11756–11768 *Nucleic Acids Research, 2012, Vol. 40, No. 22* *Published online 15 October 2012*
doi:10.1093/nar/gks933

AIRE-PHD fingers are structural hubs to maintain the integrity of chromatin-associated interactome

Massimiliano Gaetani[1], Vittoria Matafora[2], Mario Saare[3], Dimitrios Spiliotopoulos[1], Luca Mollica[1], Giacomo Quilici[1], Francesca Chignola[1], Valeria Mannella[1], Chiara Zucchelli[1], Pärt Peterson[3], Angela Bachi[2,*] and Giovanna Musco[1,*]

[1]Biomolecular NMR Laboratory, Center of Translational Genomics and Bioinformatics, Dulbecco Telethon Institute c/o S. Raffaele Scientific Institute, [2]Biomolecular Mass Spectrometry Unit, Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Via Olgettina 58, 20132, Milano, Italy and [3]Department of Molecular Pathology, University of Tartu, 50411 Tartu, Estonia

Received May 3, 2012; Revised September 11, 2012; Accepted September 14, 2012

ABSTRACT

Mutations in autoimmune regulator (AIRE) gene cause autoimmune polyendocrinopathy candidiasis ectodermal dystrophy. AIRE is expressed in thymic medullary epithelial cells, where it promotes the expression of peripheral-tissue antigens to mediate deletional tolerance, thereby preventing self-reactivity. AIRE contains two plant homeodomains (PHDs) which are sites of pathological mutations. AIRE-PHD fingers are important for AIRE transcriptional activity and presumably play a crucial role in the formation of multimeric protein complexes at chromatin level which ultimately control immunological tolerance. As a step forward the understanding of AIRE-PHD fingers in normal and pathological conditions, we investigated their structure and used a proteomic SILAC approach to assess the impact of patient mutations targeting AIRE-PHD fingers. Importantly, both AIRE-PHD fingers are structurally independent and mutually non-interacting domains. In contrast to D297A and V301M on AIRE-PHD1, the C446G mutation on AIRE-PHD2 destroys the structural fold, thus causing aberrant AIRE localization and reduction of AIRE target genes activation. Moreover, mutations targeting AIRE-PHD1 affect the formation of a multimeric protein complex at chromatin level. Overall our results reveal the importance of AIRE-PHD domains in the interaction with chromatin-associated nuclear partners and gene regulation confirming the role of PHD fingers as versatile protein interaction hubs for multiple binding events.

INTRODUCTION

Mutations in the autoimmune regulator (AIRE) gene coding for the multidomain protein AIRE (~58 kDa) cause the autosomal recessive syndrome, autoimmune polyendocrinopathy ectodermal dystrophy (APECED) (OMIM No. 240300) (1,2). APECED is recognized as a unique model for molecular studies of autoimmunity, since the disease is monogenic and characterized by multiple features of abnormal immunological tolerance leading to destructive autoimmune reactions in several organs (3). AIRE is predominantly expressed in thymic medullary epithelial cells (mTECs), where it promotes the ectopic expression of a repertoire of peripheral-tissue antigens (PTAs) to mediate deletional tolerance, thereby preventing self-reactivity (4). In absence of functional AIRE, many PTAs are under-expressed and cannot be presented to thymocytes (5). This leads to the escape of self-reactive T-lymphocytes to the periphery and ultimately to pathogenic autoimmune reactions. The mechanisms dictating AIRE role in transcriptional activation of PTAs are still under investigation. AIRE structural organization reflects its role as transcriptional

*To whom correspondence should be addressed. Tel: +39 02 26434824; Fax: +39 02 26434824; Email: musco.giovanna@hsr.it
Correspondence may also be addressed to Angela Bachi. Tel: +39 02 26434927; Fax. +39 02 26436585; Email: bachi.angela@hsr.it
Present addresses:
Massimiliano Gaetani, Department of Laboratory Medicine and Advanced Biotechnologies, University of Pittsburgh Medical Center, ISMETT (Mediterranean Institute for Transplantation and Advanced Specialized Therapies), Via Tricomi 5, 90127 Palermo, Italy.
Luca Mollica, Protein Dynamics and Flexibility, NMR Institut de Biologie Structurale, UMR 5075, CNRS/CEA/UJF, Grenoble, France.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

activator: it harbors a N-terminal/CARD domain, a nuclear localization signal, four LXXLL motifs, an atypical SAND domain likely mediating DNA binding and two plant homeodomain (PHD)-type zinc fingers (6). Both the PHD fingers contain the sites of several pathological missense mutations or are absent in several APECED-causing truncation mutants (7). Recently, it was shown that AIRE binds directly to hypomethylated histone H3 in position K4 (H3K4me0) through its first PHD finger (AIRE-PHD1), thus activating gene expression (8–11). Notably, the second AIRE PHD finger (AIRE-PHD2), whose structure has been so far unknown, does not interact directly *in vitro* with histone H3 (8,9), even though transactivation properties have been ascribed to this domain (12). In normal conditions, AIRE localizes predominantly in punctate nuclear bodies, where it partners with a number of chromatin related proteins, including CBP (13), P-TEFb (14), DNA-PK1 (15,16). In addition, a recent important screening for proteins interacting with AIRE yielded a vast array of possible AIRE interactors involved in chromatin structure and DNA-damage response, gene transcription, and RNA processing (17). Most likely, distinct regions of AIRE participate in various protein–protein interactions, thus subtending the formation of different complexes. AIRE-PHD fingers, which are strongly related to AIRE transcriptional activity, are also expected to be crucial in the formation of multimeric protein complexes at chromatin level. As a further step toward the understanding of the role of AIRE-PHD fingers in normal and pathological conditions, we investigated by a proteomic and structural approach the impact of pathological point mutations targeting AIRE-PHD fingers. We show that in the context of a tandem domain construct, the two AIRE-PHD fingers have independent non-interacting folds connected by an unstructured proline-rich linker. Our nuclear magnetic resonance (NMR) studies showed that in contrast to D297A (8) and to the patient mutation V301M (18,19), the C446G mutation on AIRE-PHD2 (20) destroys the fold, thus causing AIRE aberrant localization and reduction of transcriptional activation of representative AIRE target genes. Moreover, using two different SILAC approaches, we show that mutations on AIRE-PHD1 strongly affect the formation of multimeric protein complexes at chromatin level. Overall our results underline the importance of AIRE-PHD domains in the interaction with chromatin-associated nuclear partners and gene regulation.

MATERIALS AND METHODS

Sample preparation for NMR

Human AIRE-PHD2 and AIRE-PHD1–PHD2 tandem domain constructs (residues Q425-E485 and N295-E485, respectively, GenBank AB006682) were prepared in modified pET-24d vectors (Novagen Inc., Madison, WI, USA) by PCR cloning from an AIRE template. Site-directed mutations were made by standard overlap extension methods. Full-length AIRE WT and the mutant form C446G were also cloned into modified pET-24d vectors to enable expression as GST-fusion proteins. The DNA sequences of all constructs were verified in-house and the molecular weights were verified by mass spectrometry (MALDI). The modified pET-24d vectors express proteins with N-terminal 6-His tags, removable by cleavage with TEV (tobacco etch virus) protease, enabling use of non-tagged proteins in NMR studies.

The protein purification strategy was described previously (19). The purified sample was exchanged into 50 mM sodium phosphate pH 6.5, 0.15 M NaCl, 5 mM DTT, 50 μM $ZnCl_2$, 0.02% NaN_3, 10% (v/v) D_2O. For the isotope-labeled proteins used in NMR studies, uniformly ^{15}N- and $^{13}C/^{15}N$-labeled proteins were prepared using M9 minimal growth media appropriately supplemented with ^{15}N-labeled ammonium chloride and ^{13}C-labeled glucose. NMR experiments were performed with a protein concentration of 0.4–0.8 mM.

NMR spectroscopy and resonance assignments

NMR experiments were performed at 22°C on a 600-MHz spectrometer (Bruker Avance 600 Ultra Shield TM Plus, Bruker BioSpin) equipped with a triple-resonance TCI cryoprobe with a z-shielded pulsed-field gradient coil. Data were processed with NMRPIPE (21) and analyzed using CCPnmr (22). The ^{1}H, ^{13}C and ^{15}N chemical shifts of the backbone resonances were obtained from sensitivity-enhanced 3D HNCA, CBCA(CO)NH and CBCANH experiments. The side chain signals were assigned from 3D HCCONH-TOCSY, CCONH-TOCSY and HCCH-TOCSY experiments (23). The tautomeric state of the histidine ring (protonation of Nϵ2) was determined performing a 2D ^{1}H–^{15}N HMQC experiment optimized to detect J-couplings in histidine side chains (24). Proton–proton distance constraints were obtained from two ^{15}N and ^{13}C separated 3D NOESY spectra employing 100 ms mixing times. $^{3}J(H^{N}, H^{\alpha})$ coupling constants were measured to derive restraints for φ dihedral angles. Additional φ/ψ restraints were obtained from backbone chemical shifts using TALOS+ (25). Hydrogen bond restraints were defined from slow-exchanging amide protons identified after exchange of the H_2O buffer to D_2O. Heteronuclear ^{1}H–^{15}N nuclear Overhauser enhancements (NOEs) as well as longitudinal and transversal ^{15}N relaxation rates were measured using standard 2D methods (26). The relaxation delays were applied in an interleaved manner. The T1 and T2 decay curves were sampled at 14 different time points (50–1600 and 12–244 ms, respectively). The heteronuclear NOE experiments were run twice in an interleaved fashion with and without (reference spectrum) proton saturation during proton recovery delay. The relaxation experiments were analyzed with the program Nmrview.5.03 (27).

