Rheumatoid arthritis (RA) has been described in the medical literature for over two hundred years, but its etiology remains unknown. RA displays phenotypic heterogeneity, and it is a relatively prevalent clinical entity: it affects approximately 1% of the population, resulting in enormous pathologic sequelae. Earlier studies targeting the cause(s) of RA suggested potential infectious involvement, whereas more recent reports have focused on a genetic origin of the disease. However, neither infection nor genetics, nor any other single factor is currently accepted as causative of RA. In this article we review studies relating to the etiology of RA, and those of several related matters, and we conclude that the literature indeed does provide insight into the causes underlying the initiation of RA pathogenesis. Briefly, given the remarkable phenotypic variation of RA, especially in its early stages, as well as a number of other characteristics of the condition, we contend that RA is not a discrete clinical entity with a single etiological source. Rather, we argue that it represents a common clinical endpoint for various starting points, each of which is largely guided by as yet poorly understood aspects of the genetic background of the affected individual. Adoption of this alternative view of the origin of RA will have significant consequences for future research and for development of new therapeutic interventions for this burdensome condition.

**Keywords:** rheumatoid arthritis pathogenesis etiologic studies genetic background

**Abstract:** Rheumatoid arthritis (RA) has been described in the medical literature for over two hundred years, but its etiology remains unknown. RA displays phenotypic heterogeneity, and it is a relatively prevalent clinical entity: it affects approximately 1% of the population, resulting in enormous pathologic sequelae. Earlier studies targeting the cause(s) of RA suggested potential infectious involvement, whereas more recent reports have focused on a genetic origin of the disease. However, neither infection nor genetics, nor any other single factor is currently accepted as causative of RA. In this article we review studies relating to the etiology of RA, and those of several related matters, and we conclude that the literature indeed does provide insight into the causes underlying the initiation of RA pathogenesis. Briefly, given the remarkable phenotypic variation of RA, especially in its early stages, as well as a number of other characteristics of the condition, we contend that RA is not a discrete clinical entity with a single etiological source. Rather, we argue that it represents a common clinical endpoint for various starting points, each of which is largely guided by as yet poorly understood aspects of the genetic background of the affected individual. Adoption of this alternative view of the origin of RA will have significant consequences for future research and for development of new therapeutic interventions for this burdensome condition.

**Keywords:** rheumatoid arthritis pathogenesis etiologic studies genetic background

**Introduction**

In company with tuberculosis, trachoma, and a few other diseases, the rheumatic diseases are among the oldest known afflictions of mankind. Importantly, in the latter group rheumatoid arthritis (RA) ranks as perhaps the earliest to be clearly described and documented by physicians. Indeed, the first description of RA in what can be considered the modern clinical literature was published by Landre-Beauvais in 1800; the clinical designation “rheumatoid arthritis” was coined by Garrod in 1859. However, the disease itself is unquestionably far older than its official clinical description. For just one example, some researchers speculate that clinical manifestations of RA are present in people depicted in 16th century paintings, which show finger and hand abnormalities resembling those characteristic of RA. Other data suggest that RA has existed in North America for 3000 years or more, and it is not at all surprising that a variety of rheumatic diseases have been misclassified and misdiagnosed during recorded medical history because of the substantial overlap in clinical signs and symptoms among many of them.

Recent estimates indicate that approximately three million individuals in the United States suffer with RA, with similarly large numbers of patients diagnosed in Europe.
Without question, RA imposes a crushing burden on those it afflicts. Indeed, recent epidemiologic studies indicate that more than half of all patents with RA are not able to work ten years after initial diagnosis. This clinical entity thus imposes a significant burden on the families of RA patients, as well as on the patients themselves. With an estimated 3.6 billion US dollars spent annually on treatment for RA patients, the condition further imposes a major burden on the health care system and thus on society in general. For all these reasons, RA has been the subject of intensive research for many decades, but after more than two centuries of investigation the etiology of RA remains elusive.

As reviewed below, some studies have suggested an infectious origin for RA, with more recent reports indicating, in contrast, that the pathogenesis characteristic of the condition is of genetic origin. However, neither infection nor genetics, nor any other single factor currently is accepted as causative in RA. A question of moment therefore is: why have decades of intensive research failed in such a singular fashion to identify the causative factor or factors that engender this important clinical entity? In this article we re-examine studies targeting causality and related aspects for RA. On the basis of results from those studies, we contend that compelling evidence regarding the etiology of this clinical entity is, in fact, present in the literature, but that it has been misunderstood and misinterpreted as a result of the inappropriate application of paradigms underlying current clinical and biological research and thinking. The ultimate, and in our view inescapable, conclusion from the detailed review of the literature which follows is that RA is not a discrete disease with a single, unique etiologic source. Rather, as we have argued in the context of multiple sclerosis, we contend that the condition currently diagnosed as RA is in reality a common endpoint for a number of diverse starting points, rather than a single unique disease. As developed below, this new view of the origin of RA pathogenesis has significant consequences for future research, as well as for the development of new therapeutic interventions to treat RA.

**Clinical presentation and diagnosis of rheumatoid arthritis**

Before reviewing currently available information regarding etiology, it is useful to summarize relevant clinical and other aspects of RA so as to provide a thorough foundation for a reassessment of its origins. RA is considered to be a chronic progressive systemic condition that affects mainly diarthrodial joints. The most common, but by no means universal, mode of onset involves symmetrical pain and swelling in small joints of hands and feet, and its inception can and does vary widely among individuals. RA may begin with marked systemic symptoms including fatigue, fever, and weight loss, and slowly over weeks to months present the more classic symptoms of joint pain and swelling. A palindromic pattern of onset, commonly referred to as palindromic rheumatism, also can occur. However, this type of presentation is uncommon; it consists of recurrent episodes primarily of monoarthritis, with abrupt joint pain and swelling in one joint at a time that comes on quickly, then abates within one or two days. Importantly for the purposes of what follows below, only about 50% of patients presenting with palindromic rheumatism will actually go on to develop full RA. Fewer than half of all RA patients demonstrate extra-articular symptoms including rheumatoid nodules, which are subcutaneous granulomatous lesions with central fibrinoid necrosis surrounded by a layer of histiocytes. Some common extra-articular manifestations of RA include parenchymal lung disease, secondary Sjögren’s syndrome, cutaneous vasculitis, and pericarditis.

