisms that underlie the disorder’s formation. As noted above, T1D is considered, with near uniformity, to represent a disorder “autoimmune” in nature—meaning that patients often express features reflective of an immunological contribution to their disease pathogenesis (e.g., autoantibodies, genes associated with immune-related genetic susceptibility, etc.). Yet, not all T1D patients possess these characteristics; leading some to the proposed classification of type 1A (autoimmune) diabetes (for review, see Eisenbarth 2007) for the 70%–90% of T1D patients having these immunological self-reactive properties, with type 1B (i.e., idiopathic) representing the remainder whose specific pathogenesis remains unclear (Imagawa et al. 2000). Interestingly, despite the proposal for adopting this new set of definitions for subgrouping T1D several years ago, few efforts do so today in terms of their descriptions of those with the disease (i.e., T1A and T1B diabetes are not commonly used terms in medical research publications).

Other potential factors of influence, either toward the age of T1D disease onset, its heterogeneity, or diagnosis, would include, but not be limited to, the growing problem of obesity (both childhood and adult) and health care provider recognition, as well as an increasingly diverse genetic admixture due to migration and/or social changes (Knip et al. 2005). These examples of influential changes likely reflect a series of additional variables (e.g., pathological studies of pancreata that are suggestive of various patterns of islet histology among supposed T1D patients, alterations in what comprises genetic susceptibility for the disorder) that require consideration when one ponders the pathogenesis as well as the clinical presentation of this disease.

Clearly, we have much to learn with respect to the question of, “What is T1D?” Answers to this question will likely come from improvements in our understanding of the pathogenesis and natural history of T1D, the focus of this article (as a whole) and other efforts within this collection (as individual points of perspective).

EPIDEMIOLOGY: INCIDENCE AND PREVALENCE

Irrespective of the aforementioned controversies, T1D is without question one of the most common chronic diseases of childhood (Karvonen et al. 2000; Gale 2005). Here too, a variety of epidemiological notions (i.e., dogmas) appear, at least on their surface, firm in their proposition, whereas other concepts are less sure.

First, it does appear that two peaks of T1D presentation occur in childhood and adolescence—one between 5 and 7 yr of age, with the other occurring at or near puberty (Harjutsalo et al. 2008). Beyond this, although many autoimmune disorders disproportionately affect women, T1D appears to affect males and females equally; but controversy does exist whether a
modest excess of T1D cases occurs in males in early age or signs of autoimmunity are increased with male gender (Weets et al. 2001; Krischer et al. 2004). In addition, the incidence of T1D varies as a function of seasonal changes, higher in autumn and winter and lower in the summer months (Moltchanova et al. 2009). The pathogenic mechanisms that underlie these seemingly sure observations are unclear, but, interestingly, recent studies assessing the development of T1D-associated autoimmunity (i.e., the formation of autoantibodies characteristic for the disease, discussed below) in the months to years before the onset of symptomatic T1D also show a degree of synchronization (Kukko et al. 2005), akin to the aforementioned seasonality, supporting a theoretical role for an environmental agent driving the pathogenesis of the disorder.

Less clear to this field of investigation is knowledge related to several variances that occur with respect to the incidence and prevalence of T1D, across both geographic populations as well as within different racial/ethnic groups. To begin, for years, data regarding the incidence and prevalence for T1D were far more predominant from studies performed in Europe rather than in the United States, where such information (in the latter instance) was largely dependent on extrapolation of data obtained from a small and limited number of somewhat localized studies (e.g., Alleghany County Pennsylvania, Colorado Diabetes Registry) rather than whole-country data (Kostraba et al. 1992a; Libman et al. 1998). However, this situation has been subject to recent improvement with the formation of the SEARCH for Diabetes in Youth Consortium within the United States (Dabelea et al. 2007, 2011), a multicenter study whose goals include identifying the number of children under the age of 20 with diabetes (either T1D or T2D), to understand the influence of race/ethnicity on the disease, and to address how T1D and T2D differ in this U.S. population (www.searchfordiabetes.org). In addition to this United States–based effort, years ago, at a global level, the WHO formed the Multinational Project for Childhood Diabetes known as the DIAMOND Project, an effort that followed the highly successful and often cited EURODIAB effort (EURODIAB ACE Study Group 2000; DIAMOND Project Group 2006).

Among the most significant findings the SEARCH effort has noted thus far, approximately 215,000 youth less than the age of 20 have diabetes (both T1D or T2D), representing ~0.26% of all people within this age group (Dabelea et al. 2007). During 2002–2005, 15,600 youth were diagnosed with T1D annually in the United States. Interestingly, among youth <10 yr in age, the rate of new T1D cases was 19.7 per 100,000 each year, whereas for those >10 yr of age, the rate was 18.6 cases per 100,000. In terms of ethnicity, SEARCH showed that non-Hispanic whites showed the highest rate of new onset T1D (24.8 per 100,000 per year among those <10yrofage) (Dabelea et al. 2007, 2011). Hence, now more than ever, and perhaps for the first time ever, accurate estimates can be provided as to the incidence rates in the United States versus other geographic populations.