Structure calculations

Structures were calculated using ARIA 2.3.2 (28) in combination with CNS using the experimentally derived restraints (Table 1). All NOEs were assigned manually and calibrated by ARIA 2.3.2. The automated assignment of ARIA was not used. A total of eight iterations

Table 1. Summary of conformational constraints and statistics for the 20 best structures of AIRE-PHD2

Restraints used for structure calculations[a,b]	
All	1070
Sequential ($\|i-j\| = 1$)	237
Medium range ($1 < \|i-j\| \leq 4$)	81
Long range ($\|i-j\| > 4$)	252
Intraresidual	494
Hydrogen bonds	6
Dihedral angles	80
Deviation from idealized covalent geometry	
Bonds (Å)	0.006 ± 0.0002
Angles (°)	0.175 ± 0.028
Coordinate precision (Å)[c]	
N, C^{α}, C′	0.77 ± 0.14
All heavy atoms	1.26 ± 0.14
Structural quality[d]	
% Residues in most favored region of Ramachandran plot (%)	86.40
% Residues in additionally allowed region (%)	13.60

[a]No distance restraint in any of the structures included in the ensemble was violated by >0.4 Å.
[b]No dihedral angle restraints in any of the structures included in the ensemble was violated by >5°.
[c]RMS deviation between the ensemble of structures and the mean structure calculated on residues 432–477.
[d]Calculated on residues 432–477.

(20 structures in the first six iterations) were performed: 100 structures were computed in the last two iterations. The ARIA default water refinement was performed on the 20 best structures of the final iteration. In ARIA 2.3.2, the geometry of the Zn coordination is fixed through covalent bonds and angles in the CNS parameters; the tetrahedral angles and distances for Zn coordinating residues were maintained also after water refinement. Structural quality was assessed using PROCHECK-NMR (29). Molecular images were generated using molecular visualization system Pymol, Delano Scientific (http://www.pymol.org/). Structure similarity searches were performed using the DALI server (http://www.ebi.ac.uk/dali/index.html) (30). The family of the 20 lowest energy structures has been deposited in the PDB (2lri). Chemical shift and restraints lists that were used in the structure calculations have been deposited in BioMagResBank (accession number 18374).

Plasmids, antibodies

The plasmid encoding AIRE wild-type or AIRE carrying the single point mutations V301M, C446G and D297A were produced from pcDNA3-AIRE vector by PCR site-directed mutagenesis and verified by DNA sequencing. The GST-tagged full-length AIRE has been described before (8). To generate the GST-tagged AIRE-PHD2 mutant, AIRE C446G cDNA was amplified from pcDNA3-AIREC446G and cloned into EcoRI/HindIII sites of pGEX-1ZT-SH3 (a gift from Dr K. Saksela, University of Tampere).

Antibodies specific to the following proteins were used: AIRE (goat; Everest Biotech), RuvBL-1 (Sigma-Aldrich), RuvBL-2 (polyclonal Ab; Sigma-Aldrich), tubulin (mAb; Sigma-Aldrich), KAP-1 (Sigma-Aldrich), DNA-PK1 (Santa Cruz Biotechnology, USA), ubiquitin (mAb; Santa Cruz Biotechnology), histone H3 (polyclonal Ab; Abcam, Cambridge, MA, USA).

GST pull down assays

The GST-tagged proteins were expressed in NovaXG *Escherichia coli* strain by addition of 0.8 mM IPTG and 0.2 mM $ZnCl_2$. The proteins were purified using Glutathione Sepharose 4B (GE Healthcare) as described earlier (8). Sepharose-bound proteins (2–10 μg) were incubated with 20 μg calf thymus histones (Sigma) in binding buffer [50 mM Tris–HCl pH 7.5, 1 M NaCl, 1% NP-40, 0.5 mM ethylenediaminetetraacetic acid, 1× protease inhibitor cocktail (Sigma)] for 4 h at 4°C and washed eight times with binding buffer. The bound proteins were visualized by separation on 15% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) followed by Coomassie staining.

Cell culture

HEK293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% FCS and 100 U/ml penicillin and streptomycin (Invitrogen, USA). Stable isotope labeling was carried out essentially as described previously (31) using [^{12}C6, ^{14}N4] arginine (referred to as Arg0), [^{12}C6, ^{14}N2] lysine (referred to as Lys0), [^{13}C6, ^{15}N4] arginine (referred to as Arg10) and [^{13}C6, ^{15}N2] lysine (referred to as Lys8) (Cambridge Isotope Laboratories, Cambridge, MA, USA). The labeled cells, for each experiment, were mixed in a 1:1 ratio (4×10^6 cells each).

Transfection

HEK293T cells were transfected for 24 h with pCDNA3 empty vector, pCDNA3-AIRE, pCDNA3-AIRED297A, pCDNA3-AIREV301M, pCDNA3-AIREC446G plasmids using Lipofectamine™ 2000 Transfection Reagent (Invitrogen) according to the manufacturer's protocol.

Immunofluorescence

HEK293T cells were grown on sterile 13-mm coverslips and then transfected with the constructs as previously indicated. Comparison of AIRE staining for the wild-type and mutated forms was performed. Cells were fixed with phosphate-buffered saline (PBS), 4.0% paraformaldehyde for 15 min at room temperature and then permeabilized with 0.2% Triton X-100, 300 mM sucrose, 20 mM HEPES, pH 7.4, 50 mM NaCl, 3 mM $MgCl_2$ for 3 min at 4°C. HeLa cells were incubated with anti-AIRE antibodies in blocking buffer (0.2% bovine serum albumin in PBS) for 1 h at 37°C, rinsed with PBS, incubated with purified Alexa Fluor 488-conjugated anti-goat immunoglobulin G antibodies, rinsed and mounted with immuno-fluore mounting medium (ICN Biomedicals, Costa Mesa, CA, USA). The nuclei were visualized by Hoechst 33258 (Sigma-Aldrich) staining for 3 min at room temperature. Fluorescence was visualized with an inverted fluorescence microscope (DM IRBE;

Leica, Wetzlar, Germany) and captured with a TCS-NT argon/krypton confocal laser microscope (Leica).

RNA purification and quantitative real-time PCR

Forty-eight hours after transfection, total RNA from transfected HEK293 cells was purified with TRIzol® (Life Technologies) according to manufacturer's instructions. Total RNA of 3 μg was used for cDNA synthesis with SuperScript™ III First-Strand Synthesis System for reverse transcriptase (RT)–PCR (Life Technologies) according to manufacturer's instructions. Quantitative real-time PCR was performed with the ViiA™ 7 Real-Time PCR System (Life Technologies) using the Maxima Probe/ROX qPCR Master Mix (2×) (Thermo Scientific). All reactions were performed in triplicates and experiments were repeated at least twice. Transfections with an empty vector were used as a negative control. Relative amount of mRNA was calculated using the comparative C_t method (Life Technologies): $2^{-\Delta\Delta C_t} = 2^{-[(C_{t_{target}} - C_{t_{HPRT}})_{sample} - (C_{t_{target}} - C_{t_{HPRT}})_{control}]}$, where C_t is the threshold cycle and HPRT is the housekeeping gene. At least one primer from each pair was designed to cover an exon–exon boundary. Primers for the expression analysis were: (i) HPRT (as housekeeping gene): HPRTexon6 GACTTTG CTTTCCTTGGTCAGG, HPRTexon7 AGTCTGGCTT ATATCCAACACTTCG; (ii) Involucrin: hINV-for GCC TTACTGTGAGTCTGGTTGACA, hINV-rev GGAGG AACAGTCTTGAGGAGCT; (iii) S100A8: S100A8-for CTCAGTATATCAGGAAAAGGGTGCAGAC, S100A8-rev CACGCCCATCTTTATCACCAGAATGA G; (iv) HBG2: HBG-for TGAGAACTTCAAGCTCCT GG, HBG2-rev GACAGGGCACTGGCCACTC; (v) BAT2 (AIRE-independent gene): hBAT2-fexpr-for CCA GAGCAAATCTTACCCAG, hBAT2_expr-rev TACTG TAACCTGCTGGAGAG and (vi) S100A10 (AIRE-independent gene): S100A10-for TCCCTGGATT TTTGGAAAATCAAAAAGACC, S100A10-rev CCCG CAATTAGGGAAAAGAAGCTCTGGAA.