Criteria for the classification and diagnosis of RA were developed by the American College of Rheumatology (ACR) in 1987 and to date remain the standard guidelines employed for diagnosing this condition. According to these criteria, RA is diagnosed if a given patient satisfies at least four of seven clinical criteria, and the initial four must have been present in the patient for at least 6 weeks; this latter specification is intended to exclude the several types of virally induced arthritis that can resemble RA. Studies evaluating the success and sensitivity of these criteria in normal clinical situations show them to be reliable and responsive, but the goal is to avoid confusing RA with other conditions. Such confusion is a significant diagnostic possibility due to the often severe systemic effects seen in RA and importantly, its significant overlap in clinical signs and symptoms with other rheumatic diseases. The primary therapeutic strategy is to detect the relevant synovial and other pathogenic processes early in their development, and to treat quickly and aggressively to prevent long term joint damage. A full metabolic panel and complete blood count with differential and inflammatory biomarkers (erythrocyte sedimentation rate [ESR]/C-reactive protein) are often done, and they often are elevated in patients with RA. Two antibodies, rheumatoid factor (RF) and anticyclic citrullinated peptide (anti-CCP) also appear elevated and accompany the diagnostic protocol; however RF, and to a lesser extent anti-CCP (see below), are sometimes seen in other disease states or even in healthy...
individuals, and thus these may not serve as completely specific markers for RA in each and every case.\textsuperscript{11-13}

The critical point is that, while the pathogenic characteristics of patients with established RA are somewhat more uniform, the clinical picture of RA in its early stages is extremely diverse and heterogeneous, and an RA patient’s prognosis depends on the ability of the treating physician to make a rapid and accurate diagnosis and to implement treatment promptly. This is often a difficult task, since many patients present with divergent symptoms which manifest at different rates. The ACR criteria, which are essentially the only diagnostic criteria available, must be as specific to RA as possible. This has proved to be somewhat impractical since causation has not been identified, and few diagnostic markers exist which are completely and reliably unique to RA.

**Epidemiology and pathogenesis of rheumatoid arthritis**

Epidemiologic data indicate that the prevalence of RA in North America is between 0.5% and 1%.\textsuperscript{3} The prevalence rate is 2%–3% for first degree relatives, and genetic concordance in monozygotic twins is approximately 15%–20%.\textsuperscript{14} Patients with RA have a considerably increased risk of death compared with age- and sex-matched individuals without RA. This clinical entity is equally prevalent in Europe and many other geographic regions, although interestingly, recent studies indicate that Japan and China display a lower rate of 0.2%–0.3%.\textsuperscript{15} RA is more prevalent, at a rate of greater than 5%, in American Indian and Alaska natives, suggesting some genetic bias which has yet to be fully identified. RA shows little geographical, seasonal, or other variation, although a recent study by Ikuni and colleagues in Tokyo indicated a distinct seasonal difference in this region, with disease activity higher in the Spring and lower in the Fall.\textsuperscript{16}

RA can occur in individuals of virtually every age group, although the age of onset for a majority of the population is between 40–70 years. A US National Health Examination Survey (1960–1962) performed by Engel and colleagues\textsuperscript{17} showed the incidence of RA to be only 0.3% in adults younger than 35 years. Although it is rare, RA also affects children in the form of juvenile RA,\textsuperscript{18} now termed juvenile idiopathic arthritis. RA displays little or no important ethnic or racial bias overall, although studies of Hispanic and African American populations and some other groups remain rather limited.\textsuperscript{19,20} A strong gender bias does exist, of course, with women displaying a two- to threefold higher prevalence of RA than do men. An explanation for this strong bias towards females has been investigated extensively but with only limited success, an issue discussed in detail below. A number of behavioral, dietary, and other external factors have been associated with increased/ decreased risk for RA, and these also are discussed later in this article. Thus, overall the available epidemiologic data clearly demonstrate that RA is anything but rare in the US and elsewhere, but those data by themselves do not provide significant insight into the causative factors underlying the initiation of pathogenesis.

The pathogenesis of RA remains somewhat unclear, which also contributes to the relative obfuscation of causation underlying the initiation of the process. However, strong clinical and research emphasis has been placed on inflammation, and on the role of autoantibodies and immune complexes in initiation of the pathogenic process. Pathogenesis begins with inflammation of the synovium in small diarthrodial joints in the hands and feet, and in some larger weight bearing joints. The initial inflammation leads to progressive bone and cartilage damage, which in time destroys joint function. In a healthy joint, a thin synovial membrane containing two basic cell types lines the joint.\textsuperscript{21} One of these types is the synoviomacrophage, which possesses antigen presenting ability; the second cell type, fibroblast-like synoviocytes, produce a specialized matrix and synovial fluid. In RA, increased cellular proliferation and decreased apoptosis is often observed in this layer, and this results in hyperplasia and inflammation. Data in a recent article demonstrated that inhibition of synovial leukocyte apoptosis in the early phases of RA pathogenesis actually may distinguish RA from other arthritides.\textsuperscript{22} Findings such as this that demonstrate a difference between RA and other rheumatic diseases still are quite limited, and more study is required to improve and refine diagnostic criteria.

