The Mechanism of Tissue-Restricted Antigen Gene Expression by AIRE

Kristina Zumer, Kalle Saksela and B. Matija Peterlin

*J Immunol* 2013; 190:2479-2482;
doi: 10.4049/jimmunol.1203210
http://www.jimmunol.org/content/190/6/2479

**References** This article cites 58 articles, 24 of which you can access for free at:
http://www.jimmunol.org/content/190/6/2479.full#ref-list-1

**Subscription** Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions** Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts** Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
The Mechanism of Tissue-Restricted Antigen Gene Expression by AIRE

Kristina Žumer,* Kalle Saksel,* and B. Matija Peterlin*†

The autoimmune regulator (AIRE) gene was identified by positional cloning of the genetic locus linked to a rare autoimmune disease, autoimmune-polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (1, 2). The encoded AIRE protein is expressed primarily in medullary thymic epithelial cells (mTECs) (3). AIRE is also expressed in peripheral lymphoid tissues (4), where its contribution to tolerance remains to be investigated. The mouse model of the human APECED disease, the Aire−/− mouse, was instrumental in identifying the cellular role of AIRE, which is to regulate promiscuous expression of tissue-restricted Ag (TRA) genes in mTECs (5). The key finding was that mTECs from Aire−/− versus wild-type mice expressed fewer TRAs. Although this promiscuous gene expression in mTECs had been recognized previously (6), the underlying mechanism remained elusive. To date, AIRE is the only identified transcription factor that regulates this process. Among others, AIRE activates insulin, interphotoreceptor retinoid-binding protein A, and mucin 6 genes, all of which could be linked to autoimmunity in humans or AIRE-deficient mice (7, 8). Expressed TRAs are then processed and presented on the surface of mTECs or taken up by dendritic cells (9, 10). Exposure of maturing T cells to these Ags is critical for negative selection of autoreactive T cells.

AIRE is a transcription factor that assembles into oligomers and forms punctate structures that colocalize with CREB-binding protein, P-TEFb, and small nuclear ribonucleoproteins in the nucleus (23–25). The estimated number of AIRE-regulated genes ranges from several hundred to thousands. They have diverse promoters and are regulated by distinct transcription factors in their corresponding tissues. These findings raise the conundrum of how AIRE can regulate such a large repertoire of divergent genes. Recent work from others and us has addressed this question and revealed the mechanisms of action of this enigmatic protein (26–31).

The AIRE protein has a predicted molecular mass of 58 kDa (Fig. 1A) (3). It forms large oligomers, which are detected in a >670-kDa fraction by gel filtration (32). The N terminus of AIRE contains the homogeneously staining region (HSR) and the Sp100, AIRE-1, NucP41/75, and DEAF-1 (SAND) domain (Fig. 1A). Mutations in HSR from APECED patients
disrupt multimerization, which is essential for AIRE function (32, 33). The AIRE SAND domain is also involved in this oligomer formation. It lacks the canonical KDWK motif, which is required for DNA binding of other SAND domains (34). Although DNA binding of the AIRE SAND domain had been reported in vitro (35, 36), it appears to be non-specific (37) and may be irrelevant for its recruitment to target genes in vivo. Indeed, we found that mutating KNKA residues in the AIRE SAND domain, which correspond to the KDWK motif in other members of this family, had no effect on AIRE-induced expression of a plasmid reporter gene (26).

In some APECED patients an autosomal-dominant G228W mutation was found in the SAND domain (Fig. 1A), which causes the wild-type AIRE protein to coaccumulate in larger structures that do not colocalize with sites of active transcription (25). Thus, dominant-negative effects of this G228W mutation could be due to an increased affinity for the wild-type AIRE protein. The N terminus of AIRE is required for nuclear localization (38). The nuclear localization signal was mapped to basic residues from positions 131 to 133 (Fig. 1A) (39).

AIRE also contains two plant homeodomains, plant homology domain (PHD)1 and PHD2 (Fig. 1A). PHDs are zinc fingers closely related to RING domains, which are common in proteins involved in ubiquitylation (40). PHDs are protein–protein interaction modules, which can mediate binding to nucleosomes (41, 42). PHDs bind to N-terminal tails of histone H3, discriminating between those methylated (H3K4me3) or unmodified (H3K4) at Lys4. AIRE PHD1 is most closely related to the BHC80 PHD that interacts selectively with the unmodified H3K4 (42). Indeed, AIRE PHD1 also binds to the unmodified H3K4 (Fig. 1A), and this interaction is required for AIRE to activate transcription of genes in chromatin (26, 30, 37). The AIRE PHD2 sequence is divergent and does not interact with nucleosomes, but it contributes structurally to the activation of TRA genes by AIRE (27, 31, 43). AIRE PHD1 also binds to the DNA-dependent protein kinase (DNA-PK) (Fig. 1A) (44). DNA-PK is a nuclear kinase, which not only functions to repair DNA double-strand breaks and mediate V(D)J recombination, but it also supports transcription and chromatin remodeling (45). DNA-PK also phosphorylates two sites in the N terminus of AIRE (44). However, pharmacological inhibition of DNA-PK and bringing AIRE to a promoter via heterologous DNA tethering in cells lacking DNA-PK revealed that the kinase activity of DNA-PK is dispensable for gene activation by AIRE. Instead, these interactions represent a key mechanism for the recruitment of AIRE to its target genes (26). An important aspect of
this targeting is that AIRE interacts with DNA-PK, which is associated with the histone variant H2AX phosphorylated at Ser139 (γH2AX) that marks DNA double-strand breaks (26).