At a global level, the incidence and prevalence rates for T1D are exceptionally interesting because they vary quite dramatically, with more than a 350-fold variation in incidence among reporting countries (Vandewalle et al. 1997; Patterson et al. 2009). Although clear exceptions to this rule exist, it does remain noteworthy that the incidence of T1D is positively related to distance north of the equator (i.e., the so-called North–South Gradient) (Karvonen et al. 2000). In terms of extremes, T1D is uncommon in China, India, and Venezuela, where the incidence is only 0.1 per 100,000 per year (for review, see Maahs et al. 2010). In contrast, the disorder is far more common in Finland, with recent incidence rates of more than 60 cases per 100,000 per year being noted, and to a slightly lesser degree, Sardinia, with rates approximating 40 per 100,000 per year. Rates of more than 20 cases per 100,000 per year are observed in Sweden, Norway, Portugal, Great Britain, Canada, and New Zealand (Vandewalle et al. 1997; Patterson et al. 2009; Maahs et al. 2010).

Interestingly, wide variations in incidence have been noted to occur between neighboring areas in both Europe and North America. For example, Estonia, separated from Finland by <75 miles, has a T1D incidence less than...
one-third that of Finland. Puerto Rico has an incidence similar to the mainland United States (i.e., 17 per 100,000 per year), whereas neighboring Cuba has an incidence of less than three cases per 100,000 per year. The mechanisms underlying these variances are unknown but, as is discussed below in this article, have largely been ascribed to the all-encompassing bin of “environment.”

Beyond current rates of T1D frequency, for reasons that remain unknown, the incidence of T1D has apparently been increasing throughout the world, for decades (Gale 2002b). For example, Sweden and Norway have reported a 3.3% annual increase in T1D rates, whereas Finland has observed a 2.4% annual rise in incidence (Thunander et al. 2008; Patterson et al. 2009), and to be clear, like examples exist across the globe. These increases have largely been ascribed to some unknown change in environmental constituents because notions of genetic alterations or improvements in delivery rates of offspring from T1D mothers could not in and of themselves explain these rates of increase (Herrmann et al. 2003; Soltesz et al. 2007). That said, after years of reports suggesting increases, at least one country, Sweden, has recently and quite unexpectedly noted that its incidence rates may have reached a “plateau” (Berhan et al. 2011). If confirmed in other populations, this would be cause for optimism because current incidence rates, if they were to continue on their existing path, would suggest a near doubling of T1D cases over the next decade (Patterson et al. 2009).

It is also important to note that these increases in incidence rates have not occurred equally across all age groups; that is, the most profound elevations in incidence rates have been observed in the youngest individuals (i.e., those <5 yr of age) (EURODIAB ACE Study Group 2000; DIAMOND Project Group 2006), as well as in young children from countries with historically high incidence rates (e.g., children <5–7 yr of age in Norway). Finally, T1D appears to have seen an increase in populations whose genetic susceptibility for the disease, in previous generations, would have been considered “lower.” Put another way, less genetic predisposition to T1D (i.e., class II alleles of the major histocompatibility complex, or MHC) appears to be required in order to develop the disease now, versus decades ago. This notion finds support with at least two studies, one in Europe, the other in the United States (Gillespie et al. 2004; Steck et al. 2011).

NATURAL HISTORY OF TYPE 1 DIABETES

Over the past three decades, the ability to understand the natural history of T1D has improved dramatically through the combined use of genetic, autoantibody, and metabolic markers of the disease (Atkinson 2003). Indeed, in the mid-1980s, a now oft-cited model was developed that attempted to integrate each of these three features (Eisenbarth 1986). This model for the natural history of T1D suggests that genetically susceptible individuals with a fixed number of β cells are exposed to a putative environmental trigger, which induces β-cell autoimmunity. This process, marked by the development of islet reactive autoantibodies, portends the development of activated autoreactive T cells capable of destroying β cells, resulting in a progressive and predictable loss in insulin secretory function. With this model, clinical (i.e., symptomatic) T1D does not present until >80%–90% of the β cells have been destroyed, and there is a marked gap between the onset of autoimmunity and the onset of diabetes.

Clearly, this model has served the community well over the years, providing a road map for investigations that have transformed our understanding of the natural history for this disease. However, recently, some aspects of the classical model have been modified to update knowledge gains (Fig. 1) (Atkinson 2001). For example, there are data to suggest that pancreatic β cells may persist in some individuals with T1D for an extended period of time (i.e., never reaching zero in many established T1D patients) (Meier et al. 2005). In addition, the degree of β-cell destruction required for symptomatic onset is also of growing question, with recent studies suggesting that 40%–50% β-cell viability may be present at the onset of hyperglycemia (Akirav et al. 2008), an aspect that may be related to subject age, among other factors (e.g., body...
mass index, physical activity, etc.) (for review, see Matveyenko 2008). This may explain why, despite persistent autoimmunity, insulin secretory function can remain stable for long periods of time in persons with T1D. That said, a loss of first-phase insulin response is usually followed by a period of glucose intolerance and a period of clinically “silent” diabetes (Sosenko et al. 2010). Finally, the “slope” reflective of β-cell loss in the pre-diabetic period has also recently been subject to considerable debate, with some proposing that the disorder may see its symptomatic onset only following a period of relapsing/remitting like autoimmunity (Fig. 2) (von Herrath et al. 2007).

An improved understanding of the natural history of pre-diabetes remains critical for directing future studies aimed at the prevention of T1D. Indeed, continued identification of genes controlling disease susceptibility, improved understanding of autoimmunity/mechanisms underlying loss of immune regulation, and further identification of environmental agents influencing the disease are all examples of information needed to impact efforts toward the goal of disease prevention; each is discussed below. Likewise, understanding events (e.g., rate of C-peptide loss, the presence of residual β cells, etc.) following symptomatic onset are also of importance because many ongoing efforts are actively seeking to reverse the disorder in those previously diagnosed with the disease.