The highly proliferative tissue lining is called the pannus, and it persists in the joints of RA patients due to the expression of anti-apoptotic cytokines, two examples of which are interleukin 2 (IL-2), and IL-4, and a lack of tumor suppressor gene function.\textsuperscript{23} Synoviomacrophages and fibroblast-like synoviocytes expand and infiltrate the surrounding cartilage and bone; these cells are the primary source of inflammatory mediators, including proinflammatory cytokines, chemokines, and adhesion molecules. The cytokine tumor necrosis factor alpha (TNFα) is present in the RA synovium, and anti-TNFα therapy has proved to be effective in ameliorating disease progression in many RA patients.\textsuperscript{24,25} TNFα can influence other immune mediators, and thus blockade of this molecule in the RA synovium by biological modifiers has wide-ranging effects.\textsuperscript{26-30} Cytokine expression also...
activates chondrocytes, cells present in cartilage that produce cartilaginous matrix, and osteoclasts, which are involved in bone breakdown and reabsorption. Osteoclasts secrete acid proteases (eg, phosphatases, cathepsin K) that dissolve bone minerals and degrade bone matrix, resulting in erosion and ultimately loss of joint function.31

Although RA pathogenesis includes a significant inflammatory aspect, an important autoimmune component is also characteristic. One possible explanation for the latter resides in the idea that populations of T cells which recognize self tissue escape clonal deletion in the thymus and thereafter are available to react with self peptides in the RA joint. Autoreactive T cells are known to persist in normal individuals, and under the appropriate stimulus such autoreactive cells can initiate antiself-immune responses.32 Further, pathogen peptides which have a similar sequence and/or structure to host peptides may be expressed in the synovia of some RA patients where they are recognized by autoreactive T cells.33 This concept is designated molecular mimicry, and it currently has strong influence in some research contexts relevant to RA; molecular mimicry will be discussed in detail below in relation to one aspect of RA etiology.

Malfunctioning B cells also can produce antibodies directed against host peptides. For example RF, an antibody directed against the Fe portion of immunoglobulin G (IgG) molecules, is present in approximately two thirds of patients with RA.12 Although the presence of RF is used for diagnostic purposes, 10% of healthy individuals and many patients with other autoimmune disorders, including Sjögren’s syndrome, systemic lupus erythematosus (SLE), and mixed connective tissue disease, also express RF.13,34 In addition, approximately 70% of patients with disease due to chronic hepatitis C virus are RF positive.35 The diagnostic protocol for RA has included laboratory testing for RF for more than 50 years, but the presence of RF as a marker for RA in patients is considered unreliable due to its presence in the general population and in other autoimmune and infectious diseases. Decline in RF levels may be useful, however, as an indicator of a given patient’s response to treatment when various disease-modifying antirheumatic drugs and other biologics, such as infliximab or rituximab, are administered.36

Interestingly, recent studies suggest that autoantibodies to citrullinated proteins provide a useful and quite specific diagnostic indicator for RA. The anti-CCP antibody is present in approximately 80% of RA patients and is quite specific for the disorder.11 Moreover, this autoantibody is detected in less than 1% of healthy individuals; it can even appear before RA is clinically evident and detectable. Studies indicate that citrullination of possible autoantigen peptides can take place in synovial fluid,37 and that anti-CCP antibodies contribute importantly to the maintenance of joint pathogenesis in patients with established RA who are positive for them.38,39 Citrullination is a post-translational modification involving the deamination of arginine residues by the enzyme peptidylarginine deiminase (PAD). Hydrolysis of the amino group (NH2) of arginine to a neutral oxygen group by PAD results in the formation of the amino acid citrulline containing the neutral oxygen. This neutral oxygen is the moiety recognized by the autoantibody. Many protein antigens can undergo citrullination; for example, both the alpha and beta chains of fibrin in the joint have been shown to undergo citrullination, and autoantibodies that recognize and react with citrullinated fibrin have been isolated from RA patients.40 Interestingly, increased citrullination of myelin basic protein has also been reported in MS.39,41

Autoimmunity as a component of rheumatoid arthritis pathogenesis

As indicated above, RA includes an important autoimmune component, but whether a loss of tolerance to self tissue actually comprises a primary causative factor in RA pathogenesis remains to be established. T and B lymphocytes, dendritic cells (DC), natural killer cells, and polymorphonuclear leukocytes all are massed in the joint lining of RA patients, and they play an obvious role in chronic inflammation. T cell involvement in RA pathogenesis has been studied at length but remains unclear. Early in life T cells reactive to host tissue normally are deleted in the thymus, thereby preventing any immune response to self. However, T cells can escape deletion and persist in normal individuals.32 CD4+ T cells have been extensively studied, especially those that produce interferon gamma (IFNγ), and these are the dominant T cell subtype found in RA joints.42 The cytokine environment controls the differentiation of T cells into two additional effector types, T helper cells (Th), Th1 and Th2. Th1 is the main phenotype observed in autoimmune diseases including RA, but new T cell subsets have emerged.43

Th17 and T regulatory (Treg) T cells both have been documented in the synovia of RA patients, and currently available data indicate that both can contribute to RA pathogenesis.42–44 A recently published report by Shen and colleagues45 shows an increase in the frequency of Th17 cells, resulting in an increase in IL-17 secretion from peripheral blood monocytes from RA patients. IL-17 plays a crucial role in mediating the inflammatory aspect, an important autoimmune component.
response and may control osteoclastogenesis. Findings by Kirkham and colleagues support a role for IL-17 in RA; synovial membrane IL-17 mRNA levels may be predictive of joint injury and damage progression. Other important cytokines, such as IL-1β and of course TNFα also are prognostic for joint damage and are found at elevated levels in RA joints. Interestingly IL-10, a cytokine considered to be anti-inflammatory in its actions, also was shown to be produced at high level in the RA synovium in one study, and it too may be thus predictive for joint damage; in this context, IL-10 may be a surrogate marker for some other as yet undefined mediator or, more interestingly, may be involved directly or indirectly in increasing production of anti-CCP antibodies. More research is needed to address this question.

