

The Interaction of Natural Selection and GC Skew May Drive the Fast Evolution of a Sand Rat Homeobox Gene

Yichen Dai¹ and Peter W.H. Holland^{*,1}

¹Department of Zoology, University of Oxford, Oxford, United Kingdom

*Corresponding author: E-mail: peter.holland@zoo.ox.ac.uk.

Associate editor: Bing Su

Abstract

Several processes can lead to strong GC skew in localized genomic regions. In most cases, GC skew should not affect conserved amino acids because natural selection will purge deleterious alleles. However, in the gerbil subfamily of rodents, several conserved genes have undergone radical alteration in association with strong GC skew. An extreme example concerns the highly conserved homeobox gene *Pdx1*, which is uniquely divergent and GC rich in the sand rat *Psammomys obesus* and close relatives. Here, we investigate the antagonistic interplay between very rare amino acid changes driven by GC skew and the force of natural selection. Using ectopic protein expression in cell culture, pulse-chase labeling, in vitro mutagenesis, and drug treatment, we compare properties of mouse and sand rat *Pdx1* proteins. We find that amino acid change driven by GC skew resulted in altered protein stability, with a significantly longer protein half-life for sand rat *Pdx1*. Using a reversible inhibitor of the 26S proteasome, MG132, we find that sand rat and mouse *Pdx1* are both degraded through the ubiquitin proteasome pathway. However, in vitro mutagenesis reveals this pathway operates through different amino acid residues. We propose that GC skew caused loss of a key ubiquitination site, conserved through vertebrate evolution, and that sand rat *Pdx1* evolved or fixed a new ubiquitination site to compensate. Our results give molecular insight into the power of natural selection in the face of maladaptive changes driven by strong GC skew.

Key words: mutation, biased gene conversion, genome evolution, protein evolution, *Pdx1*, homeodomain.

Introduction

Vertebrate genomes are far from uniform in their proportions of A, C, G, and T nucleotides and are mosaics of more AT-rich and more GC-rich regions (Eyre-Walker and Hurst 2001). The processes that generate these patterns are complex but include variation in the mutability of certain nucleotides (Ehrlich and Wang 1981), differences in the probabilities of each transition and transversion mutation (Petrov and Hartl 1999; Schaibley et al. 2013), and the influence of chromosome pairing at meiosis through both recombination-associated AT skewed point mutation (Arbeithuber et al. 2015) and by GC-biased gene conversion (gBGC; Galtier 2001; Galtier and Duret 2007; Arbeithuber et al. 2015). The latter is considered to be strong force acting to elevate GC content locally. gBGC is a process in which GC pairs are preferred over AT pairs at polymorphic sites and occurs by biased repair of mismatched DNA heteroduplexes formed by strand invasion after pairing of homologous chromosomes; it leads to gradual accumulation of GC in the affected genomic region (Galtier 2001; Galtier and Duret 2007). gBGC has been shown to be widely present in vertebrates and is a prominent effect leading to local accumulation of GC content, especially in intergenic regions and synonymous third codon positions (Figuat et al. 2015).

The predominant localization of GC skew to synonymous sites (Eyre-Walker and Hurst 2001) implies that natural

selection is usually effective at removing GC changes that negatively affect protein function. However, an unusual and extreme case has been discovered recently in gerbil genomes whereby very high GC skew is associated with radical coding sequence divergence in localized regions of the genome (Hargreaves et al. 2017). In these animals, several highly conserved proteins have diverged dramatically over a short evolutionary period through amino acid changes almost always associated with changes from AT- to GC-rich codons. This finding allows us a rare insight into the inevitable conflict between GC skewed nucleotide changes and protein function. For example, some highly conserved sites, inferred to be under strong purifying selection, have undergone unique changes in gerbils. These rare cases offer the chance to explore how molecular evolution might compensate for maladaptive functional changes.