Conventional nuclear extraction, mChIP extraction, immunoprecipitation, protein electrophoresis, immunoblotting, PAM-MAP SILAC

For the conventional nuclear extract, HEK293T cells were separated into cytoplasmic and nuclear fractions using a previously published protocol (32). For the modified chromatin immunoprecipitation protocol (mChIP), HEK293T cells were separated into cytoplasmic and nuclear fractions as described before but an additional step of sonication wad added to the nuclear extract. Before the centrifugation, nuclear extract was sonicated three times for 30 s with at least 1 min on ice between each pulse. Nonidet P-40 was added to a final concentration of 0.2% and the sample was mixed for 30 s. The extract was gently clarified by centrifugation at 1800 relative centrifugal force for 10 min (4°C) and the supernatant was transferred to a fresh tube (33). AIRE target proteins were immunoprecipitated using anti-AIRE antibody that was incubated with protein G-Sepharose 4 Fast Flow beads (GE Healthcare/Amersham Biosciences) for 1 h at 4°C. Nuclear or mChIP lysates were adjusted to IBB buffer (10 mM Tris–HCl, pH 8, 0.2% Nonidet P-40, 150 mM NaCl) and pre-cleared with protein G-Sepharose beads for 1 h at 4°C. The clarified supernatants were incubated with anti-AIRE coupled to protein G resin for 2 h at 4°C. The beads were rinsed several times with IBB buffer, re-suspended in Laemmli buffer, heated at 85°C and centrifuged at 10 000 *g*. Depending on the type of SILAC experiment, the heavy and light protein extracts were mixed before or after the IP. In particular, in the MAP SILAC, the isotopes were mixed just before the elution, during the last washing step, while for the PAM SILAC, the isotopes were mixed at the level of the extracts and then immunoprecipitated. In the latter case, the fast on/off rates of dynamically interacting proteins will result in equilibrium between the heavy and light form of the proteins that are bound to the bait. The relative ratios of these interactors depend on the kinetic parameters of the individual binding partners and will be similar to those of background proteins after a certain time interval. Therefore, dynamic but specific interactors can be effectively distinguished by measuring their ratio variability depending on the isotope mixing before (MAP-SILAC) or after the purification (PAM-SILAC) (34,35). AIRE IP eluates were separated by 4–12% SDS–PAGE (Invitrogen), stained with Coomassie Brilliant Blue (Bio-Rad) and excised in eight slices for LC-MS/MS analysis. In immunoblotting experiments, proteins separated by SDS–PAGE were subsequently transferred onto nitrocellulose membranes (GE Healthcare/Amersham Biosciences). These membranes were incubated with specific antibodies.

Mass spectrometry and data analysis

Mass spectrometry analysis was performed by LC-MS/MS using an LTQ-Orbitrap mass spectrometer (Thermo-Scientific, Bremen, Germany). Tryptic digests for each band were first cleaned using Stage Tips as described previously (36) and then injected in a capillary chromatographic system (EasyLC, Proxeon Biosystems, Odense, Denmark). Peptide separations occurred on a homemade column obtained with a 10-cm fused silica capillary (75 m inner diameter and 360 m outer diameter; Proxeon Biosystems) filled with Reprosil-Pur C18 3-m resin (Dr Maisch GmbH, Ammerbuch-Entringen, Germany) using a pressurized 'packing bomb'. A gradient of eluents A [distilled water with 2% (v/v) acetonitrile, 0.1% (v/v) formic acid] and B [acetonitrile, 2% (v/v) distilled water with 0.1% (v/v) formic acid] was used to achieve separation from 8% B (at 0 min, 0.2 ml/min flow rate) to 50% B (at 80 min, 0.2 ml/min flow rate). The LC system was connected to the orbitrap equipped with a nanoelectrospray ion source (Proxeon Biosystems). Full-scan mass spectra were acquired in the LTQ-Orbitrap mass spectrometer in the mass range m/z 350–1500 Da and with the resolution set to 60 000. The 'lock-mass' option was used for accurate mass measurements. The 10 most intense doubly and triply charged ions were automatically selected and fragmented in the ion trap. Target ions already selected for the MS/MS were dynamically excluded for 60 s (37).

Protein identification and quantification were achieved using the MaxQuant software version 1.0.13.13 (38). Mass spectra were analyzed by Mascot (version 2.2.2) against a concatenated forward and reversed version IPI.human.v3.68.decoy.fasta database. The initial mass tolerance in MS mode was set to 7 ppm and MS/MS mass tolerance was 0.5 Da. Cysteine carbamidomethylation was searched as a fixed modification, whereas *N*-acetyl protein and oxidized methionine were searched as variable modifications. Labeled arginine and lysine were also specified as variable modifications. The resulting Mascot '.dat' files were loaded into the MaxQuant software together with the raw data for further analysis. SILAC peptide and protein quantification were performed automatically with MaxQuant using default settings as parameters. Protein quantification was based on extracted ion chromatograms of contained peptides. Peptide assignments were statistically evaluated using a Bayesian model on the basis of sequence length and Mascot score. Peptides and proteins were accepted with a false-discovery rate of 0.01, two minimum peptides identified per protein of which one unique. The experiments were done in biological duplicate performing two technical replicate for each biological one.

RESULTS

AIRE-PHD2 is a bona fide PHD finger

We solved AIRE-PHD2 3D structure by multidimensional heteronuclear NMR spectroscopy (Figure 1A). The recombinant protein (residues Q425-E485) behaves as a monomer in solution, as assessed by the rotational correlation time ($\tau_c \sim 5$ ns), determined from ^{15}N relaxation data. This is in agreement with the expected value for a folded 7-kDa protein. Residues R433-S476 adopt a well-defined tertiary structure with an RMSD of 0.77 Å for backbone atoms and have all residues in the allowed regions of the Ramachandran plot (Table 1). In contrast, the N- and C-terminal residues are disordered as indicated by the small number of NOEs and negative heteronuclear NOE (Figure 1B). AIRE-PHD2 adopts the typical PHD finger fold characterized by a short two-stranded anti-parallel β-sheet (residues L444-R445 and A452-A453) followed by an α-helical turn and a C-terminal α-helix (residues R473-S476) (Figure 1A). The two Zn^{2+} ions are coordinated in a cross-braced scheme by C432, C437, H454, C457 and C446, C449, C472, C475, respectively. The domain is further stabilized by a network of conserved hydrophobic interactions composed by V443, L444, A452, F453, W455, F459, L470 (Figure 1A). Overall, the variable loops L1 (G338-R445) and L2 (A450-R456) are well defined, whereas L3 (H458-R471) is less ordered, in agreement with the paucity of inter residual NOEs in this region (Figure 1A and Supplementary Figure S1A). Consistently, residues in L3 show a reduction of the heteronuclear NOE intensities, thus indicating internal motions on a rapid time scale (pico- to nanosecond range) (Figure 1B). We observed similar dynamics in the corresponding loop in AIRE-PHD1 (19). Structural comparison performed with DALI server reveals high structural homology with AIRE-PHD1, with a *Z*-score of 3.9 and a RMSD of 1.6 Å over the C^{α} atoms of equivalent residues (Supplementary

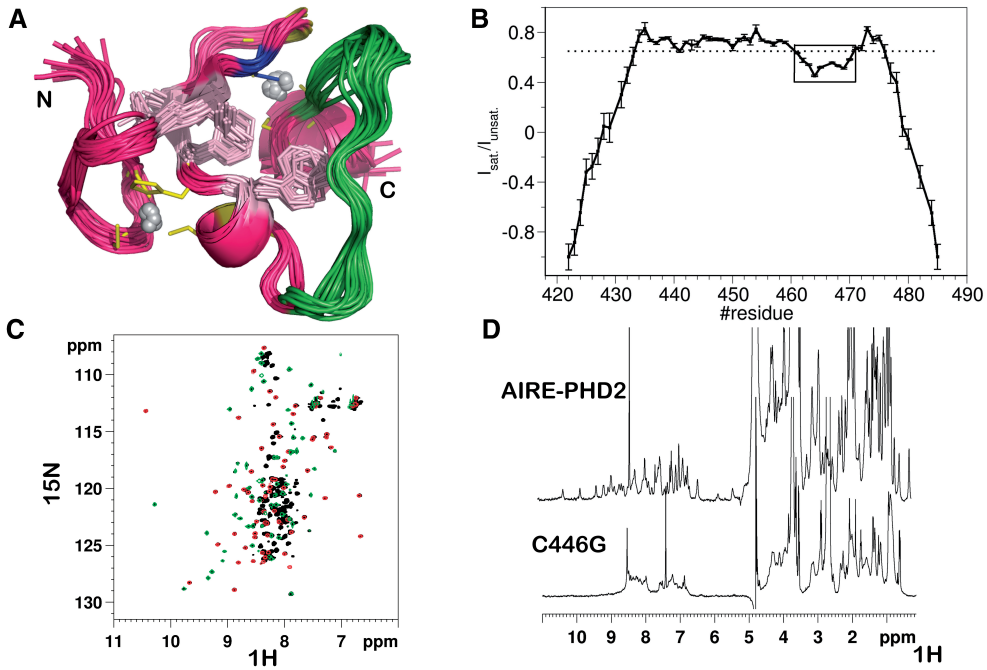

Figure 1. (**A**) Solution structure of AIRE-PHD2. Superposition of the best 20 structures. Zn^{2+} (gray spheres) binding residues, Cys446, conserved hydrophobic amino acids and L3 are represented in yellow, blue, pink and green, respectively. (**B**) Backbone dynamics of AIRE-PHD2. The black box indicates residues of L3 displaying heteronuclear NOEs <0.65 (dotted line). (**C**) Superposition of ^{1}H ^{15}N HSQC spectra of AIRE-PHD1–PHD2 tandem domain (black), of AIRE-PHD1 (red) and of AIRE-PHD2 (green). (**D**) 1D ^{1}H NMR spectra of AIRE-PHD2 wild-type and of C446G mutant.