Treg cells normally suppress inflammation by inhibiting proliferation and cytokine production from T cells. Treg cells from the synovia of RA patients appear unable to suppress cytokine production. They are similarly unable to control T cell activation, which may engender a breakdown in tolerance. Interestingly, Treg cell function is restored in many patients following anti-TNFα therapy, which may explain in part the therapeutic benefit of these biologic modifiers. While some researchers and clinicians argue that the intense rise in levels of mediators of inflammation such as TNFα and IL-10 is a consequence of malfunctioning Treg cells, few studies examining the role of these cells in RA have been published; thus, firm conclusions cannot be drawn at this point.

The T cell has been the focus of autoimmune research for decades, but use of the anti-CD20 antibody (rituximab), which is B cell-depleting, has focused new attention on that cell type. B cells produce antibodies and can, through the activation of complement, clear many pathogens. They secrete a variety of cytokines, express major histocompatibility (MHC) molecules on their surfaces, act as antigen presenting cells, and are largely functionally controlled by T cells. B cells may become autoreactive by escaping deletion in a similar fashion to T cells, but that process for B cells takes place in bone marrow. Normally, self-reactive B cells which are not deleted are sustained in an inactive state (anergy) by T helper signals; thus a disruption in T cell function could liberate the suppression of autoreactive B cells and lead to destruction of self tissue. As discussed above, activation of self-reactive B cells by T cells also stimulates antibody production (eg, RF and/or anti-CCP) directed against self peptides. In addition, B cells are capable of producing cytokines such as IL-2 and IFNγ, and they thus are able to stimulate the differentiation of naïve T cells into activated the Th1 and Th2 subtypes.

Much of what we currently understand regarding T and B cell involvement in RA pathogenesis has emerged from the successful use of biological modifiers in treating the condition. For one example, abatacept, which is a selective co-stimulatory modulator and inhibitor of T cells, often improves RA symptoms, and on this basis many researchers conjecture that T cells are primarily responsible for RA pathogenesis. We note, however, that the therapeutic agents in current use for treatment of patients with RA, including those targeting TNFα or its receptor, lack universal efficacy, indicating yet again significant heterogeneity in the detailed pathogenic processes among those patients.

Studies targeting the etiology of rheumatoid arthritis: infection

The idea that microbial species, including bacteria, viruses, and parasites, represent major environmental stimuli that can initiate and sustain antiself-immune responses in genetically susceptible individuals has long been, and continues to be, a central force in autoimmune disease research. Thus, the idea that infectious agents may play an etiologic role in the initiation and maintenance of RA pathogenesis has been promulgated extensively over the years. In some articles, the case is made for causation by a single organism, while in others multiple organisms are argued to be etiologic for RA (see Table 1). As developed below, we favor the multiple organism hypothesis, although in a rather different

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Organisms associated with the etiology of rheumatoid arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteria</strong></td>
<td><strong>Escherichia coli</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Klebsiella pneumoniae</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Mycobacterium</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Mycoplasma</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Proteus mirabilis</strong></td>
</tr>
<tr>
<td><strong>Viruses</strong></td>
<td><strong>Adenovirus</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Cytomegalovirus</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Epstein–Barr virus</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Hepatitis B</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Hepatitis C</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Human parvovirus B19</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Paramyxoviruses</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rubella</strong></td>
</tr>
</tbody>
</table>
Several organisms have been studied as potential causative agents in RA, including various retroviruses, the bacterium Proteus mirabilis, Epstein–Barr virus (EBV), herpes simplex virus 6 (HSV-6), various Mycoplasma species, Escherichia coli, and parvovirus B19. Among these microbes possibly associated with induction of RA, EBV and P. mirabilis have been extensively studied and present perhaps the most consistent data in support of infection as a causative factor in RA. Elevated antibodies specifically targeting EBV and P. mirabilis have been identified in RA samples. Moreover, both of these infectious agents contain antigen proteins/peptides that share a similar molecular structure to that of host tissue peptides, and which therefore can perpetuate inappropriate immune reactions to self. EBV is a virtually ubiquitous virus which induces vigorous antibody responses, but the virus is rarely if ever eradicated; it has been one of the strongest candidates to cause chronic immune diseases, including RA, for more than 25 years. Patients with RA have high levels of specific anti-EBV antibodies compared to those of healthy control subjects. Epstein–Barr nuclear antigen (EBNA), a protein transcription factor involved in the regulation of latent gene expression, viral capsid antigen, and early antigen, both of the latter of which are expressed during the lytic phase of the EBV infection cycle, show elevated levels in most RA patients. Using polymerase chain reaction (PCR) and other assay methods, researchers have identified the presence of EBV antigens in RA joints; higher EBV viral load is also found in peripheral blood lymphocytes of RA patients. An increase in the prevalence of EBV in RA joints versus healthy controls is controversial; that is, studies describe conflicting observations. Edinger and colleagues reported that EBV gene expression is not altered in RA synovia even in the presence of EVB antigen-specific T cell clones. These investigators also found no difference in EBV gene expression in peripheral blood when they compared RA with other diseases such as osteoarthritis, Gulf War syndrome, and others. In addition, Mousavi-Jazi and colleagues detected EBV DNA using PCR in only two of 31 patients with chronic RA compared to 14 control subjects. A direct causal link between EBV and RA has yet to be established because of the inconsistency of these observations.