Genetics

Many components related to the natural history and pathogenesis of T1D are reviewed in detail elsewhere within this collection, one facet being that of genetics. In short, despite being strongly influenced by genetic factors, T1D does not fit any simple pattern of inheritance and is considered a complex, multifactorial disease (see Noble and Erlich 2012). Early familial aggregation and twin studies supported the aforementioned importance for both genetic and environmental risk factors in T1D (Tattersall 1972), because individuals in the United States having a first-degree relative with T1D have an approximately 1 in 20 risk of developing T1D, whereas the general population of the United States have a one in 300 risk (Redondo et al. 2001). In addition, monozygotic twins have historically been considered to have a disease concordance rate of 30%–50%, with dizygotic twins having a concordance of 6%–10%. This said, one recent study suggests that were one to follow twins throughout their lifetimes, the percentage reaching concordance for T1D would come exceedingly close to being uniform (Redondo et al. 2008). All of this said, a strange curiosity...
Figure 2. Model for type 1 diabetes as a relapsing–remitting disease. (A) Graph showing the stepwise, nonlinear decline of β-cell mass over time, as well as the development of autoantibodies that are associated with hyperglycemia, that is, the onset of T1D. (B) The immunological response to T1D is cyclic. An increase in the numbers of autoreactive effector T cells is controlled by an increase in the number of regulatory T cells. However, over time, a gradual disequilibrium of the cyclical behavior could occur, leading to the number of autoreactive effector T cells surpassing the number of regulatory T cells, which would no longer be capable of containing autoreactive effector T-cell responses and thereby lead to a decline in pancreatic islet function. (C) β-Cell proliferation increases in a cyclical fashion over time. This figure indirectly depicts the biological trends of the development of T1D, which may be attributed to the cyclical nature of the immunological events that lead to the attack or protection of β cells. Such a phenomenon is usually the result of feedback-loop mechanisms, which, in the case of T1D, could be due to misdirected effector T cells that are not easily controlled by regulatory T cells. The inflammatory process of the pancreatic islets themselves may enhance β-cell proliferation and antigenic presentation, ultimately leading to the generation of more effector and regulatory T cells. In addition, as β-cell mass declines, the pressure on each β-cell to produce insulin increases, which may be sufficient to alter the recognition of β cells by the immune system and to alter their ability to regenerate and increase insulin production. (Adapted from von Herrath et al. 2007; with permission.)
remains that 85% of new T1D cases reside in individuals with no known family history for the disease (Hämäläinen and Knip 2002). Another interesting observation is that differences in risk are also dependent on which parent has diabetes—children of T1D mothers have only a 2% risk of developing T1D, whereas children of T1D fathers have a 7% risk (Redondo et al. 2001).

Following decades of effort to unravel the “enigma” of T1D genetics, nearly 50 loci have (thus far) been associated with susceptibility to the disease (Fig. 3) (Cooper et al. 2008; Concannon et al. 2009; Pociot et al. 2010). Nevertheless, no single gene is in-and-of-itself either necessary or sufficient to predict the development of T1D. The first T1D susceptibility locus identified, the Human Leukocyte Antigen (HLA) complex, provides the greatest contribution (i.e., ~60%) to the overall genetic susceptibility. There are three classes of HLA genes, with class II genes having the strongest association with T1D (Redondo et al. 2001). Because class II HLA genes encode for molecules that participate in antigen presentation, the effect of MHC allelic variability on T1D risk may, for example, be explained by differences in the presentation of β-cell antigens, either by promoting anti-self-reactivity or by the failure to impart regulated immune responses (Mallone et al. 2005). The great majority of T1D patients carry the HLA-DR3 or -DR4 class II antigens, with ~30% being DR3/DR4 heterozygous. In Caucasians, the DR3/DR4 genotype confers the highest T1D risk, followed by DR4 and DR3 homozygosity, respectively. Conversely, the class II allele, DQB1*0602, in linkage disequilibrium with DR2, is associated with protection from the development of T1D and is found in <1% of patients with T1D (Redondo et al. 2001).

Once one moves beyond HLA, the depth of genetic contributions to T1D becomes what some consider a notion of diminishing returns, at least at the level of individual odd ratios (OR) for disease risk. For example, the IDDM2 locus (i.e., the terminology used to define regions of the human genome providing susceptibility to T1D) has an odds ratio of ~6.8. On the x-axis are indicated possible candidate genes within genomic regions in which convincing associations with T1D have been reported. On the basis of the known functions of these candidate genes, the corresponding bars in the graph depicting odds ratios have been color-coded to suggest possible roles of these loci in susceptibility to T1D. At IL2RA and TNFAIP3, there is evidence of two independent effects on risk with different odds ratios; thus these loci both appear twice in the figure. An excellent resource for current information on all aspects of genes implicated in T1D is T1DBase (www.t1dbase.org). (Adapted from Concannon et al. 2009; with permission.)