The *pancreatic duodenum homeobox 1* gene (*Pdx1*, also referred to as *Xlox*, *lpf1*, *ldx1*, *Stf1*, *XIHbox8*, or *Lox*) is a member of the ParaHox gene cluster and one of the most highly conserved of all homeobox genes (Wright et al. 1989; Brooke et al. 1998; Ferrier and Holland 2002). The *Pdx1* protein is the key transcription factor initiating vertebrate pancreatic development and is also essential for maintaining pancreatic β -cell identity in adulthood (Jonsson et al. 1994; Ahlgren et al. 1996; Ahlgren et al. 1998). There are two key functional domains in the *Pdx1* protein: the hexapeptide domain (PFPWMK)

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

responsible for dimerization with PBC proteins (Liberzon et al. 2004), and the 60 amino acid homeodomain responsible for binding to DNA and sequence-specific recognition of promoter sequences (Lu et al. 1996). Throughout vertebrate evolution, the Pdx1 homeodomain has been highly conserved, retaining 100% amino acid identity between catshark, coelacanth, *Xenopus*, chicken, cat, mouse, and human, and just two to four amino acid changes seen in teleost fish and *Anolis* lizard. This extreme conservation, for over 420 My, suggests that every residue in the 60 amino acid homeodomain is likely to be under strong purifying selection, not just those encoding the three alpha helices key to homeodomain folding.

The importance of the homeodomain sequence is also indicated by human phenotypes. A single amino acid change in the third alpha helix, R197H, is associated with drastically lowered levels of insulin expression and a propensity for heterozygous carriers to develop type II diabetes (Macfarlane et al. 1999). The most extreme phenotype reported is absence of a pancreas in an infant carrying one *Pdx1* allele with a E164D point mutation in the homeodomain and a second allele with the mutation E174K (Schwitzgebel et al. 2003). These changes did not alter protein localization, binding to the *insulin* gene promoter, or interaction with known Pdx1 cofactors, but both mutant proteins had significantly shortened protein half-life as measured in cell culture assays. Given that previous work has shown that Pdx1 is degraded via the ubiquitin 26S-proteasome pathway (Claiborn et al. 2010), it is likely that these mutations altered ubiquitination sites and/or the interaction between Pdx1 and ubiquitin pathway proteins.

The only known exceptions to the extreme conservation of vertebrate *Pdx1* are rodents of the gerbil subfamily. It was recently shown that *Pdx1* is highly divergent in two gerbil species: *Psammomys obesus* (sand rat) and *Meriones unguiculatus* (Mongolian gerbil) (Hargreaves et al. 2017). In these animals, the normally invariant 60 amino acid homeodomain has experienced 15 and 14 amino acid changes, along with drastic change outside the homeodomain. This implies that radical Pdx1 sequence divergence occurred in the ancestry of gerbils. Although amino acid changes would normally be indicative of either reduced selection pressure or positive adaptive change, neither seems likely in this case. Instead, detailed analysis in *Psammomys obesus* reveals that the changes have been driven by extreme accumulation of GC nucleotides across a very extensive genomic region, possibly by gBGC, and that the GC bombardment of the gene may have driven maladaptive changes to fixation. Similar substitutions are not seen in any other animal lineage. This unusual situation gives an opportunity to examine how natural selection responds to intense mutation accumulation. First, we ask whether amino acid changes, driven by GC skew, have affected protein stability, by analogy to some human mutations. Second, we use a pharmacological inhibitor to test if, despite divergence, sand rat protein stability is regulated by the ubiquitin-proteasome system (UPS). Third, we use in vitro mutagenesis to identify which amino acid residues are targeted for ubiquitination and activation of UPS in mouse and sand rat proteins and uncover possible compensatory mutations. Our data suggest that

extreme accumulation of GC changes removed a critical and conserved ubiquitination site, and that this maladaptive change was overcome by emergence of a novel compensatory change. Hence, despite an onslaught of GC skew, some critical protein properties are maintained.

Results

Altered Stability of Gerbil Pdx1 Protein

The divergence of sand rat and Mongolian gerbil *Pdx1* homeobox sequence, compared with the usual extreme conservation, raises the question of whether the gene is nonfunctional in these species. We have previously shown that this is unlikely since the *Pdx1* coding sequence is depleted in polymorphisms relative to the 3' untranslated region, suggesting presence of purifying selection (Hargreaves et al. 2017). We also note that five of the six hexapeptide residues are unchanged. As for the homeodomain, the amino acid changes in helix 3, which makes close contact with the major groove of DNA, are between chemically similar amino acids; there is one chemically significant change (Leu to Arg) in helix 1 (fig. 1). This pattern of change suggests that the gene is functional, and that purifying selection has resisted substitutions of residues that affect interactions with DNA. The nonhelix sites that have changed in gerbils are also normally conserved, but in these cases, substitutions driven by GC skew have somehow been tolerated in gerbils only. To test if substitutions are adaptive, we analyzed the hexapeptide and homeodomain regions of the 13 species shown in figure 1, and fitted site, branch, and branch-site models using codeml in PAML (Yang 1997). We found no support for positive selection occurring within the gerbil lineage (Supplementary Material online). A two-ratio branch model gave a dN/dS ratio of 0.0197 (dN = 0.1280, dS = 6.0599) on the gerbil lineage, indicating a high rate of synonymous substitution and presence of purifying selection. We suggest the observed amino acid changes in gerbil *Pdx1* are maladaptive changes.