The bacterium P. mirabilis is a second well studied organism postulated to be involved in RA etiology. Elevated IgG and IgM titers specific for this organism have been documented in blood and urine samples of RA patients compared to those of control individuals. Molecular similarities have been identified in amino acid sequences...
present in hemolysins and urease enzymes of *P. mirabilis* and in host tissue components. *P. mirabilis* also is a cause of urinary tract infections, which interestingly enough are ten times more common in women than in men. Some investigators have suggested that this may underlie the two to threefold higher risk for RA development in women compared to men; however, this causal relationship remains speculative at this point.

Many other microbes have been reported to be associated with or responsible for pathogenesis in RA, only to be subsequently ruled out as etiological agents because data supporting causality for each of them has proved to be inconsistent among screening reports.\(^6^9,7^0\) Retroviruses, for example, have been suspected as activating agents in several autoimmune diseases, including RA. An article by Brand and colleagues\(^6^6\) reported that human retrovirus 5 proviral DNA was detected in 53% of synovial samples extracted from arthritic joints of RA patients, but a contemporaneous study from another group found no association between human retrovirus 5 and RA.\(^7^1\) Interestingly, a recent study demonstrated that autoantibodies to citrullinated α-enolase peptide 1 in a significant proportion of patients with RA cross-react with enolase from the bacterium *Porphyromonas gingivalis*; this latter is an oral pathogen that causes periodontitis with similarities to RA.\(^7^2\) These autoantibodies were identified in only 2%-3% of individuals with no or other forms of arthritis. It will be important to continue research along this line to determine whether this organism is involved either directly or indirectly in the elicitation of antibodies to citrullinated peptides/proteins in RA patients.

As mentioned, criteria have been formulated, in our view insightfully, which specify the characteristics that any single organism must exhibit to be the etiologic agent causing RA.\(^6^0\) In order for infection to be etiologic in RA, the causative organism should be ubiquitous geographically, since RA is common around the globe; the infection should be persistent in humans, since the disease is usually chronic; the microbial agent must elicit an immune response, and lastly, the infection must be present only in diseased tissue and not in healthy controls. None of the pathogens discussed above satisfy all these criteria, and more importantly even when researchers have found organisms that they argue do more or less fit these criteria, their work often has failed to be replicated by other laboratories.\(^7^3,7^4\) Thus, infection by any of the organisms indicated above and several others as the sole causative agent for RA essentially has been dismissed.\(^6^7\) However, this view of infectious causation results from use of an otherwise extremely successful biologic paradigm, Koch’s postulates, which we argue below is inappropriately applied to RA, as well as to other currently idiopathic chronic diseases.

**Genetics as the etiologic basis for rheumatoid arthritis**

Strong evidence exists supporting the presence of a significant genetic component involved in susceptibility of RA. However, it is estimated that genetic polymorphisms or mutations explain only approximately 50%-60% of the genetic effect, since genetic predisposition can be influenced by so many other factors, including environment, smoking, lifestyle and even infection.\(^7^5\) A 1.5-fold increased rate of RA occurrence in first degree relatives has been observed, but as discussed above, studies of monozygotic twins show concordance rates of only 12%-15%; the occurrence in dizygotic twin’s drops dramatically to only about 3.5%.\(^7^6\) Currently, five genes are most frequently associated with increases in RA susceptibility: *HLA DRB1*, *PTPN22*, *6q23*, *STAT4* and *TRAF1/C5* (*PA DI 4* asian descent only).\(^1^4\) See Table 2 for a list of genes/loci associated with the etiology of RA.\(^1^4,7^7-7^9\) Variants within two genes, *HLA DRB1* and *PTPN22* appear to present the greatest influence on

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Genes/loci associated with the etiology of rheumatoid arthritis(^1^4,6^8-7^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genetic/Loci</strong></td>
<td><strong>Abbreviations</strong>: HLA, human leukocyte antigens; PTPN22, protein tyrosine phosphatase nonreceptor 22; CS-TRA F1, compliment component TNF-S receptor-associated factor 1; CTLA 4, cytotoxic lymphocyte antigen 4; STAT 4, signal transducer and activator of transcription factor 4; PAD 4, peptidylarginine deiminase 4; Fc-receptor-like gene 3; TNFRSF14, tumor necrosis factor receptor superfamily member 14; TNFAIP3, tumor necrosis factor alpha-induced protein 3; PRKCQ, protein kinase 3 theta; KIF5A, kinesin family member 5A; RUNX1, runt-related transcription factor 1; CIITA, class II, major histocompatibility complex, transactivator; IL-10, interleukin 10.</td>
</tr>
<tr>
<td>HLA</td>
<td></td>
</tr>
<tr>
<td>PTPN22</td>
<td></td>
</tr>
<tr>
<td>CS-TRA F1</td>
<td></td>
</tr>
<tr>
<td>CTLA 4</td>
<td></td>
</tr>
<tr>
<td>STAT 4</td>
<td></td>
</tr>
<tr>
<td>PAD 4</td>
<td></td>
</tr>
<tr>
<td>FRLG1</td>
<td></td>
</tr>
<tr>
<td>TNFRSF14</td>
<td></td>
</tr>
<tr>
<td>TNFAIP3</td>
<td></td>
</tr>
<tr>
<td>PRKCQ</td>
<td></td>
</tr>
<tr>
<td>KIF5A</td>
<td></td>
</tr>
<tr>
<td>CD40</td>
<td></td>
</tr>
<tr>
<td>RUNX1</td>
<td></td>
</tr>
<tr>
<td>CIITA</td>
<td></td>
</tr>
<tr>
<td>IL-10</td>
<td></td>
</tr>
</tbody>
</table>
increasing susceptibility and are therefore discussed in some detail here.