The C terminus of AIRE does not show obvious homology with functional domains in other proteins, but it is highly conserved between human and mouse AIRE proteins. It serves as a transcriptional activation domain (TAD) (Fig. 1A) (27). It binds to P-TEFB and brings it to TRA genes (23, 27). This leads to the phosphorylation of Ser2 in the RNAPII CTD and productive elongation with cotranscriptional processing of nascent mRNA species. The key role of this domain is underscored by an APECED patient mutation that affects only the extreme C terminus in AIRE (Fig. 1A) (46) but completely abolishes its function (27).

AIRE contains four LXXLL motifs, two in the HSR, one in a proline-rich region between the two PHDs, and one in the TAD (Fig. 1A). LXXLL motifs are known to mediate protein–protein interactions between nuclear receptors and their coactivators (47). Their role for AIRE remains to be established.

Armed with all of this knowledge of interacting proteins and chromatin targeting strategies of AIRE, we can now propose a model for the AIRE-regulated promiscuous TRA gene expression in mTECs. The defining feature of AIRE target genes is that although they are inactive, they have already engaged RNAPII on their promoters (Fig. 1B) (23, 27, 29). This means that the basal transcriptional machinery is already in place, but RNAPII is stalled or generates only sterile and unprocessed transcripts, which are unstable (48, 49). Because this RNAPII is not phosphorylated at its CTD, it also fails to recruit chromatin-modifying machineries. Thus, H3K4 remains unmodified, leading to PHD1-mediated recruitment of AIRE to such inactive genes (Fig. 1B) (30, 37). Initiation of transcription has been associated with topoisomerase II–catalyzed formation of transient double-strand DNA breaks, leading to DNA-PK recruitment to initiating RNAPII (50, 51). DNA-PK–associated AIRE is thereby brought into close proximity of the initiating RNAPII (Fig. 1B). Because the binding between AIRE PHD1 and unmodified H3K4 is insufficient to target AIRE to specific TRA genes (52), only the combinatorial interactions of AIRE with all its partners, the unmodified H3K4, DNA-PK, and engaged RNAPII, lead to the recruitment of AIRE with associated P-TEFB to these transcription units (Fig. 1C). Levels and distribution of these factors at distinct TRA genes will vary from cell to cell, thereby giving rise to the seemingly stochastic nature of TRA gene expression by AIRE.

However, several other aspects of AIRE remain to be investigated. For example, how is its transcription regulated in mature mTECs or in peripheral lymphoid tissues (53, 54)? There exist three splice variants of human AIRE, some of which lack the N-terminal domains required for the formation of oligomers (2). Do they play any role in central tolerance? Furthermore, several posttranslational modifications of AIRE have been reported (44, 55, 56). How do they affect the function of AIRE and who directs these transcriptional and posttranslational events to direct central tolerance? Finally, P-TEFB, the essential coactivator of AIRE, is itself tightly regulated in cells. Among other stimuli, P-TEFB can also be activated by cellular stress (57). Thus, it is likely that increased numbers of DNA breaks in mTECs also activate P-TEFB (28), thereby further enhancing the expression of TRA genes to optimize the establishment of central tolerance.

Conclusions
AIRE is a transcription factor that activates the expression of TRA genes in mTECs. Their promoters must be occupied by RNAPII, unmodified H3K4, and DNA-PK. Sufficient levels of these proteins ensure that AIRE is recruited to these sites. AIRE oligomers then bring P-TEFB to RNAPII, which leads to its extensive phosphorylation. Thus modified, RNAPII is competent for elongation and cotranscriptional processing of target genes, which leads to the expression of TRAs and their presentation to T cells via MHC class II determinants.

Disclosures
The authors have no financial conflicts of interest.
The monopartite nuclear localization signal of autoimmune regulator mediates its nuclear import and interaction with multiple importin α molecules. *FEBS J.* 273: 315–324.