To assess if Pdx1 protein stability has been altered in gerbil evolution, we conducted pulse-chase labeling experiments of ectopically expressed proteins in a rat pancreatic cell line (INS-1). To allow tracking of protein degradation in individual cells, mouse and sand rat *Pdx1* genes were cloned into pSNAP_f vectors, after first lowering the GC content of the sand rat codons ("mousifying") to optimize translation efficiency (Supplementary Material online). These constructs produce Pdx1 fusion proteins with a C-terminal SNAP-tag motif, which will link covalently to a red fluorescent tag (SNAP-Cell TMR-Star) enabling protein stability to be measured after washing out unbound red tag. After transfection into INS-1 cells, pulse-chase was carried out and fluorescence intensity of individual transfected cells recorded at intervals up to 70 h (fig. 2a). Individual cells were identified under brightfield imaging, and cells that exhibited ambiguity in identity were excluded from analysis (fig. 2b). Data from 89 ($n = 43$ transfected with mouse *Pdx1*, $n = 46$ transfected with sand rat *Pdx1*) individual cells were used to plot decrease in intensity per cell, revealing a faster rate of intensity decline for mouse Pdx1 than for sand rat Pdx1 protein (fig. 2c).

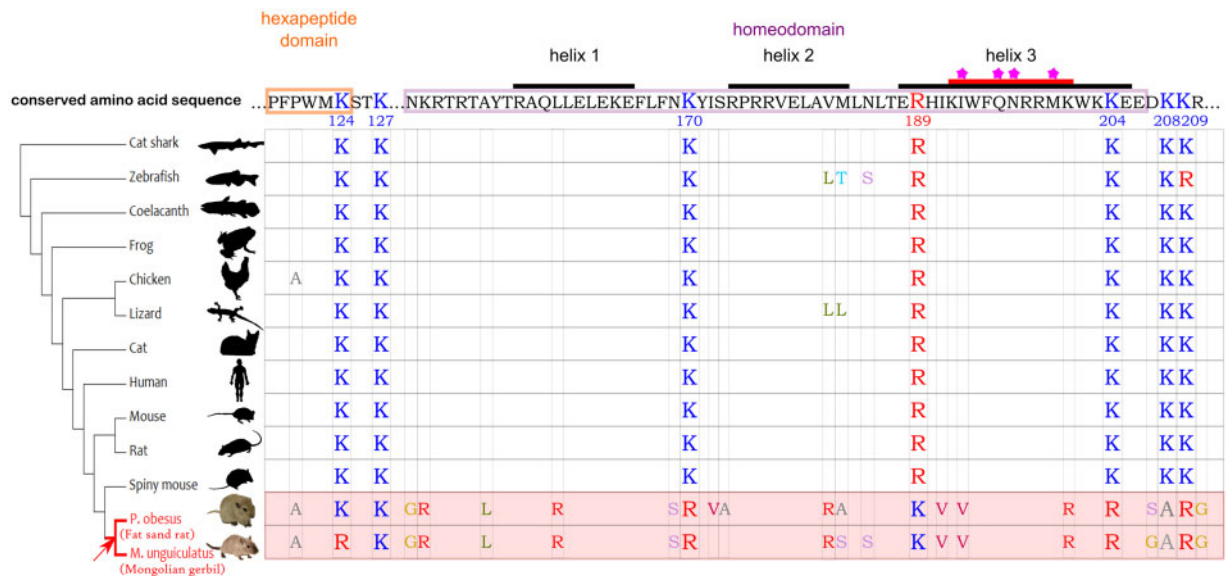


FIG. 1. Alignment of Pdx1 hexapeptide and homeodomain sequence across vertebrates. The human and mouse amino acid sequence, conserved in most vertebrates, is shown at the top. The three alpha helices are labeled, and key amino acids known to directly interact with DNA bases are highlighted with a star (Longo et al. 2007). Lysine/Arginine residues key to the current research are shown for all species. For other residues, only those that deviate from the conserved sequence are shown.