**HLA-DRB1**

Variation in the MHC, or human leukocyte antigen (HLA), locus on chromosome 6 has been linked reliably to RA across all populations. Approximately 30%–50% of the genetic component of RA has been associated with the HLA-DRB1 gene. A conserved amino acid sequence has been identified in the third hypervariable region of the DR beta chains at amino acids 70–74. This region is designated the “shared epitope” (SE) and contains the sequence glutamine-lysine-arginine-alanine-alanine (QKRAA). Expression of the SE is associated with increased susceptibility to RA, and it also correlates with more severe pathogenesis. The SE may serve multiple roles in the pathogenesis of RA. For example, arthritogenic peptides from a self antigen or a microbial pathogen could bind to the SE and lead to positive or negative selection of autoimmune T cells in the thymus. In turn, this may engender an inadequate number of Treg cells, which normally inhibit T cell activation and control immune responses, thereby activating autoreactive T cells. In addition, the SE itself could become the target for T cells via molecular mimicry. However, other published studies do not link the SE directly with RA pathogenesis, and 20% of RA patients lack the SE altogether. Interactions between the shared epitope and B cells may induce production of autoantibodies, which can perpetuate an immune response that can lead to RA pathogenesis.

**PTPN22**

The second most compelling documented genetic risk for RA is a single nucleotide polymorphism (SNP) in the gene encoding the protein tyrosine phosphatase nonreceptor 22 (PTPN22) on chromosome 1p13. Tyrosine phosphatase is a lymphoid enzyme which negatively regulates T cell activity and which may also regulate other cells due to its expression in B cells and myeloid cells. The risk of developing RA is two- to fourfold higher in individuals who carry this SNP; the variant has been confirmed in UK, Swedish, Finnish, German, Dutch, Spanish, and Canadian populations. A study of Japanese individuals identified a significantly lower allele frequency compared to those discussed above, which is of interest because RA shows a low prevalence in this population, as discussed above. The PTPN22 variant also has been associated with several other diseases, including juvenile idiopathic arthritis, Graves’ disease, SLE, and type 1 diabetes. Other genetic factors have been investigated and their associated risks documented, but HLA-DRB1 and PTPN22 appear to generate the greatest associated risk and the maximum consistency compared with other genetic variants. Clearly, some genetic component(s) promoting susceptibility to RA pathogenesis does/do exist, but no single variant has been identified to date that causes the disease. We argue below that the challenge is to identify any remaining genetic effects and to explore how these variants interact with one another, as well as how they interact with environmental and other factors to induce development of RA.

**Linkage**

Whole genome scans for RA have been carried out to identify HLA and non-HLA loci involved in RA pathogenesis and susceptibility. The results of most such studies have not been replicated and are considered unreliable. Many researchers argue that this inconsistency is a result of small sample sizes, low statistical power, and clinical or genetic heterogeneity, but we suggest that this evidence simply may be misunderstood. An alternative approach to analyzing susceptibility genes was reported in a recent article by Goh and colleagues. This paper presented a global view of what was designated a human “diseasome”, linking various diseases to one another through common mutated or alternately expressed genes. In the case of RA, four genes are included in the diseasome and are linked with other diseases; those genes are PTPN22, that encoding IL-10, that specifying runt-related transcription factor 1 (RUNX1), and class II transactivator (CIITA) (see Figure 1).

In our view, one attractive possibility for expanding our understanding of pathogenesis in RA, and in other clinical entities as well, is to assess the means by which genes associated with this clinical entity manifest in other diseases. For example, SLE is a chronic inflammatory connective tissue disorder that can involve many organs, including joints. As with RA the cause of SLE is currently unknown, but also as with RA an important autoimmune component exists in the disease. Further, SLE displays a significant gender bias towards women. Some patients with SLE will develop a symmetrical polyarthritis while others might not experience any arthritis at all. Could “genetic overlap” of these inflammatory disorders explain these clinical similarities, and likewise the dissimilarities? The real world utility of the Goh and colleagues’ strategy remains to be established in a rigorous manner at this point, as does the precise and most illuminating means of applying it to disease causality. Regardless, in our view the study of differing effects of disease-related genes in other, more or less related pathogenic
contexts ultimately should provide an important alternative strategy for the analysis of genetic risks, along with the many factors associated with RA pathogenesis.78

The contribution of other factors to initiation and maintenance of rheumatoid arthritis pathogenesis

It follows from the observations reviewed above that neither infection nor specific single or multiple genetic attributes can be held uniquely responsible for the initiation and maintenance of RA pathogenesis. In turn as suggested by us and others,7,8,9 some combination of genetic attributes with infectious agent(s) or other environmental factor(s) must constitute the etiologic basis for that pathogenic process. The latter, environmental, factors largely remain to be defined, however. Some studies indicate that diet plays a role in the development of RA. That is, caffeine intake, low levels of antioxidants in the diet, and diets high in red meat have been associated with increased risk of RA, while vitamin D may show some promise for decreasing the overall prevalence of this clinical entity; alcohol consumption may act as a protective agent.5,9,10 Such reports are limited, however.3

One quite well studied environmental influence that increases risk for development of RA is tobacco smoking.5,9,11–93 Reports from a number of large-scale studies both in North America and Europe have indicated that duration and intensity of smoking (eg, the number of cigarettes smoked/day) both increase risk.9,93 Interestingly, one study suggested that the risk of developing RA is not increased in passive smokers,93 although more research is required in this area. The risk of developing RA for smokers, while increased, appears to be relatively modest in individuals who do not possess the HLA DRB1 SE (see above, Autoimmunity as a component of RA pathogenesis), but in those who do possess this allele risk can be increased several-fold.89,94–96 In contrast, no increased risk for RA appears to exist in smokers possessing the PTPN22 allele. Importantly, smoking has been shown to be involved in engendering antibodies to cyclic citrullinated peptides/proteins in patients with RA,95–98 and studies demonstrate that the risk for RA is significantly increased in smokers who have the HLA DRB1 SE and anticitrullinated protein antibodies.88 Smoking may interact in some manner with P. mirabilis infection of the urinary tract to increase the risk of RA development, although again this is an area requiring additional study.99