**Figure 3.** Putative functions of non–HLA-associated loci in type 1 diabetes. The y-axis indicates the best estimate of the odds ratio for risk alleles at each of the indicated loci on the basis of currently published data. Although not shown, the HLA region has a predicted odds ratio of ~6.8. On the x-axis are indicated possible candidate genes within genomic regions in which convincing associations with T1D have been reported. On the basis of the known functions of these candidate genes, the corresponding bars in the graph depicting odds ratios have been color-coded to suggest possible roles of these loci in susceptibility to T1D. At IL2RA and TNFAIP3, there is evidence of two independent effects on risk with different odds ratios; thus these loci both appear twice in the figure. An excellent resource for current information on all aspects of genes implicated in T1D is T1DBase (www.t1dbase.org). (Adapted from Concannon et al. 2009; with permission.)
T1D) has been mapped to a variable number of tandem repeats (VNTR) region located upstream of the insulin gene. Disease association studies in case-control and family cohorts have shown that the number of tandem repeats is associated with T1D risk; Shorter repeats confer higher risk with longer repeats conferring lower risk (Pugliese et al. 1997). Yet, despite it being the genetic region having the second highest impact on disease development, its overall contribution is low (i.e., OR of 1.5). Another non-HLA gene associated with T1D is CTLA-4 (cytotoxic T lymphocyte associated-4). First identified in a family study of T1D (Nisticò et al. 1996), this gene encodes a molecule that plays an important role in the regulation of T-cell functionality and, hence, overall immune responsiveness, but once again its OR for T1D risk is low (Cooper et al. 2008). Other specific genes finding some degree of support for their influence on T1D include PTPN22 as well as CD25, each thought to provide influence to immune responsiveness (Concannon et al. 2009). Perhaps most interesting, however, and as pointed out by others (Concannon et al. 2009; Pociot et al. 2010), is that the vast majority of disease risk loci ascribed to human T1D are tied with functions related to immune responsiveness. Indeed, with this observation, we have likely entered a new era in the genetics of T1D in which genotype–phenotype studies will become the standard for investigation, in addition to entry into new and growing areas such as epigenetics, transcriptomics, and interactomics (Todd 2010). As our understanding of the function of susceptibility and resistance genes for T1D grows, we will continue to gain new insights into the relationship between genetic risk and the autoimmunity that culminates in the formation of this disorder.

Autoimmunity, Autoantibodies, and Cellular Immunity

As previously indicated, T1D is an autoimmune disease culminating in destruction of the pancreatic β cells, characterized histologically by insulitis (i.e., islet cell inflammation) and associated β-cell damage. Curiously, it remains unclear why the autoimmunity in T1D is specific to the insulin-producing β cells (Atkinson et al. 2011). Beyond this, the specific mechanisms responsible for inducing the autoimmunity in T1D also have yet to be elucidated (La Torre 2010).

Over the years, many different theories have been promulgated to explain this induction of β-cell autoimmunity including molecular mimicry (i.e., sharing of antigenic properties, including amino acid sequences between β cells and possible environmental agents) leading to the generation of an immune response, alteration of self-antigens to a now antigenic self, defective MHC expression on cells of the immune system, breakdown in central tolerance (i.e., a failure to establish immunity to self-antigens in early life), deleterious trafficking of dendritic cells from β-cells to pancreatic lymph nodes, sensitivity of the β cells to free-radical or cytokine-induced damage, the ever-elusive local viral infection, defects in peripheral tolerance (i.e., aberrant T-cell activation), and more (for review, see Bluestone 2010; Atkinson et al. 2011). Indeed, even the basic role for the cellular immune response, long thought key to the pathogenesis of T1D, has been remarkably controversial if not elusive (Roep 2003). Given that nearly all studies performed to date have involved immunological characterizations far from the site of tissue injury (i.e., analyses performed on cells obtained from peripheral blood and not isolated from pancreatic islets or lymphoid organs, including the pancreatic lymph node).

Regardless of the proposed cause(s) of the autoimmunity that results in T1D, pancreatic histology, either through postmortem examination or via biopsy, represents the only true means of directly showing β-cell injury (Atkinson 2009; In’t Veld 2011). To be clear, human insulitis is an elusive lesion, of which we know very little. One important reason relates to the limited number of cases that have been available for study (i.e., less than 150) (In’t Veld 2011). Of these, few have been studied in depth, and most lack the techniques of modern technologies. What evidence does exist suggests that insulitis is usually limited to β-cell-containing islets and that the facet disappears when these endocrine cells are absent, leaving a situation of so-called pseudo-atrophic islets devoid of β cells.
The inflammatory lesion within islets of those with T1D is typically characterized by a decrease (or absence) of insulin-producing β-cells along with a pancreatic islet cell infiltrate composed of T-lymphocytes, B-lymphocytes, macrophages, and lesser numbers of other cells representing the immune response (for review, see Foulis 2008). Because pancreatic biopsies have not, in most settings, been considered ethically feasible (for safety reasons) and autopsy tissue from subjects recently diagnosed with T1D is rare, major programs have recently been established to obtain these tissues for research. Most notable among such examples are the Belgium T1D registry, PEVNET (Finland), and the Network for Pancreatic Organ donors with Diabetes (In’t Veld et al. 2007; Oikarinen et al. 2008; Atkinson 2009). With time, it is anticipated that tissues obtained from these programs will provide major guidance toward questions related to the pathogenesis and natural history for this disorder.

With this limitation (i.e., lack of access to the pancreas), autoimmunity in T1D has typically been identified by the presence of autoantibodies to islet and/or β-cell antigens, which in addition to their presence at the time of diagnosis, can often be detected long before the disease becomes clinically evident (for review, see Ziegler 2010). Among a list of T1D-associated autoantibodies that actually has more than two dozen members are islet cell autoantibodies (ICAs), autoantibodies to glutamic acid decarboxylase (GADAs), insulin autoantibodies (IAAs), and autoantibodies to transmembrane tyrosine phosphatase (IA2As), as well as those against the ZnT8 molecule (ZnT8As). Although these are the five most prevalent and best characterized, the potential for other autoantibody/autoantigen combinations remains (Taplin 2008; Zhang et al. 2008).