Figure S1C). In agreement with previous homology modeling studies (39), the analysis of the electrostatic surface potential revealed a large positively charged surface which is not suited for favorable interactions with the positively charged histone tails (Supplementary Figure S1B), indicating that the two PHD fingers—though structurally very similar—strongly diverge in function. Importantly, when linked in the context of the double domain (residues N295–E485), the two domains maintained their folds, behaved as structurally independent and non-interacting domains, as assessed from the ^{1}H–^{15}N HSQC spectrum of the tandem PHD1–PHD2 domain, which was essentially the sum of the ^{15}N HSQC spectra of the single PHD domains (Figure 1C and Supplementary Figure S2A). Additional cross peaks, which belong to the linker region, were clustered in the middle of AIRE-PHD1–PHD2 spectrum, implying that the linker is largely unstructured. T1, T2 and {1H}15 N heteronuclear NOE experiments performed on the double domain construct show that the correlation time of the two PHD fingers within the tandem is similar (5.8 ± 0.2 ns; 6.2 ± 0.4 ns, for the first and second PHD finger, respectively) to the correlation time of the single domains ($\tau_{c\_PHD1} = 5.0 \pm 0.4$ ns; $\tau_{c\_PHD2} = 5.7 \pm 0.5$ ns) and the $R2/R1$ ratio follows the same trend as for the isolated domains (Supplementary Figure S2B). This indicates that the two PHD finger domains within the context of the tandem display independent overall tumbling and that they are not connected through tight contacts or interactions between each other, as the presence of a defined domain arrangement would lead to a change in the tensor components of the rotational diffusion tensor and to more native-like dynamics with significantly larger overall correlation time (overall $\tau_{c\_PHD1–PHD2} = 6.05 \pm 0.4$ ns AIRE). The {1H}15 N heteronuclear NOE values of both AIRE-PHD1 and AIRE-PHD2 in the tandem construct were similar to the values observed in the isolated domain (Supplementary Figure S2C), as expected, since the domains maintain their fold in the tandem construct and have therefore similar picosecond–nanosecond dynamics. Taken together, these data indicate that the two domains do not interact with each other and tumble independently in solution. Overall these results are in line with previous NMR studies in which titrations using the single domains did not highlight any intermolecular interactions (40). Finally, structural comparison with members of the RING finger family shows poor structural alignment (Supplementary Figure S1D and E), indicating structural differences between these two classes of zinc-binding domains.

Structural and functional analysis of pathological C446G mutation

We previously showed the structural features of D297A and V301M mutations in AIRE-PHD1 (8,18). D297A impairs histone H3 binding, thus reducing the expression of AIRE target genes, whereas the pathological V301M mutation preserves both domain fold and the binding to histone H3. An APECED causing mutation (C446G) has been identified in AIRE-PHD2 with a relatively strong phenotype including Type 1 diabetes (20). To elucidate its structural impact, we analyzed by NMR spectroscopy the corresponding AIRE-PHD2 mutant. The C446G mutation, located on the second Zn^{2+} binding site (Figure 1A), impairs metal binding, thus completely destroying the fold of the domain, as shown by the increased line width of the ^{1}H 1D spectrum and by the loss of chemical shift dispersion (Figure 1D). Importantly, in the context of the full-length protein, C446G does not affect the binding with histone H3 in GST-pull down experiments (Supplementary Figure S3), showing that the structural integrity of AIRE-PHD2 is not required for histone H3 binding *in vitro*. This result is in agreement with previous studies performed on deletion mutants of AIRE, which showed that AIRE-PHD2 was dispensable for histone binding *in vitro* (9).

We next investigated whether the C446G mutation might affect AIRE subcellular distribution when transiently expressed into human embryonic kidney epithelial cell line (HEK293T). This cell line has been widely used to demonstrate AIRE dimerization, subnuclear localization, transactivation potential and induction of target gene activation (8,9). The subcellular distribution of this mutant was compared to the wild-type protein and to mutations targeting AIRE-PHD1 (D297A and V301M). Importantly, AIRE-PHD1 mutations did not alter the normal AIRE subcellular distribution (Figure 2A–C), as these mutants mainly localized into the nucleus forming discrete speckles. Conversely, C446G mutant showed an altered subcellular distribution as it accumulated in discrete structures in the cytoplasm and was scarcely detectable in the nucleus (Figure 2D). Consistently, western blot analysis of nuclear and cytosolic extracts confirmed the presence of AIRE wild-type and D297A and V301M mutants in the nucleus, whereas the C446G mutant was barely detectable both in the cytosolic and nuclear extracts. This is in line with the formation of insoluble aggregates, most likely due to the misfolding of AIRE-PHD2 (Figure 2E).

To further investigate the effects of C446G, D297A and V301M mutations on AIRE subnuclear localization, we compared by western blot analysis the AIRE wild-type and mutants in terms of amount of recovered protein after the chromatin enrichment through a mChIP (33) (see further). The amount of C446G mutant recovered after the chromatin solubilization was lower with respect to AIRE wild-type and V301M and D297A mutants, thus confirming the detrimental role of C446G mutation in AIRE sub-nuclear localization and binding to chromatin (Figure 2E, lower panel).

We therefore hypothesized that the misfolding induced by C446G, along with its altered nuclear localization, might also affect the expression of AIRE target genes. Indeed, C446G mutation strongly reduced the transcriptional activation of AIRE target genes, such as involucrin, S100A8 and HBG2. Although AIRE V301M did not alter histone H3 tail binding *in vitro* (8), it strongly reduced AIRE target gene activation *in vivo,* in agreement with previous findings (9) (Figure 2F). Importantly, this result suggests that binding to histone H3 is necessary but not sufficient to activate gene expression and that

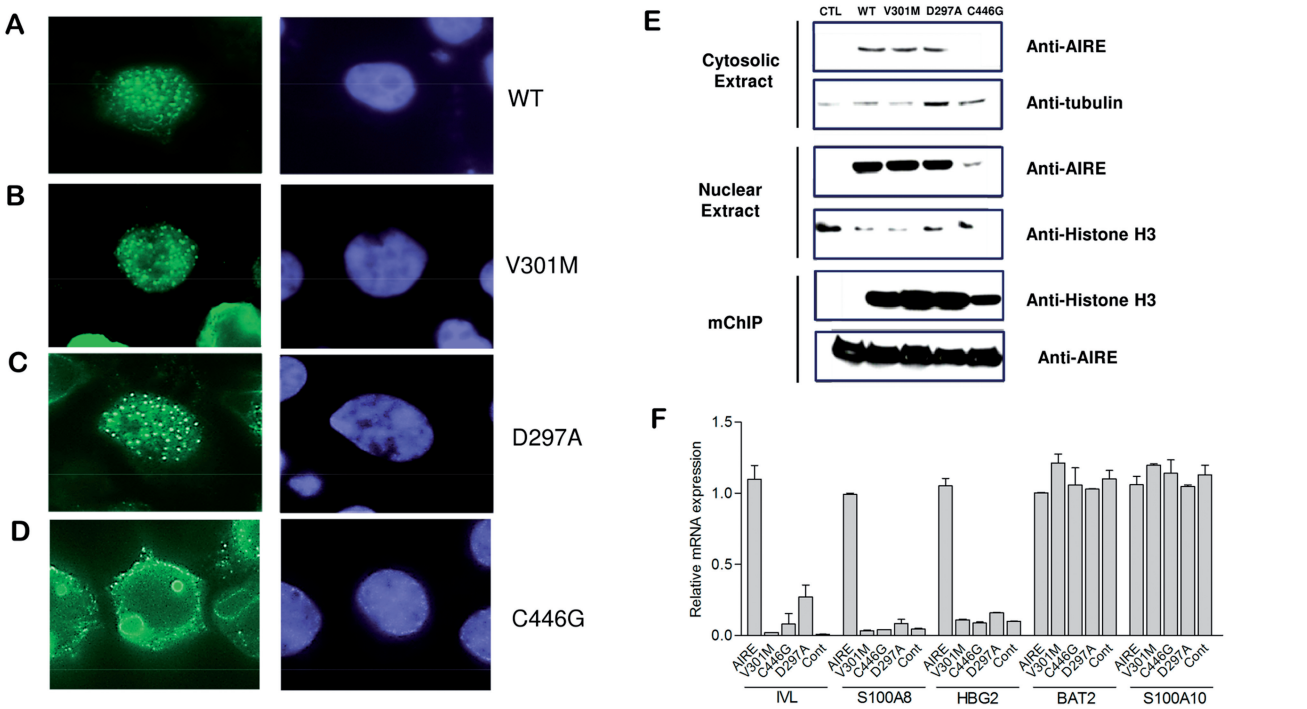

Figure 2. Effects of PHD finger mutations on AIRE cellular localization and on AIRE-dependent gene expression. AIRE staining (green) in HEK293T transfected with (**A**) AIRE wild-type, (**B**) AIRE-V301M, (**C**) AIRE-D297A, (**D**) AIRE-C446G. The nucleus was stained with Hoechst (blue). AIRE wild-type, D297A and V301M mutants localize to nuclear bodies, whereas C446G accumulates into extranuclear aggregates. (**E**) Western blot (using anti-AIRE antibody) of transfected HEK293T with AIRE wild-type and mutants. V301M and D297A mutants recovery in nuclear extracts after chromatin enrichment is comparable to AIRE wild-type, whereas AIRE-C446G recovery is highly reduced. (**F**) Effect of mutations on the activation of AIRE target genes. Relative expression level of AIRE regulated Involucrin (IVL), S100A8 and HBG2 genes and AIRE independent BAT2 and S100A10 genes 48 h after transfection. Quantitative RT–PCR results are presented as means of two independent experiments ± SEM. Cont, negative control transfection.