As outlined above, it has been firmly established that autoimmune diseases in general, and RA in particular, affect women more often than they do men, and interestingly, women appear to be more susceptible to autoimmune disease following their reproductive years. The reason for this large gender bias remains unexplained, but steroid hormones have long been postulated to play a key role.93,100,101 Changes in the female hormonal environment due to menstruation or

---

**Figure 1** Linkage diagram showing potential genetic associations with several autoimmune diseases. Copyright © 2007, National Academy of Sciences U.S.A. Adapted with permission from Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. Proc Natl Acad Sci U S A. 2007;104:8685–8690.

**Abbreviations:** MS, multiple sclerosis; BLS, bare lymphocyte syndrome; AML, acute myelogenous leukemia; GVHD, graft vs host disease; HIV, human immunodeficiency virus; PDD, platelet defect deficiency; IDDM, insulin-dependent diabetes mellitus; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; RUNX1, runt-related transcription factor 1; CIITA, Class II, major histocompatibility complex, transactivator; IL-10, interleukin 10; PTPN22, protein tyrosine phosphatase nonreceptor 22.
during pregnancy, while breast feeding, and/or through the use of oral contraceptive pills (OCP), appear to have a role in RA pathogenesis. Increased breast feeding duration is protective for RA development, while the literature evaluating OCP use and parity is inconsistent, demonstrating protection in some studies and no associated risks in others.

The female systemic immune system is considered by some investigators to be more “reactive” than those of men. This contention is supported by observations demonstrating higher immunoglobulin concentrations, enhanced cytokine production, greater resistance to infections, and stronger T cell responses in women, along with other characteristics not mentioned. This overall increase in immune function may explain in part the heightened susceptibility to autoimmune disease.

Pregnancy has been shown to ameliorate symptoms associated with RA. According to Nelson and Ostensen, symptoms improve in approximately 70% of pregnant RA patients, but the condition often returns within three months postpartum. It is important to note, though, that the mother is exposed to paternal gene products expressed by the fetus, and her immune system is clearly tolerant. The mechanism by which this occurs is not well understood, but the quiescent state of the immune system during pregnancy seems most likely to be responsible for the symptomatic improvements experienced by patients with RA.

Elevated female immune function in combination with the amelioration of RA symptoms during pregnancy has led to the idea that estrogen may stimulate immune reactions and initiate autoimmunity in selected populations. This mechanism is reviewed in a recent article by Kovats and Carreras in which the regulation of DC differentiation and function by estrogen receptor ligands is discussed. Since they govern APCs, DC’s play a vital role in innate and adaptive immunity, and in immune tolerance. DC’s express estrogen receptors on their surfaces and have a relatively short lifespan, indicating that short-term exposure to various estrogens may modulate DC activation.

Ex-vivo studies by Paharkova-Vatchkova and colleagues showed that 17ß-estradiol (E) promotes the differentiation, not the proliferation, of functional DC’s from murine bone marrow precursor cells. Even more interesting is the observation that ex vivo, DC differentiation in E-deficient medium is inhibited by the addition of an estrogen receptor antagonist such as tamoxifen, but differentiation is restored when the inhibitor is removed and E supplemented.

If estrogen has a pathological affect on a woman’s immune system it would be logical to posit that postmenopausal women would experience some protection from development of RA, since circulating estrogen levels are much lower at this time. This is not the case, however, since several epidemiologic studies have demonstrated that the incidence of RA in women continues to increase with advancing age. The effects of post-menopausal hormone replacement therapy (HRT) on the incidence and severity of RA was evaluated by the Women’s Health Initiative in randomized controlled trials using historical and medication data. No statistically significant differences were identified relating to the risk of developing RA, or relating to the overall severity of the disease if induced, between HRT-treated and control groups. Another setback to the idea that estrogen exacerbates immune reactions in women is that RA symptoms often get worse in the postpartum period, when estrogen levels have fallen. Research suggests that low estrogen levels may stimulate the immune system, whereas high concentrations may be inhibitory; this clearly is an area that deserves further study.

An alternative concept has emerged, as recent observations indicate that cell trafficking between the fetus and the mother takes place during pregnancy. Many studies show that a low number of fetal cells can and often do persist in the maternal circulation for years after delivery, a concept designated microchimerism. This may perpetuate an immune response and could provide an additional explanation for the strong gender bias exhibited in RA. As mentioned, variations in the HLA-DRB1 gene appear to provide the greatest genetic risk factor for development of RA. This variant is not universally expressed in all RA patients, though, which makes it a challenge to associate HLA-DRB1 possession with RA etiology. However, recent research suggests that women who do not express the HLA allele SE may acquire it later as a result of microchimerism. A study by Rak and colleagues targeted two groups of women, one with, and a control group without, RA, all patients lacking any RA-associated HLA-DRB1 mutations or associated variants. When compared with healthy women, 42% of women with RA had a higher frequency of, and higher levels of, the DRB1 microchimerism versus 8% of controls. This difference was not statistically significant, but it does provide another attractive mechanism for interactions between genetics and environmental stimuli.

Unfortunately, both fetal and maternal microchimerism have been documented in healthy individuals, suggesting a possible advantage of some kind to the host. More studies need to be done in order to distinguish among these possibilities.