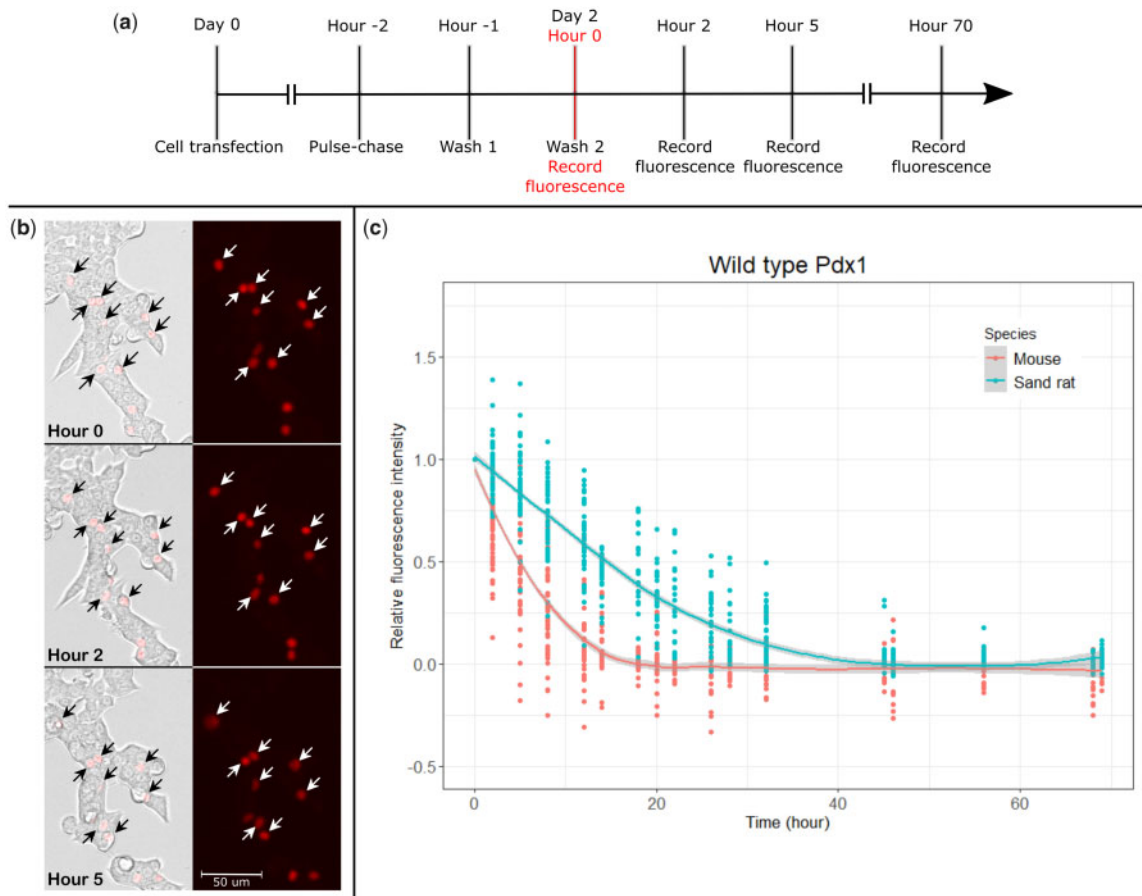


FIG. 2. Measurement of protein stability. (a) Experimental design showing timing of cell transfection, pulse-chase, and fluorescence recording. (b) An example of cell tracking over 5 h. The labeled cells are overexpressing mouse Pdx1-SNAP. (c) Intensity decline for mouse Pdx1 and sand rat Pdx1 proteins. Each dot represents an individual cell. The curve shown was generated using the Local Regression (loess) function in the ggplot2 package in R (Wickham 2016).