additional interactions are required for AIRE-mediated transcriptional activation of these genes. This hypothesis is in agreement with the notion that AIRE is part of a macromolecular complex which is recruited to chromatin to activate gene expression, whereby the PHD finger constitutes a molecular platform for the binding of proteins important for AIRE transcriptional activity.

Mutations targeting AIRE-PHD domains alter the interactome of chromatin bound AIRE

To investigate the role of AIRE-PHD fingers in the assembly of AIRE-mediated macromolecular complexes on chromatin, we compared the chromatin-associated interactome of AIRE wild-type with the one generated by AIRE-PHD finger mutants (D297A, V301M, C446G). To this aim, we applied a SILAC-based immunoprecipitation approach coupled to MS (Figure 3A). mTECs are an extremely rare cell population and their numbers are insufficient for a proteomic approach. The MS screening for AIRE's partners was therefore performed in HEK293T cell line transfected with AIRE wild-type and mutant constructs. In order to enrich AIRE complexes on chromatin structures, we performed mChIP to favor the identification of chromatin-related interactors. The protocol allows the efficient purification of protein–DNA macromolecules, enabling subsequent protein identification by MS (33) (see Supplementary Data). The enrichment of protein–DNA complexes was confirmed by the increased recovery of both histone H3 and AIRE protein in mChIP extracts when compared with conventional nuclear extracts (Supplementary Figure S4A and B). In two independent SILAC experiments, HEK293T cells cultured in light and heavy SILAC medium were transfected with empty pcDNA3 or with the pcDNA3-AIRE expression vector, respectively. After 24 h, the cells were lysed and their nuclei were isolated as described in (33). The mChIP extracts were immunoprecipitated with anti-AIRE antibody; bound proteins were then mixed 1:1, eluted and separated by SDS–PAGE. Gel pieces were cut, digested with trypsin and analyzed by LC-MS/MS (Figure 3A). Raw data sets of two independent experiments analyzed with MaxQuant led to the identification of 243 proteins. This list was scrutinized according to the following specificity criteria: (i) the ratio (Heavy/Light) H/L 1.3 (confidence interval >80%) was considered as the interactors' specificity threshold, notably nine-known AIRE interactors fall within this threshold (Supplementary Figure S4C and D; Supplementary Dataset S1); (ii) only proteins with ratio H/L concordant in both biological replicates were considered (Supplementary Figure S5); (iii) for each biological replicate, two technical replicates were analyzed, to increase the robustness of the quantitative data. This filter narrowed the list of putative AIRE-interactors to 24 best candidates (Table 2). Remarkably, most of the interactors could not be detected in conventional nuclear

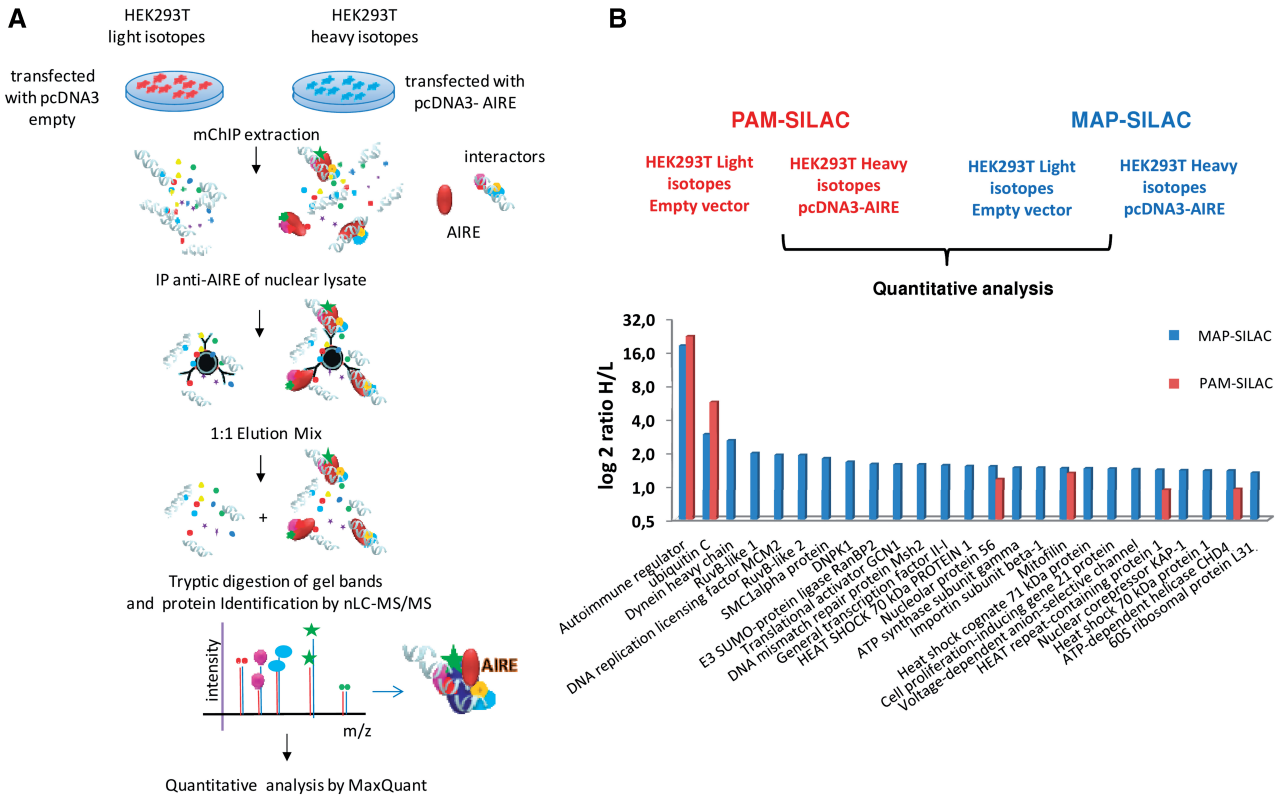

Figure 3. (**A**) Experimental workflow of SILAC combined to mChIP approach. (**B**) Comparison of H/L ratio of interactors using MAP-SILAC and PAM-SILAC. AIRE interactors ratios are higher in the MAP-SILAC than in the PAM-SILAC experiment, confirming the dynamic nature of the interactions.

Table 2. AIRE interactors

Uniprot	Protein names	Peptides (seq)	PEP	Ratio H/L	SD
O43918-1	**AIRE**	32	0	18.16	1.07
A8K674	Ubiquitin	5	3.57×10^{-27}	2.9	1.32
Q14204	Dynein heavy chain, cytosolic	4	1.44×10^{-10}	2.55	0.2
Q9Y265-1	RuvB-like 1	7	3.82×10^{-34}	1.96	0.38
B7Z8Z6	DNA replication licensing factor MCM2[a]	2	3.23×10^{-13}	1.89	0.31
Q9Y230	RuvB-like 2[a]	7	1.49×10^{-61}	1.88	0.49
Q14683	SMC1alpha protein	5	7.10×10^{-05}	1.76	0.47
P78527-1	DNA-dependent protein kinase catalytic subunit[a]	22	9.63×10^{-100}	1.64	0.26
P49792	E3 SUMO-protein ligase RanBP2[a]	6	2.09×10^{-24}	1.56	0.34
Q92616	Translational activator GCN1[a]	2	7.12×10^{-08}	1.55	0.09
P43246	DNA mismatch repair protein Msh2[a]	2	0.022905	1.55	0.27
P78347-1	General transcription factor II-I	5	2.14×10^{-23}	1.52	0.17
B3KTT5	Heat shock 70 kDa protein 1	28	0	1.5	0.03
O00567	Nucleolar protein 56[a]	6	4.40×10^{-36}	1.49	0.58
P36542-1	ATP synthase subunit gamma	2	0.000536	1.45	0.05
Q14974	Importin subunit beta-1[a]	8	9.42×10^{-27}	1.45	0.21
Q16891-1	Mitochondrial inner membrane protein, Mitofilin	4	2.82×10^{-12}	1.44	0.41
P11142-1	Heat shock cognate 71 kDa protein	32	0	1.43	0
P63244	Cell proliferation-inducing gene 21 protein	12	9.75×10^{-99}	1.42	0.19
P45880-1	Voltage-dependent anion-selective channel protein 21	5	5.94×10^{-26}	1.4	0
Q9H583	HEAT repeat-containing protein1	7	3.14×10^{-39}	1.38	0.29
Q13263-1	Nuclear corepressor KAP-1[a]	2	0.001523	1.38	0.17
P08107	Heat shock 70 kDa protein 1	31	0	1.37	0.04
Q14839-2	ATP-dependent helicase CHD4	9	2.12×10^{-40}	1.37	0.19
P62899	60S ribosomal protein L31	3	2.97×10^{-09}	1.31	0.33

Proteins with a SILAC ratio >1.3 (AIRE IP versus control) in both biological replicates are reported. Uniprot accession numbers, with protein name, number of sequenced peptides, posterior error probability, average ratio and SD are reported.
[a]AIRE interactors identified in (17).

extract immunoprecipitation but were identified only after chromatin enrichment (Supplementary Figure S6).