However, it is critical to note that “times are changing” with respect to the potential role for autoantibodies, or their cellular source (i.e., B-lymphocytes), in the pathogenesis of T1D. For decades, the predominant dogma was that autoantibodies possessed no known etiological role in the disease and, simply put, were thought to represent the “smoke of the fire” in the pancreas and not the fire itself. However, recent studies in animal models of T1D purporting a crucial role for B-lymphocytes in disease development have opened the door for a previously unappreciated role for autoantibodies in the presentation of self-antigens to the cytotoxic T cells responsible for β-cell destruction (Maríno et al. 2011). This concept has also drawn support in human T1D studies in which therapeutic benefits were seen, over the short term, in recent-onset T1D patients treated with the B-lymphocyte-depleting agent anti-CD20 (rituximab) (Pescovitz et al. 2009). Hence, it is currently topical not only to portend the potential “diagnostic and predictive value” for B-lymphocytes and autoantibodies in T1D, but also to identify the role for this immune lineage in disease pathogenesis (Clynes 2010).

In terms of that diagnostic and predictive role, T1D-associated autoantibodies are typically present in 70%–80% of patients newly diagnosed with the disease (Bingley 2010), likely forming the aforementioned T1A population. In contrast, 0.5% of the general population and 3%–4% of relatives of patients with T1D are autoantibody-positive (Knip et al. 2010b). However, it is important to note that a wide variety of factors contribute to these percentages, including geographic population where studied, age and gender of the individual tested, race and ethnicity, quality and format for the autoantibody assay, and more (Tsirigiani et al. 2009; Bingley 2010). In stating that autoantibodies are surrogate measures for β-cell autoimmunity, autoantibody titer as well as the absolute number of autoantibodies (i.e., one, two, etc.) are both independent predictors of T1D risk (Skyler 2007). Specifically, when present at higher titers, at a younger age, or with the high-risk HLA genes, autoantibodies allow for a more accurate prediction of T1D risk. For example, one of the first demonstrations of this notion, developed years ago, was the observation that ICA titers of >40 Juvenile Diabetes Foundation (JDF) units carried a 60%–70% risk of developing T1D over the ensuing 5–7 yr (Schatz et al. 1994). Since that time, however, more and more studies (note: in fact, now
nearly all do so) use the so-called biochemical autoantibodies (i.e., GADAs, IAAs, IA2As, ZNT8As) for purposes of both diagnosis and prediction of T1D (Bingley 2010). Indeed, biochemical autoantibodies, when present in combination, increase the risk for T1D significantly. For example, in the large, NIH-funded Diabetes Prevention Trial–Type 1 (DPT-1), the 5-yr risk of T1D was 20%–25% for subjects with one autoantibody, 50%–60% for subjects with two autoantibodies, nearly 70% for those with three autoantibodies, and almost 80% for those with four autoantibodies (Winter 2011) (note: this study did include ICAs as part of its performance, in addition to biochemical autoantibodies). This ability to use autoantibodies for predicting future cases of T1D has also been supported by several large natural history trials including the NIH TrialNet and TEDDY efforts, as well as in a number of general population-based efforts (Miao et al. 2007; Orban et al. 2009; Knip et al. 2010b).

Although one would consider T1D-associated autoantibodies a relatively easy biomarker for studies of the disease, the reality for such a notion in everyday practice has often proved otherwise. For example, one peculiar aspect of IAAs is that they must be measured within 1 wk of the start of exogenous insulin therapy, because insulin antibodies (i.e., antibodies against therapeutic injected insulin) will also be detected and are indistinguishable from IAAs in current assay systems (Winter 2011). Similarly, although IAAs have shown themselves to be highly specific and sensitive for T1D, their assay (in terms of a methodology) has perhaps proven itself to be the most problematic of all of the biochemical autoantibodies, requiring high serum volumes and perhaps being composed of subsets having differing capacities for disease prediction (Bonifacio 2010). These have formed somewhat of a practical limitation in that IAAs have, in many ways, proved themselves the most important of the T1D-associated autoantibodies, at least in terms of predictive value. GADAs, like ICAs, are observed in 60%–70% of new-onset T1D patients (Bingley 2010; Winter 2011). But, unlike ICAs, GADAs have shown themselves to be the most predominant autoantibody in those with the aforementioned disorder LADA, perhaps forming a meaningful way to diagnose that disorder in adults, were it to find a true and meaningful basis (Leslie et al. 2008). IA-2 has an extracellular, transmembrane, and cytoplasmic domain, and autoantibodies to several forms of IA-2 have been observed in persons with T1D (Torii 2009; Bonifacio 2010). Here, some debate exists as to the significance of the various forms of autoantibodies to the IA-2 antigen and moreover, like GAD, which is expressed in many tissues including brain, pituitary, and pancreas, questions exist for the potential pathogenic significance for immune system constituents reactive with these entities.

The identification and description of autoantibodies in T1D have allowed us to gain remarkable insight into the natural history of this disease and, in fact, may comprise the greatest research success story in the 40 years of research investigations into the autoimmune nature of the disease (Nierras et al. 2010). In combination with a growing understanding of genetic susceptibility, autoantibodies allow for us to accurately predict which patients will develop T1D, from an early age (Fig. 4) (Bonifacio 2010; Ziegler 2010). Indeed, based on studies of large populations (both general population and families) for metabolic, genetic, and immune markers, levels of risk (i.e., low to very high) can be defined (Table 1). Nevertheless, until methods are developed that prove capable of preventing the development of T1D, it is likely that the utility for the use of these autoantibodies outside of research settings (i.e., in public health care settings) will be limited.