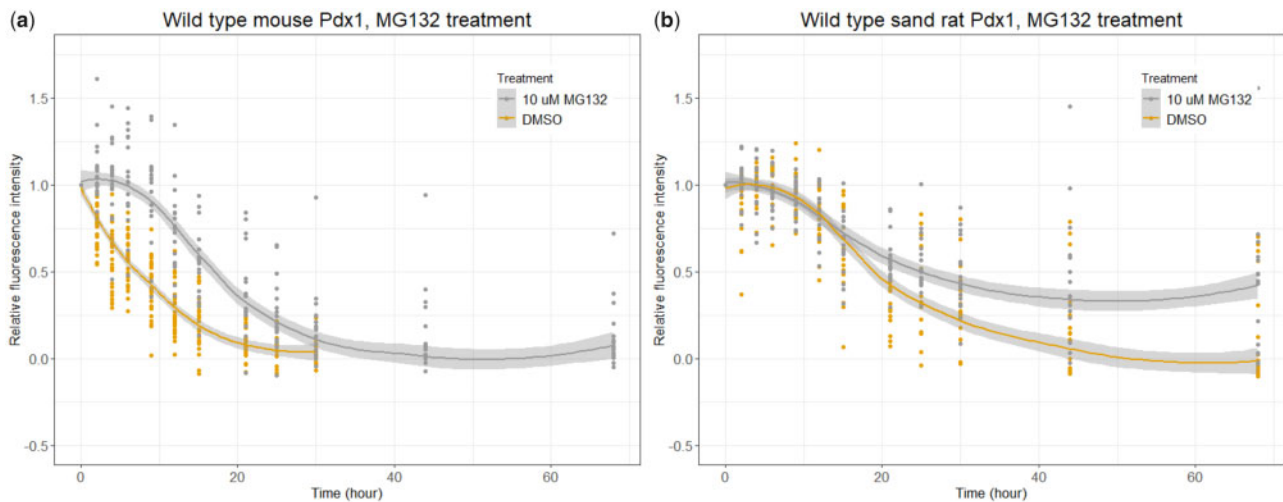


Fig. 3. Effect of inhibiting UPS on Pdx1 protein stability. Graphs show the decline in fluorescence intensity for (a) mouse Pdx1 and (b) sand rat Pdx1 in cells treated with DMSO (control) or 10- μ M MG132 (inhibitor of UPS), in both cases showing shift of protein decay curve. Statistical modeling confirmed that in both cases the effect is significant (mouse $P = 1e-04$; sand rat $P = 0.0007$); the Pdx1 protein half-lives were extended to 10.8 ± 0.9 h (mouse) and 11.3 ± 1.7 h (sand rat).

To quantify the difference in protein stability, we modeled the decrease of fluorescence intensity using the Linear and Nonlinear Mixed Effects (nlme) Models package (Pinheiro et al. 2018) in R. This revealed a significant difference between the degradation pattern of mouse Pdx1 and sand rat Pdx1 proteins ($P = 0$; Supplementary Material online). Based on this model, we estimate the half-life of mouse Pdx1 protein to be 4.7 ± 0.4 h, and the half-life of sand rat Pdx1 protein to be 13.7 ± 0.3 h.

Sand Rat Pdx1 Protein Degradation Is Regulated by the UPS

It has previously been shown that mouse Pdx1 protein stability is regulated by the UPS (Claiborn et al. 2010). To test if sand rat Pdx1 is regulated by the same biochemical pathway, we treated transfected INS-1 cells with MG132, a reversible inhibitor of the 26S proteasome. MG132 treatment slows the overall rate of mouse Pdx1 protein degradation, shifting the best-fit curve to the right (fig. 3a, $P = 1e-04$; Supplementary Material online). This is consistent with previous findings that mouse Pdx1 protein stability is regulated via UPS. We find that MG132 treatment also slows the rate of sand rat Pdx1 degradation pattern (fig. 3b, $P = 0.0007$; Supplementary Material online), although less sharply compared with mouse Pdx1. Differences in processes prior to proteasome degradation, for example, E3 ligase efficiency in recognizing and tagging the target Pdx1 protein, could contribute to subtle differences in how proteins respond to proteasome inhibition.

Identification of Pdx1 Ubiquitination Sites Suggests Compensatory Molecular Evolution

Although ubiquitin can sometimes be added to the peptide N-terminus or to cysteine residues, it is only added to lysine residues in the context of targeting for degradation in proteasomes (Kerscher et al. 2006). We therefore asked which

lysine residues were utilized by the UPS in regulating stability of mouse and sand rat Pdx1 proteins. We first used prediction software to deduce which residues had the highest probability of being a functional ubiquitination site (Chen et al. 2013; Gao et al. 2013; Wang et al. 2017). In addition, because ubiquitination acts through covalent addition of a moiety to lysine residues, we excluded key lysine residues in the homeodomain that have been proven to engage in interactions with DNA. This resulted in seven lysine residues of interest: four in the conserved Pdx1 protein (mouse, human, and most vertebrates) and three in the divergent sand rat protein.