We next investigated AIRE interactome applying a recently developed strategy which considers the highly dynamic nature of chromatin–protein interactions. This SILAC-based approach allows to identify transient but specific interactors by measuring their H/L ratio variability, depending on mixing after or before the purification, defined as MAP- (mixing after purification) and PAM- (purification after mixing) SILAC, respectively (34,35). In particular, MAP-SILAC enables the identification of highly dynamic interacting proteins (see Supplementary Data). Thus, we compared the ratios of the 24 interactors with those obtained when heavy and light isotopes-labeled proteins were mixed before the purification (PAM-SILAC). AIRE maintained similar H/L ratios both in PAM and MAP experiments, whereas several proteins, depending on the dynamics of their interaction, presented different H/L ratios. In particular, in MAP experiments (blue bars in Figure 3B), we detected interactors that were absent in PAM experiments (red bars in Figure 3B), thus confirming their transient interactions (Supplementary Dataset S2). Notably, only ubiquitin maintained its H/L ratio in both SILAC experiments, suggesting a covalent link to AIRE (Figure 3B). Identification of ubiquitin and AIRE in gel bands at high-molecular weight in the Coomassie gel supported this finding (Supplementary Figure S7).

We next investigated whether the point mutations targeting PHD fingers might affect AIRE interactors. Unfortunately, screening of AIRE-C446G interactome was hampered by AIRE partial misfolding, cytoplasmic aggregation, altered nuclear localization and non-specific association to chromatin components, which caused irreproducible results. We therefore focused our analysis on D297A and V301M mutations. In two independent SILAC experiments, HEK293T cells transfected with AIRE-D297A or AIRE-V301M were grown in light SILAC-containing medium, whereas AIRE wild-type transfected cells were grown in heavy SILAC-containing medium. Proteins were subsequently extracted with the mChIP protocol and interactors were isolated via MAP-SILAC: those having comparable affinity for wild-type and AIRE mutants had H/L equal to 1, while those having higher affinity for AIRE wild-type had H/L ratio >1 (see Supplementary Data). Importantly, most of the interactors preferentially bound AIRE wild-type, as assessed by a ratio $H/L > 1$ for both mutants (Figure 4A and Supplementary Dataset S2). Unfortunately, some of the interactors were not detected in these SILAC experiments and their ratio could not be assigned. Notably, the V301M mutation overall had a stronger impact on AIRE interactome when compared with the D297A mutant, suggesting that the interaction surface with possible AIRE interactors is located around V301, on the opposite side with respect to D297, which is a fundamental residue of the histone H3 binding surface (Figure 4B). We conclude that both AIRE-PHD1 mutations influence AIRE–protein interactions, suggesting that this domain participates not only in histone H3 binding but also in recruiting other protein complexes on chromatin structures.

Validation of candidate AIRE-interacting proteins

To confirm the specificity of the interactions between AIRE and its putative partners, we validated some of them by western blot analysis on direct and reverse co-immunoprecipitation experiments. We selected known AIRE partners, such as DNA-PK1 and KAP-1 as positive controls (17) and RuvBL-1 and RuvBL-2 as putative interactors identified in our study. Both RuvBL-1 and RuvBL-2 proteins, as well as DNA-PK1 and KAP-1, were enriched after AIRE immunoprecipitation (Figure 4C) and were thus confirmed as AIRE binders. Remarkably, the interaction between AIRE and RuvBL proteins in co-immunoprecipitations was missed in a previous proteomic study (17), strongly suggesting that our optimized protocol enables a more efficient detection of transient associations. We also confirmed that AIRE mutations diminish binding to RuvBL-1 (Figure 4D) in co-immunoprecipitation with anti RuvBL-1. However, in *in vitro* pull-down experiments, using GST-AIRE and GST-AIRE-PHD1 and recombinant RuvBL1 or *in vitro* translated RuvBL1/2, we could not detect a direct binding, suggesting that the interaction might be mediated by additional partners or might require other scaffold proteins increasing complex stability (data not shown). Finally, as KAP-1 was not detected in the MAP-SILAC IP experiments with AIRE mutants, we tested whether PHD mutations had a similar effect on its binding to AIRE. Both AIRE-V301M and wild-type co-immunoprecipitated with KAP-1, while this interaction was strongly hampered by D297A mutation (Figure 4D), suggesting that KAP-1 involvement in the macromolecular complex either depends on AIRE recruitment to histone H3 or on the direct or indirect recognition of the same PHD finger surface (Figure 4B).

DISCUSSION

PHD fingers have emerged as a robust-conserved structural scaffold with diversified activities (41,42). PHD fingers are not only epigenetic readers, able to sense the modification status of histone H3, they can also bind to non-histone proteins, thereby expanding the role of PHD fingers in diverse cellular processes, including transcriptional regulation or signal transduction. The functional twist of PHD fingers has been underlined by the discovery that the PHD finger in Pygo–BCL9 Wnt signaling complex (43) and in MLL1–CyP33 complex (44), is able to bind simultaneously histone PTM and non-histone partners by engaging different binding surfaces, thus mediating a direct crosstalk between the 'histone code' decoding and additional cellular pathways. Another dramatic functional shift is represented by the discovery that KAP-1-PHD finger can work as a SUMO E3 ligase in the SUMOylation of the adjacent bromodomain (45).

Recent work has given some insights into the function of AIRE-PHD1, showing its fundamental role in the recognition of non-methylated histone H3 and in the recruitment of AIRE to regions of relatively inactive chromatin where it might have an activating influence (8,9), even though additional proteins and/or DNA-binding must