Discussion

This article has discussed a number of factors relating directly or indirectly to the etiology and pathogenesis of RA, with the
intention to demonstrate that previous studies of infectious, genetic, or other causation taken in isolation have provided little or no useful insight into that etiology. We suggest, however, that when the existing data are taken in context and as a whole, a picture does emerge as to how RA pathogenesis is initiated. The picture indicates that, contrary to current research paradigms, RA is not a discrete clinical entity with a single, unique etiological source. Rather, we suggest that RA, as with some other currently idiopathic chronic diseases, is a common endpoint for a number of different starting points. Thus, the focused search for single etiologic agents or factors in the genesis of RA pathogenesis has been and continues to be misleading. Such a search represents a too-simple approach to a complex and multi-factorial problem.

We contend that the paradigms underlying that too-simple approach, including Koch’s postulates, should be abandoned in favor of more sophisticated approaches to understanding the initiation of joint pathogenesis in RA.

That RA is a common clinical endpoint for differing initiating factors and pathogenic mechanisms is supported by the many different clinical presentations that encompass the diagnosis of RA. Thus as indicated, Koch’s postulates, one of the most successfully employed paradigms of modern infectious disease research, may not be applicable to causation in this context of RA. That is, if RA is a consequence of the acquisition of any one or more of several different micro-organisms or environmental or other toxins during a special window of susceptibility, and if disease genesis is dependent on some aspect(s) of genetic background of the individual acquiring that particular organism(s) or toxin at that time, then the classic notion of a single specific causative agent cannot be applied to this disease. Instead, a new, far more complex paradigm is required. We note in passing that observations from other human diseases have undermined to a large extent the general utility of Koch’s postulates. For just one example, in the case of chronic Chlamydia-induced inflammatory (reactive) arthritis, the organism can only rarely be cultured from affected joints; culture of the offending organism from the site of pathogenesis is one of the four main requirements for satisfaction of the postulates, but few clinicians or others would argue that C. trachomatis is not etiologic for the inflammatory arthritis. Indeed, it is well accepted that this bacterium is viable and metabolically active in those joints, but that it exists at affected sites in an unusual biological state which does not allow completion of its standard developmental cycle to produce new infectious organisms. Moreover, an essentially identical inflammatory reactive arthritis is caused by infection with enteric pathogens such as Salmonella, and these organisms appear not to be viable at all in the joints of affected patients.

If Koch’s postulates can not be applied to elucidation of microbial involvement in the etiology of RA, and by extension the involvement of other environmental factors, then the definition of any such involvement must be approached in an entirely different manner. Future screening of materials from RA patients targeting specific organisms or toxins or other environmental agents must be done on patient samples for which detailed histopathology has been done, and for which extensive clinical histories are available. It also will be important to study patient materials from individuals early in disease development and from those with established disease. Moreover, in all cases information regarding the patients’ genetic background should be obtained, although it is anything but clear at this point which aspects of that background, in addition to those discussed above, are relevant to initiation of pathogenesis. This latter subject unquestionably will constitute a major research focus in the future. We add at this point that definition of the etiologic complexities underlying RA should provide important new insights into the relative clinical severity seen among patients with RA. Such definition also may well provide a rational explanation for the differential response of patients with RA to biological modifiers such as those targeting TNFα or its receptor and other therapeutic modalities.

As outlined here, both specific infectious agents and genetics have been dismissed as sole etiological sources for RA, since the data supporting each mechanism have proved to be inconsistent. In contrast, if one considers RA to be not a single disease but rather the common endpoint for many different potential initiating causes, the discrepancy is essentially obviated. Some recent publications favor this notion. One review by van der Helm and colleagues connects susceptibility genes for RA with anti-CCP expression and also associates an additional set of distinct genes with anti-CCP absence, all of which point to classification of RA into discrete, and perhaps differentially induced, disease subsets. In another article discussing B cell involvement in RA pathogenesis and autoimmunity, the authors contended that no single antigen or epitope exists that is specific for RA pathogenesis. They further suggested that several autoantigens exist which may be distinct throughout different populations of RA patients.

It seems an ineluctable conclusion from the summary of research presented here that many factors contribute to RA etiology, and that they do so differentially in individual
patients. Multiple infectious agents, various environmental factors, and as yet poorly understood aspects of patient genetic background all are involved in causation and pathogenesis of RA, but to what extent in any given individual remains to be determined. We contend that a reorientation of the approach to research regarding causation in RA, and in the study of etiology of other currently idiopathic chronic conditions as well, ultimately will prove to be highly productive for the ultimate goal of developing new therapies for the precise treatment, and perhaps cure, for specific RA patients. However, the development of new therapies for RA based on knowledge of the true causative factor(s) in each patient will not be, in reality, a simple matter at all. If RA has multiple initiating causes, including infectious and/or other environmental agents, as yet poorly understood aspects of human genetic background, and widely varying combinatorial complexities associated with them, then treating patients with this condition is destined to become a textbook example of personalized medicine; ie, RA or any endpoint clinical entity resulting from different initiating causative factors simply cannot be treated effectively in all patients by any one single overall therapeutic strategy. Defining complex initial causation events and agents in RA certainly will not be an easy task, especially if, as we contend here, multiple causative factors operate to engender the condition in any given patient. Regardless, we argue that this is the approach which is by far the most likely to yield meaningful new therapeutic strategies for treatment, and possibly prevention, of RA.

Acknowledgments
The authors report no conflicts of interest regarding this work. Preparation of this article was supported by research grants AR-42541 (APH), AR-48331 (JAW-H), and AR-53646 (JDC) from the US National Institutes of Health. Some portions of the manuscript were submitted in part fulfillment of the requirements for the degree of Master of Basic Medical Sciences at Wayne State University School of Medicine (JAS).

References