We used in vitro mutagenesis to replace singly each lysine codon of interest with a codon for arginine, an amino acid with similar polarity but which cannot be tagged by ubiquitin. For mouse Pdx1, we generated mutant proteins K170R, K204R, K208R, and K209R with C-terminal SNAP-tags (numbering according to mouse Pdx1). These constructs were transfected into INS-1 cells and protein stability measured through pulse-chase labeling with SNAP-Cell TMR-Star. As shown in figure 4a, mutant proteins K204R, K208R, and K209R did not have significantly different half-lives compared with the wild type control ($P = 0.1487$, 0.1341 , and 0.5847 respectively; Supplementary Material online). However, mutant protein K170R showed a very different protein decay pattern, with a much flatter best-fit decay curve and a longer half-life ($P = 0.0267$; Supplementary Material online). This reveals that lysine residue 170 is the major Pdx1 ubiquitination site. We suspect it is not the only ubiquitination site as the protein is still gradually degraded within each cell.

For sand rat Pdx1, we generated SNAP-tagged mutant proteins K124R, K127R, and K189R, transfected into INS-1 cells and measured protein stability (numbering according to alignment with the mouse Pdx1 protein). As shown in figure 4b, mutant proteins K124R and K127R did not have drastically different stability compared with wild type control ($P = 0.9649$ and 0.5538 , respectively; Supplementary Material

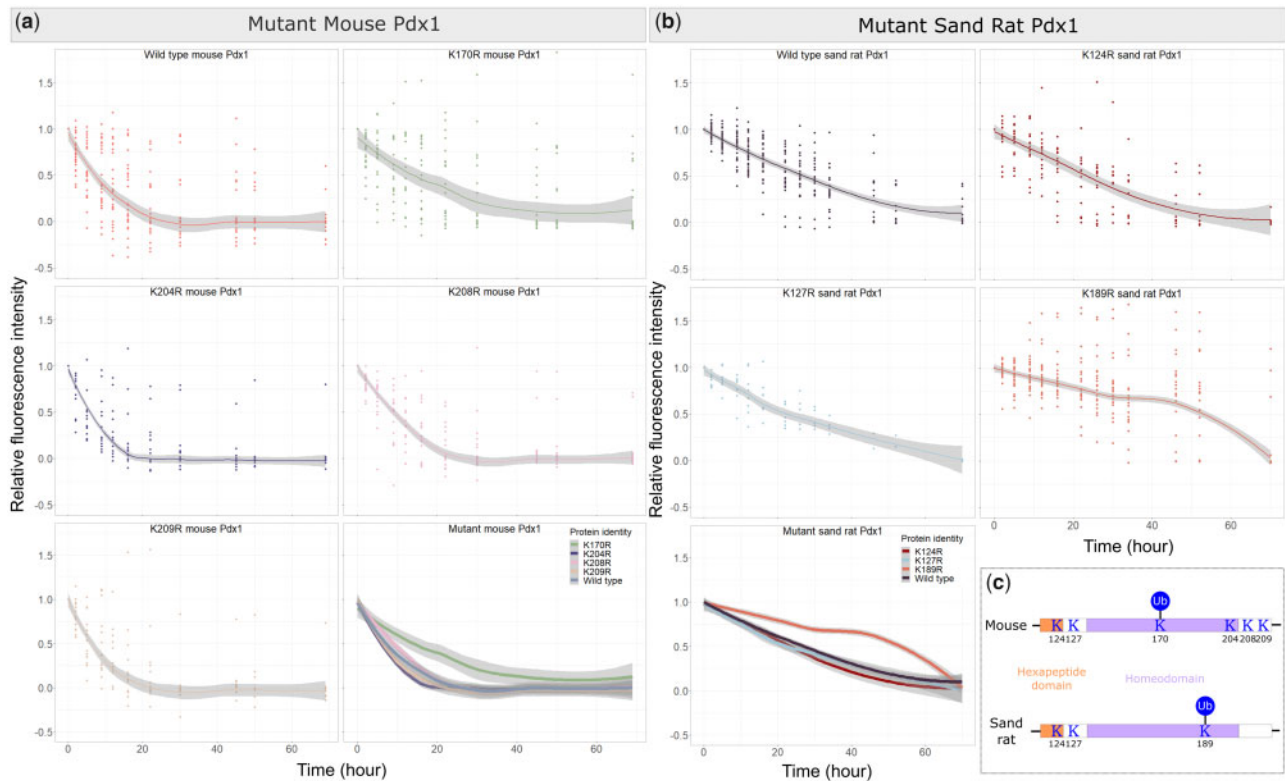


Fig. 4. Protein stability comparison between wild type and mutant proteins. (a) Graphs show decline in fluorescence intensity for mouse wild type and individual mutant proteins, with a merged graph in the bottom right corner. (b) Graphs showing decline in fluorescence intensity for sand rat wild type and individual mutant proteins, with a merged graph in the bottom left corner. (c) Summary of ubiquitination site location in mouse and sand rat Pdx1.