If autoantibodies have, for a major period of time, seen limited attention with respect to their pathogenic significance, cellular immune responses (both adaptive and innate) have been the focus of much in the way of research interest for their destructive potential. Briefly, studies of cellular immunity in T1D have been limited from a series of practical (i.e., assessments from peripheral blood) and technical (i.e., poor reproducibility, on most occasions) issues. Patients with T1D appear to have multiple defects in regulatory mechanisms that normally keep
autoreactive cells that escape negative selection within the thymus. Examples would include defective antigen-presenting cells, expansion of autoreactive T cells, and defective regulatory T cells (e.g., defective dendritic cell interactions, impaired iNKT cells, and resistance of so-called effector T cells to immune regulation). As a result of what appears to be a loss in fail-safe mechanisms at multiple levels, T1D occurs.

Environment

Although T1D knowledge gains have clearly occurred in each of the trio of factors considered responsible for the disorder’s development (i.e., genetic susceptibility, the immune system, and environment), the rate of intellectual progress, at least as might be defined by reasonably “firm” conclusions, has arguably been most limited in the later facet, that being environment. This is not to say that the environment is less important for studies on the pathogenesis and natural history of T1D because, clearly, disease discordance rates in twins, the rise in global incidence, variance in geographic prevalence, and rapid assimilation of local disease incidence rates when individuals migrate from low- to high-incidence countries all provide fundamental support to such a notion (Atkinson 2001).

Indeed, the field of T1D research is certainly not one devoid of hypothetical models to explain a role for the environment in its pathogenesis (Nerup et al. 1994; Wilkin 2001; Kukreja 2002; von Herrath et al. 2003; Bach 2005; Dahlquist 2006; Fourlanos et al. 2008; Vaara et al. 2008; Cooke 2009; Wasserfall et al. 2011). For example, the “accelerator” and “overload” hypotheses suggest that environmental stresses (specifically childhood obesity for the former) increase insulin demand, thereby overloading islet cells and accelerating β-cell damage (Wilkin 2001; Dahlquist 2006; Fourlanos et al. 2008). The “hygiene hypothesis” attributes the rising incidence of autoimmune disease in general to a reduced or altered stimulation of the immune system by environmental factors (Nerup et al. 1994; Kukreja 2002; Bach 2005; Cooke 2009). Conversely, the “fertile field hypothesis” proposes that microbial infection induces a temporary state in which other antigens can more easily react, yielding autoreactive T cells (von Herrath et al. 2003). Also implicating the gut, the “old friends hypothesis,” which is based on the role of normal gastrointestinal microbes, implicates dietary exposure as a
possible direct regulator of the immune system and of self-tolerance by altering gut microbiota and permeability (Vaarala et al. 2008). Very recently, the “threshold hypothesis” was put forward to provide a mathematical model for calculating T1D risk by considering the contributions of genetics and environment as a quantifiable function of invariables subject to calculation (Wasserfall et al. 2011). Although proposing hypotheses for T1D can be a difficult effort, proving them is even more complex.

Perhaps more numerous than the number of hypotheses that have been proposed to explain T1D are the number of environmental candidates thought to influence the disorder (Gale 2005). To be fair, a portion of the difficulties one encounters when attempting either to prove an environmental hypothesis for T1D or to identify environmental agents associated with the disorder can be attributed to the low disease risk conferred by genetic and environmental factors that have been identified thus far. Indeed, one underlying theme among all studies of environmental factors is that in the vast majority of a given population, even those with a combination of the highest-risk HLA haplotypes, most do not develop T1D. Beyond this are issues of variability for age at disease onset, a somewhat unpredictable natural history of disease, the propensity for most cases to develop in those without a family history for the disease, and more.

So, what are the environmental factors influencing T1D development? Historically, infectious agents have been the most frequently noted environmental influences for T1D (for review, see Boettler 2011). There is, however, no direct evidence that infection plays a role in the pathogenesis of this disease, albeit one agent (i.e., rubella) is often, and incorrectly, cited as evidence for this activity (Gale 2008). Beyond rubella infection, a relationship between β-cell autoimmunity and enteroviral infections has long been reported in association with the disease (for review, see Jäidane et al. 2010). This association has seen its basis at many levels: increased frequency of anti-enteroviral antibodies; immunohistochemistry of enteroviral detection, or viral RNA, in the pancreas of those with T1D (i.e., anti-VP1 staining); elevations of enteroviral RNA in peripheral blood from those with T1D; and more (for review, see Hober 2010; Taurainen et al. 2011). This association has seen its basis at many levels: increased frequency of anti-enteroviral antibodies; immunohistochemistry of enteroviral detection, or viral RNA, in the pancreas of those with T1D (i.e., anti-VP1 staining); elevations of enteroviral RNA in peripheral blood from those with T1D; and more (for review, see Hober 2010; Taurainen et al. 2011). However, the presence of antibodies against enteroviruses in those with T1D, even if true (i.e., such findings have been highly controversial), does not prove a causal relationship. For example, persons with autoimmunity may also be