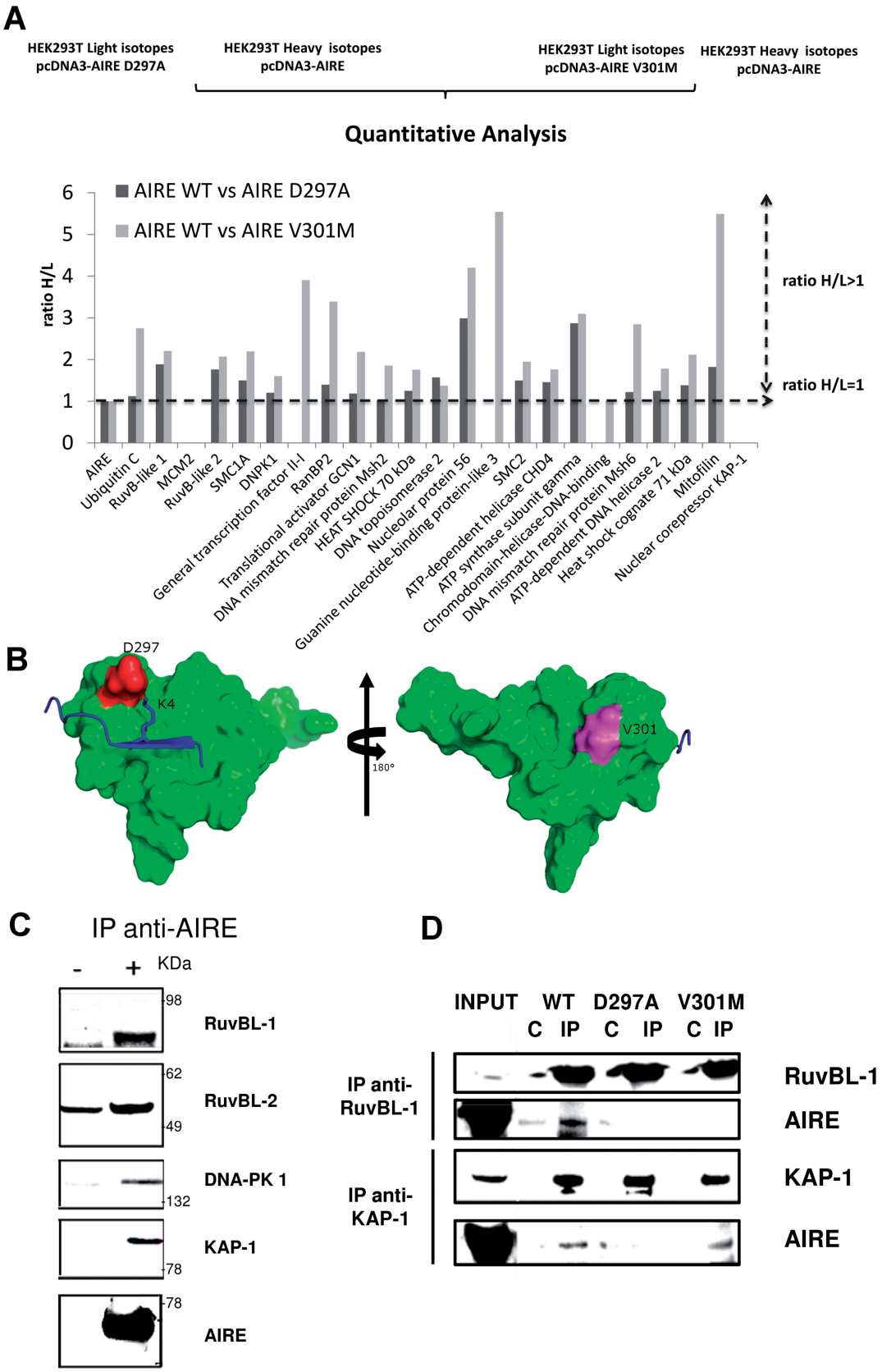

Figure 4. (**A**) Interactomic profiles of AIRE-D297A and -V301M mutants. Comparison between H/L ratios of the SILAC IP from AIRE wild-type (heavy) versus AIRE-D297A (light) and from AIRE wild-type (heavy) versus AIRE V301M (light). A missing bar indicates that the protein has not been identified in this specific experiment. (**B**) Surface representation of AIRE-PHD1 (green) in complex with H3K4me0 peptide (blue cartoon) (2ke1); D297 and V301, located on opposite domain surfaces are represented in red and magenta, respectively. (**C**) Validations of AIRE interactors. Western blot analysis of AIRE IP: DNA-PK1, RuvBL-1, RuvBL-2 and KAP-1 co-immunoprecipitate with AIRE. (**D**) Effects of D297A and V301M mutations on the interaction with RuvBL-1 and KAP-1. Both mutations impair AIRE co-immunoprecipitation with RuvBL-1, whereas only D297A mutation impairs interaction of AIRE with KAP-1.

be involved in the recruitment of AIRE to PTA genes (17). Conversely, little information is available on AIRE-PHD2, whose role has been so far structurally and functionally unexplored. Considering the high functional versatility of PHD fingers (42), it is conceivable that both AIRE-PHD fingers may play yet unidentified functions besides the role of histone readers and may constitute a structural platform for the direct or indirect recruitment of other chromatin-related proteins in an histone-independent fashion. In an attempt to further clarify the differential role of AIRE-PHD fingers both in normal and pathological conditions, we have determined the solution structure of AIRE-PHD2 and we have characterized the impact of point mutations targeting AIRE-PHD fingers on: (i) AIRE subcellular localization, (ii) on the transcriptional effects on AIRE targets genes and (iii) on the formation of the multimeric protein complex at chromatin level.

AIRE-PHD2 adopts the typical PHD finger fold and despite the low-sequence identity with AIRE-PHD1, the two domains superimpose with a RMSD of 1.6 Å over the C^{α} atoms of equivalent residues and display analogous backbone dynamics. Similarly to AIRE-PHD1, AIRE-PHD2 shows poor structural alignment with RING finger structures, thus supporting the notion that structurally AIRE-PHD2 is a *bona fide* PHD finger and does not belong to the class of RING fingers, as hypothesized by phylogenetic analysis (46). AIRE-PHD2, at variance to AIRE-PHD1, displays a positive electrostatic surface, thus providing a structural rationale why AIRE-PHD2 is not suited for interactions with the positively charged histone H3 tail. AIRE-PHD fingers do not interact in the context of a double domain construct and tumble independently in solution, strengthening the hypothesis that the two PHD fingers have evolved to different functions. In agreement with previous data excluding a direct involvement of AIRE-PHD2 in histone H3 binding *in vitro* (8,9), in the full-length context, C446G mutation does not affect binding to histone H3 in *in vitro* assays, further supporting the idea that the two PHD fingers independently exert their function(s). In line with this notion, previous studies focusing on the link between histone binding and AIRE target gene activation showed that AIRE recognition of the hypomethylated H3 tail through AIRE-PHD1 was essential for the ectopic expression of genes encoding PTAs, whereas AIRE-PHD2 seemed to be dispensable for AIRE transcriptional activation properties (9). Our results show that AIRE-PHD2 structural integrity is fundamental for the correct AIRE transcriptional activity, as the destabilization of its fold through the pathological mutation C446G drastically reduces activation of AIRE-dependent genes, suggesting that misfolding of AIRE-PHD2 has a more dramatic outcome on gene activation as compared to the effect caused by its deletion (10). In the context of APECED disease, we hypothesize that this mutation causes the under-expression of PTAs which cannot be presented to thymocytes, thus leading to the escape of self-reactive T-lymphocytes to the periphery and ultimately to pathogenic autoimmune reactions. Herein, it is conceivable that AIRE subnuclear targeting and structural integrity are crucial for its association with specific structures/non-histone-proteins on chromatin. The fact that AIRE transcriptional activity relies on additional interactions, besides binding to histone H3, is underlined also by the observation that V301M mutation, which is irrelevant for histone H3 binding reduces the activation of AIRE target genes, thus suggesting that PHD domains have the potential to interact with other partner proteins that form the fully active transcriptional machinery driving ectopic expression of genes encoding PTAs.

To gain further insights into the molecular details of this machinery, we analyzed AIRE interactome by a quantitative proteomic screening both on AIRE wild-type and on AIRE mutants carrying single point mutations on its PHD fingers. In line with a previous proteomic study (17), we confirmed that AIRE takes part in the formation of multimolecular complexes involved mainly in nuclear transport, DNA repair, mRNA processing, gene transcription and chromatin remodeling. Out of the 24 putative AIRE interactors, 9 were already identified in a previous study (17), supporting the robustness of our analysis. Herein, we provide evidence that the majority of these interactions can be efficiently retrieved only after chromatin solubilization, indicating that AIRE recruitment to DNA is accompanied by the formation of multimolecular complexes. Moreover, the MAP-SILAC strategy revealed that most of the AIRE interactors are highly dynamic, consistent with the transient interactions occurring on chromatin. The only exception was represented by ubiquitin, which might covalently interact with AIRE, as previously proposed (47).

To assess the functional role of AIRE-PHD fingers in the formation of these macromolecular complexes, we analyzed the impact of PHD finger mutations on AIRE chromatin-related interactome. Unfortunately, AIRE-C446G misfolding and subnuclear mis-localization caused irreproducible results, thus hampering a reliable proteomics screening for this mutant. On the other hand, both D297A and V301M mutations induced clear differences in AIRE interactome as compared to the wild-type protein. This is in line with the observations that the two mutations target opposite surfaces of the PHD finger and are therefore conceivably involved in different interaction networks. Overall, with the exception of KAP-1 interaction, V301M mutation had a stronger impact on AIRE interactome with respect to D297A. It mainly reduced the interactions with protein partners involved in transcription, such as general transcription factor II, DNA-PK1, DNA mismatch repair protein, Msh2, thus suggesting that this mutation might alter the formation of macro-complexes necessary to activate transcription. Notably, V301M mutation is located on the opposite surface with respect to the histone H3 interaction site (Figure 4B) and does not affect the binding to histone H3 peptide (8) suggesting a role for AIRE-PHD1 beyond the histone H3 tail interactions. Importantly mutation of V301 affects the expression of AIRE target genes, in agreement with the scenario, for which recruitment of AIRE on chromatin to activate PTAs involves different non-overlapping surfaces and also non-histone proteins, such as DNA-PK1 (16). In conclusion, our study

provides the molecular evidence that PHD-mediated co-operative binding to H3K4me0 and to other cofactors is necessary to fully activate transcription, confirming the role of the PHD finger as a versatile protein interaction hub for multiple binding events.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–7, Supplementary Datasets 1 and 2 and Supplementary References [1,2,19].

FUNDING

Fondazione Telethon [TCP99035 to G.M.]; Fondazione Cariplo [CR-5148 to G.M. and A.B.]; European Regional Fund/Archimedes Foundation (to P.P.). Funding for open access charge: Fondazione Telethon [TCP99035 to G.M.].

Conflict of interest statement. None declared.