online). Only mutant K189R displayed a completely different degradation pattern compared with the other two mutant proteins and the wild type ($P = 0.0000$; [Supplementary Material](#) online). For K189R, high levels of fluorescence were detected in cells throughout the experimental period, with many cells exhibiting rounded shapes and detaching from the substrate around hour 50. This radical disruption of regulated protein decay indicates that the K189 is the ubiquitination site for sand rat Pdx1. Mutation of this site abolishes Pdx1 degradation via the UPS pathway.

Amino acid 170, which we identify as the major functional ubiquitination site in the mouse Pdx1 protein, has been conserved across 420 My of jawed vertebrate evolution. This K170 site is invariant across vertebrates, with the exception of some teleosts and the gerbil species discussed here. In sand rat and Mongolian gerbil, the codon has undergone a unique, GC skew associated mutation to arginine ([fig. 1](#)). In contrast, stability and degradation of sand rat Pdx1 is regulated through ubiquitination of K189, a site which across the rest of jawed vertebrate evolution is otherwise conserved as arginine. These results reveal a unique shift in the location of the ubiquitination site in Pdx1 along the evolutionary lineage leading to gerbils ([fig. 4c](#)). We propose that in the course of evolution, and under extreme GC skew, the ultra-conserved lysine residue 170, possibly along with other ubiquitination sites, was removed by mutation and as a compensatory mechanism natural selection favored an otherwise

maladaptive mutation elsewhere in the protein sequence ([fig. 5](#)). The fixation of the “new” lysine 189 is consistent with natural selection acting in the face of strong GC skew.

Discussion

The sand rat and the Mongolian gerbil are both desert-dwelling rodents, the former native to North Africa and the Middle East ([Schmidt-Nielsen et al. 1964](#)) and the latter to Mongolia and China ([Mallon 1985](#)). Both are members of the Gerbillinae subfamily that diverged from the Murinae subfamily, including mice and rats, 18 ± 1.8 Ma ([Chevret and Dobigny 2005](#)). The sand rat has been of particular interest in studies of diabetes since both wild-caught and laboratory-reared animals can exhibit elevated fasting blood glucose levels, increased body weight and symptoms of type 2 diabetes after being fed a high calorie diet ([Schmidt-Nielsen et al. 1964](#); [Shafir and Gutman 1993](#)). It had been hypothesized that these characteristics are associated with loss of the *Pdx1* gene in the gerbil subfamily ([Leibowitz et al. 2001](#)), but our recent work revealed that rather than missing the gene, sand rats and Mongolian gerbils possess a highly divergent *Pdx1* ([Hargreaves et al. 2017](#)).

Two features make the sequence divergence of gerbil Pdx1 particularly intriguing. First, the divergence of this normally highly conserved protein is unique, with nothing similar seen in other vertebrates, including arid-adapted rodents such as the closely related spiny mouse *Acomys cahirinus*. Second, the