### Table 1. Type 1 diabetes risk stratification by T1D family history and HLA genotyping

<table>
<thead>
<tr>
<th>Population</th>
<th>T1D risk (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk</td>
<td></td>
</tr>
<tr>
<td>No affected FDR plus HLA protective genes</td>
<td>0.01</td>
</tr>
<tr>
<td>No affected FDR</td>
<td>0.4</td>
</tr>
<tr>
<td>Affected FDR plus HLA protective genes</td>
<td>0.3</td>
</tr>
<tr>
<td>Intermediate risk</td>
<td></td>
</tr>
<tr>
<td>No affected FDR plus HLA risk genes</td>
<td>4</td>
</tr>
<tr>
<td>One affected FDR</td>
<td>5</td>
</tr>
<tr>
<td>Mother with T1D</td>
<td>3</td>
</tr>
<tr>
<td>Father with T1D</td>
<td>5</td>
</tr>
<tr>
<td>Sibling with T1D</td>
<td>8</td>
</tr>
<tr>
<td>High risk</td>
<td></td>
</tr>
<tr>
<td>One affected FDR plus HLA high-risk genes</td>
<td>10–20</td>
</tr>
<tr>
<td>Multiple affected FDRs</td>
<td>20–25</td>
</tr>
<tr>
<td>Very high risk</td>
<td></td>
</tr>
<tr>
<td>Identical twin affected</td>
<td>30–70 (more?)</td>
</tr>
<tr>
<td>Multiple affected FDRs plus HLA risk genes</td>
<td>50</td>
</tr>
<tr>
<td>Sibling affected plus HLA risk genes, identical by descent</td>
<td>30–70</td>
</tr>
</tbody>
</table>

Adapted from Ziegler and Nepom 2010; with permission.

Increase in T1D risk is associated with progression of islet autoantibodies from single to multiple autoantibodies. Characteristics of the initial antibody response can help predict disease progression.

Abbreviations: IAA, insulin autoantibodies; GAD65, glutamic acid decarboxylase 65; IA-2, islet-associated autoantibody 2; ZnT8, zinc transporter 8; FDR, first-degree relative; HLA risk genes, HLA DQB1*0302; HLA protective genes, HLA DQB1*0602.
more prone to enteroviral infection, may have a stronger humoral response to infection because of their particular HLA genotype, or may be in a nonspecific hyperimmune state marked by elevation of antibody levels to a variety of exogenous antigens. Beyond this, questions regarding the specificity of these antibodies used to detect virus in pancreatic sections, for patients with T1D, have come under increased scrutiny (Richardson et al. 2009). As a result, much more effort will be required to link this virus or any of several other candidate viruses (e.g., rotavirus, cytomegalovirus, etc.) to the pathogenesis of this disease.

Perhaps second to viruses, nutritional influences have also often been considered in association with T1D, perhaps the most predominant being the association between the effect of breastfeeding and/or early exposure to cow’s milk on the incidence of autoimmunity and the disorder, albeit here too, the notion remains highly controversial. In support of the notion, a major meta-analysis showed a weak but statistically significant association (OR 1.5) between T1D and both a shortened period of breastfeeding and cow’s milk exposure before 3–4 mo of age (Gerstein 1994). Based on this notion, a large and well-organized effort (i.e., the Trial to Reduce IDDM in the Genetically at Risk, TRIGR) has been formed to test whether cow milk avoidance reduces T1D risk (TRIGR Study Group et al. 2011), with early evidence in terms of avoiding autoantibody markers in children on specialized infant diets appearing as promising (Knip et al. 2010a). In terms of the molecular mechanisms that underlie this association, a variety of constituents in either cow or breast milk, ranging from casein to bovine insulin, as well as bovine serum albumin (BSA), have each been subject to multiple reports touting their implication for T1D development (Borch-Johnsen et al. 1984; Karjalainen et al. 1992; Vaarala 2005; Luopajärvi et al. 2008). However, with time, these associations have either not been subject to replication by others or firm evidence for their support remains lacking. For example, no association between early exposure to cow’s milk and β-cell autoimmunity in young siblings and offspring of T1D patients has been shown in multiple natural history studies, increased breastfeeding in developed countries is inconsistent with a rising incidence of childhood diabetes, and attempts to link immunity to BSA with T1D has not been subject to marked replication in the research community (Atkinson et al. 1993; Norris et al. 1996). This said, very recently, cross-reactivity between the β-cell-specific protein (insulin) and bovine α-casein has been noted and holds interesting potential for molecular mimicry (Adler et al. 2011).

The ingestion of nutrients containing elements of plants also appears to have an effect on the development of T1D, because two studies (i.e., the Diabetes Autoimmunity Study in the Young [DAISY] and the German study of offspring of T1D parents [BABY-DIAB]) provided evidence that susceptibility to T1D is associated with the timing of exposure to cereal and gluten (Norris et al. 2003; Ziegler et al. 2003). Although both studies provide interesting findings, their conclusions were in some ways contradictory and show the need for larger collaborative investigations in order to appropriately determine how early dietary exposures affect risk for autoimmunity (Atkinson 2003). Other dietary targets that may be of heightened immunogenicity, but whose associations to T1D remain somewhat in their infancy, include wheat storage globulin, gluten/gliadin, and casein, among others (for review, see Lefebvre 2006).

As previously discussed, the highest incidence of T1D worldwide occurs in northern Europe, leading some to suggest that low serum concentrations of vitamin D may not only be associated with T1D but perhaps cause for development of the disease (for review, see Todd 2010). To this end, data exist suggesting that newly diagnosed T1D subjects had lower serum concentrations of this metabolite than healthy controls (Pozzilli et al. 2005; Littorin et al. 2006; Svoren et al. 2009). In addition, polymorphisms in the vitamin D metabolism gene have recently been implicated with this disorder (Bailey et al. 2007). The concept is an interesting one because the aforementioned “North–South Gradient Hypothesis” (Karvonen 2000) would be consistent with a notion that the amount of UV-B
exposure (a factor influencing the synthesis of vitamin D) could modulate this metabolite, one that is closely linked with immune responses. That said, not all reports find such a disease association, with at least one (Bierschenk et al. 2009) suggesting that at a population-based level, most individuals (including those with T1D) are either vitamin D deficient or insufficient. Thus, although of interest, a clear cause–effect relationship between vitamin D and T1D remains to be identified.