REFERENCES

1. Nagamine,K., Peterson,P., Scott,H.S., Kudoh,J., Minoshima,S., Heino,M., Krohn,K.J., Lalioti,M.D., Mullis,P.E., Antonarakis,S.E. *et al*. (1997) Positional cloning of the APECED gene. *Nat. Genet*., **17**, 393–398.
2. Aaltonen,J., Bjorses,P., Perheentupa,J., Horelli-Kuitunen,N., Palotie,A., Peltonen,L., Lee,Y.Su., Francis,F., Henning,S., Thiel,C. *et al*. (1997) An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. the finnish-german APECED consortium. autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy. *Nat. Genet*., **17**, 399–403.
3. Mathis,D. and Benoist,C. (2009) Aire. *Annu. Rev. Immunol*., **27**, 287–312.
4. Anderson,M.S., Venanzi,E.S., Klein,L., Chen,Z., Berzins,S.P., Turley,S.J., von Boehmer,H., Bronson,R., Dierich,A., Benoist,C. *et al*. (2002) Projection of an immunological self shadow within the thymus by the aire protein. *Science*, **298**, 1395–1401.
5. Kisand,K. and Peterson,P. (2011) Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy: Known and novel aspects of the syndrome. *Ann. NY Acad. Sci*., **1246**, 77–91.
6. Aasland,R., Gibson,T.J. and Stewart,A.F. (1995) The PHD finger: Implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci*., **20**, 56–59.
7. Halonen,M., Kangas,H., Ruppell,T., Ilmarinen,T., Ollila,J., Kolmer,M., Vihinen,M., Palvimo,J., Saarela,J., Ulmanen,I. *et al*. (2004) APECED-causing mutations in AIRE reveal the functional domains of the protein. *Hum. Mutat*., **23**, 245–257.
8. Org,T., Chignola,F., Hetenyi,C., Gaetani,M., Rebane,A., Liiv,I., Maran,U., Mollica,L., Bottomley,M.J., Musco,G. *et al*. (2008) The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep*., **9**, 370–376.
9. Koh,A.S., Kuo,A.J., Park,S.Y., Cheung,P., Abramson,J., Bua,D., Carney,D., Shoelson,S.E., Gozani,O., Kingston,R.E. *et al*. (2008) Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proc. Natl Acad. Sci. USA*, **105**, 15878–15883.
10. Koh,A.S., Kingston,R.E., Benoist,C. and Mathis,D. (2010) Global relevance of aire binding to hypomethylated lysine-4 of histone-3. *Proc. Natl Acad. Sci. USA*, **107**, 13016–13021.
11. Chignola,F., Gaetani,M., Rebane,A., Org,T., Mollica,L., Zucchelli,C., Spitaleri,A., Mannella,V., Peterson,P. and Musco,G. (2009) The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic Acids Res*., **37**, 2951–2961.
12. Meloni,A., Incani,F., Corda,D., Cao,A. and Rosatelli,M.C. (2008) Role of PHD fingers and COOH-terminal 30 amino acids in AIRE transactivation activity. *Mol. Immunol*., **45**, 805–809.
13. Pitkanen,J., Doucas,V., Sternsdorf,T., Nakajima,T., Aratani,S., Jensen,K., Will,H., Vahamurto,P., Ollila,J., Vihinen,M. *et al*. (2000) The autoimmune regulator protein has transcriptional transactivating properties and interacts with the common coactivator CREB-binding protein. *J. Biol. Chem*., **275**, 16802–16809.
14. Zumer,K., Plemenitas,A., Saksela,K. and Peterlin,B.M. (2011) Patient mutation in AIRE disrupts P-TEFb binding and target gene transcription. *Nucleic Acids Res*., **39**, 7908–7919.
15. Liiv,I., Rebane,A., Org,T., Saare,M., Maslovskaja,J., Kisand,K., Juronen,E., Valmu,L., Bottomley,M.J., Kalkkinen,N. *et al*. (2008) DNA-PK contributes to the phosphorylation of AIRE: importance in transcriptional activity. *Biochim. Biophys. Acta*, **1783**, 74–83.
16. Zumer,K., Low,A.K., Jiang,H., Saksela,K. and Peterlin,B.M. (2012) Unmodified histone H3K4 and DNA-dependent protein kinase recruit autoimmune regulator to target genes. *Mol. Cell. Biol*., **32**, 1354–1362.
17. Abramson,J., Giraud,M., Benoist,C. and Mathis,D. (2010) Aire's partners in the molecular control of immunological tolerance. *Cell*, **140**, 123–135.
18. Soderbergh,A., Rorsman,F., Halonen,M., Ekwall,O., Bjorses,P., Kampe,O. and Husebye,E.S. (2000) Autoantibodies against aromatic L-amino acid decarboxylase identifies a subgroup of patients with Addison's disease. *J. Clin. Endocrinol. Metab*., **85**, 460–463.
19. Bottomley,M.J., Stier,G., Pennacchini,D., Legube,G., Simon,B., Akhtar,A., Sattler,M. and Musco,G. (2005) NMR structure of the first PHD finger of autoimmune regulator protein (AIRE1). insights into autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) disease. *J. Biol. Chem*., **280**, 11505–11512.
20. Wolff,A.S., Erichsen,M.M., Meager,A., Magitta,N.F., Myhre,A.G., Bollerslev,J., Fougner,K.J., Lima,K., Knappskog,P.M. and Husebye,E.S. (2007) Autoimmune polyendocrine syndrome type 1 in norway: Phenotypic variation, autoantibodies, and novel mutations in the autoimmune regulator gene. *J. Clin. Endocrinol. Metab*., **92**, 595–603.
21. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
22. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
23. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. NMR Spectroscop*, **34**, 93–158.
24. Pelton,J.G., Torchia,D.A., Meadow,N.D. and Roseman,S. (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Sci*., **2**, 543–558.
25. Shen,Y., Delaglio,F., Cornilescu,G. and Bax,A. (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
26. Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D. and Kay,L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
27. Johnson,B.A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol*., **278**, 313–352.
28. Rieping,W., Habeck,M., Bardiaux,B., Bernard,A., Malliavin,T.E. and Nilges,M. (2007) ARIA2: Automated NOE assignment and

data integration in NMR structure calculation. *Bioinformatics*, **23**, 381–382.
29. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
30. Holm,L., Kaariainen,S., Rosenstrom,P. and Schenkel,A. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
31. Ong,S.E., Blagoev,B., Kratchmarova,I., Kristensen,D.B., Steen,H., Pandey,A. and Mann,M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics*, **1**, 376–386.
32. Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.*, **11**, 1475–1489.
33. Lambert,J.P., Mitchell,L., Rudner,A., Baetz,K. and Figeys,D. (2009) A novel proteomics approach for the discovery of chromatin-associated protein networks. *Mol. Cell. Proteomics*, **8**, 870–882.
34. Wang,X. and Huang,L. (2008) Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Mol. Cell. Proteomics*, **7**, 46–57.
35. Mousson,F., Kolkman,A., Pijnappel,W.W., Timmers,H.T. and Heck,A.J. (2008) Quantitative proteomics reveals regulation of dynamic components within TATA-binding protein (TBP) transcription complexes. *Mol. Cell. Proteomics*, **7**, 845–852.
36. Rappsilber,J., Mann,M. and Ishihama,Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2**, 1896–1906.
37. Olsen,J.V., de Godoy,L.M., Li,G., Macek,B., Mortensen,P., Pesch,R., Makarov,A., Lange,O., Horning,S. and Mann,M. (2005) Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics*, **4**, 2010–2021.
38. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
39. Musco,G. and Peterson,P. (2008) PHD finger of autoimmune regulator: an epigenetic link between the histone modifications and tissue-specific antigen expression in thymus. *Epigenetics*, **3**, 310–314.
40. Chakravarty,S., Zeng,L. and Zhou,M.M. (2009) Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure*, **17**, 670–679.
41. Musselman,C.A. and Kutateladze,T.G. (2011) Handpicking epigenetic marks with PHD fingers. *Nucleic Acids Res.*, **219**, 9061–9071.
42. Sanchez,R. and Zhou,M.M. (2011) The PHD finger: A versatile epigenome reader. *Trends Biochem. Sci.*, **36**, 364–372.
43. Fiedler,M., Sanchez-Barrena,M.J., Nekrasov,M., Mieszczanek,J., Rybin,V., Muller,J., Evans,P. and Bienz,M. (2008) Decoding of methylated histone H3 tail by the pygo-BCL9 wnt signaling complex. *Mol. Cell*, **30**, 507–518.
44. Wang,Z., Song,J., Milne,T.A., Wang,G.G., Li,H., Allis,C.D. and Patel,D.J. (2010) Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression. *Cell*, **141**, 1183–1194.
45. Ivanov,A.V., Peng,H., Yurchenko,V., Yap,K.L., Negorev,D.G., Schultz,D.C., Psulkowski,E., Fredericks,W.J., White,D.E., Maul,G.G. *et al.* (2007) PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell*, **28**, 823–837.
46. Saltis,M., Criscitiello,M.F., Ohta,Y., Keefe,M., Trede,N.S., Goitsuka,R. and Flajnik,M.F. (2008) Evolutionarily conserved and divergent regions of the autoimmune regulator (aire) gene: a comparative analysis. *Immunogenetics*, **60**, 105–114.
47. Akiyoshi,H., Hatakeyama,S., Pitkanen,J., Mouri,Y., Doucas,V., Kudoh,J., Tsurugaya,K., Uchida,D., Matsushima,A., Oshikawa,K. *et al.* (2004) Subcellular expression of autoimmune regulator is organized in a spatiotemporal manner. *J. Biol. Chem.*, **279**, 33984–33991.