Beyond these, associations between specific agents with T1D become much rarer. Toxic doses of nitrosamine compounds can also cause diabetes through the generation of free radicals, but the effect of dietary nitrate, nitrite, or nitrosamine exposure on human T1D risk remains unclear (Kostraba et al. 1992b). Several perinatal risk factors for childhood diabetes are also associated with the development of T1D (Dahlquist et al. 1999). For example, the effect of maternal–child blood group incompatibility is fairly strong (both ABO and Rh factor, with ABO > Rh). Other perinatal factors conferring increased risk include pre-eclampsia, neonatal respiratory distress, neonatal infections, Caesarian section, birth weight, gestational age, birth order, and maternal age (Flood et al. 1982; Blom et al. 1989; Patterson et al. 1994; McKinney et al. 1997). In the end, it will be important to determine whether these factors ultimately contribute to T1D themselves, or how they may be confounded by other unknown risk factors.

Finally, it has also been argued that the rising incidence of T1D could be accounted for by protective factors in the environment that have been lost (Todd 1991). In support of this, there has been a parallel rise in the rates of asthma and allergy to that in T1D. The aforementioned hygiene hypothesis proposes that early exposure to infective agents in early childhood is necessary for maturation of the neonatal immune response. In the absence of such exposure, the model states this allows for a failure of early immune regulation that may permit, depending on genetic susceptibility, the development of autoimmunity (i.e., Th1) or allergic (i.e., Th2) disease (Holt 2000). This model is consistent with the concept that T1D is less likely to develop in the presence of environmental factors eliciting strong Th2-like immunity (e.g., pinworms and other infections), yet specific evidence for the hygiene hypothesis in human T1D is minimal, but attractive (Gale 2002a).

In sum, the list of potential environmental triggers and regulators of disease in T1D remains considerable. It stands likely that only through continued efforts within large, prospective, multicenter screening programs will specific environmental factors (and the influence of genetic and immunologic factors on them) truly associated with the development of the disease be identified. In addition, the complexity of these efforts must go beyond the simplicity of previous efforts and delve into areas not subject to much in the way of previous investigation (e.g., antibiotic use, fever, exposure to environmental toxins, etc.), throughout pregnancy and into early infancy. Indeed, one more aspect does appear clear with respect to environment: T1D-associated autoantibodies are often, but not exclusively, observed within the first 2–3 yr of life in many developing the disease (Bonifacio 2010), supporting the notion that environmental agents could operate early in disease pathogenesis. This (and there needs to be an increased appreciation that environmental factors) likely interacts with genetic factors, affording either susceptibility or resistance to the disease (Fig. 1), resulting in a modulation in the rate of T1D development and not merely that of initiators of the disease (Hermann et al. 2003).

CONCLUDING REMARKS

Much has been learned in the last 40 years regarding the pathogenesis and natural history of T1D. That said, a major motivation driving research efforts in these areas was a belief that such gains would result in a means to prevent as well as to reverse the disease (Skyler 2011). Sadly, despite the performance of an impressive number of clinical trials—ranging from small pilot efforts to large, multicenter consortium-based efforts—no means has been identified that meets this purpose, especially one applicable to a public heath care setting. Because of this,
future efforts will likely benefit from continuing improvements in knowledge related to the question of how T1D develops.

ACKNOWLEDGMENTS

The thoughts put forward in this work have been collected over many years with the gracious support of the National Institutes of Health, the Juvenile Diabetes Research Foundation, the American Diabetes Association, the Brehm Coalition for Type 1 Diabetes Research, and the Jeffrey Keene Family Professorship.

REFERENCES


Pathogenesis and Natural History of T1D
W. A. Atkinson


Cite this article as Cold Spring Harb Perspect Med 2012;2:a007641
Pathogenesis and Natural History of T1D


M.A. Atkinson


The Pathogenesis and Natural History of Type 1 Diabetes

Mark A. Atkinson

Cold Spring Harb Perspect Med 2012; doi: 10.1101/cshperspect.a007641

Subject Collection Type I Diabetes

The Pathogenesis and Natural History of Type 1 Diabetes
Mark A. Atkinson

Humoral Autoimmunity in Type 1 Diabetes: Prediction, Significance, and Detection of Distinct Disease Subtypes
Massimo Pietropaolo, Roberto Towns and George S. Eisenbarth

Endoplasmic Reticulum Stress, Pancreatic β-Cell Degeneration, and Diabetes
Feroz R. Papa

Islet Autoantigens: Structure, Function, Localization, and Regulation
Peter Arvan, Massimo Pietropaolo, David Ostrov, et al.

Environmental Triggers of Type 1 Diabetes
Mikael Knip and Olli Simell

Generating β Cells from Stem Cells—The Story So Far
Matthias Hebrok

Antigen Targets of Type 1 Diabetes Autoimmunity
Bart O. Roep and Mark Peakman

For additional articles in this collection, see http://perspectivesinmedicine.cshlp.org/cgi/collection/

Copyright © 2012 Cold Spring Harbor Laboratory Press; all rights reserved