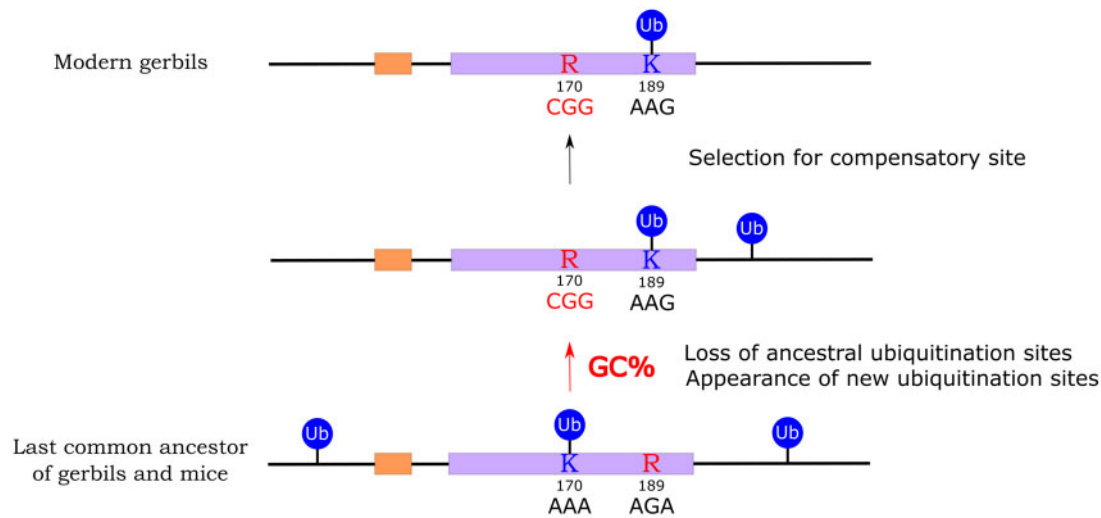


FIG. 5. Summary of the possible evolutionary pathway leading to emergence of a compensatory ubiquitination site and loss of conserved sites in gerbil *Pdx1*. This process is associated with GC skew.

divergence of the gene is clearly associated with a strongly pronounced GC skew that encompasses several genes in a large genomic region (Hargreaves et al. 2017). Our working hypothesis to explain these findings is that focused and extreme gBGC has driven accumulation of GC nucleotides throughout this region, generating maladaptive substitutions. If correct, this hypothesis predicts that natural selection would act to fix compensatory mutations, both within and outside the GC skewed region, to restore biological function.

Before considering the existence of compensatory mutations, we first ask if amino acid changes in gerbil *Pdx1* are indeed maladaptive. The evidence from comparative analysis is consistent with this view since large changes to the homeodomain, such as in gerbils, have not been tolerated across the rest of vertebrate evolution or indeed across the Bilateria. In addition, we do not find support for positive selection acting on the hexapeptide region and homeodomain in the gerbil lineage. Direct tests of function, such as gene deletion or experimental manipulation in gerbils, would not answer the question if compensatory changes have occurred elsewhere. Gene replacement in mouse would not give a clear answer for the same reason. Instead, we took a biochemical approach and asked whether protein stability was altered, as reported for some human variants associated with disruption of pancreatic function. Using a cell culture expression system, we found that sand rat *Pdx1* does have a different half-life to mouse *Pdx1*, but surprisingly the change was in the opposite direction from our original expectation. Although two different single amino acid changes in human *Pdx1* cause decreased half-life, the radically altered sand rat protein has a longer half-life. Perhaps more importantly, we noted that degradation of sand rat *Pdx1* protein follows a steady curve, indicative of actively regulated degradation rather than unregulated decay, and we obtained experimental verification that protein stability is regulated by the UPS. This regulated degradation is another indication that the sand rat *Pdx1* gene encodes a functional protein under selection rather than being a nonfunctional pseudogene. Although these assays were

conducted in a heterologous system (a rat cell line), we argue that this is not a confounding factor since the key mediators of *Pdx1* degradation, *Pdx1* C-terminal inhibiting factor 1 (*Pcif1*) and the scaffolding protein *Cul3* (Claiborn et al. 2010), are highly conserved between rat, mouse, and sand rat (>97% protein similarity between all three species).

To investigate candidate maladaptive mutations, we mapped functional sites responsible for UPS-regulated degradation of the mouse *Pdx1* protein. These mutagenesis experiments revealed that one highly conserved lysine, K170, is the critical residue for regulated degradation by UPS. Interestingly, this residue has mutated to an arginine in gerbils and could no longer be used as a site for ubiquitination. Loss of this key ubiquitination site is not seen in any other tetrapod lineage and is likely to have been driven by the gBGC phenomenon described above. However, a new ubiquitinated lysine site has evolved in gerbil *Pdx1* proteins, at a different site in the same protein domain, and it is this new site that enables UPS-regulated degradation. We suggest that the maladaptive change to a key amino acid residue, conserved for over 420 My, is compensated for by emergence of a new site fulfilling a similar function. We cannot state that the sand rat *Pdx1* protein is as efficient or “optimal” as the conserved or ancestral sequence present before the emergence of the hotspot of GC skew. Indeed, amino acid changes at other residues may affect other biochemical properties. However, we propose that as GC skew gradually altered the sequence of this critical gene during gerbil evolution, natural selection favored substitutions elsewhere in the protein to compensate for loss of a critical biochemical function (fig. 5). This gives us an insight into the power of natural selection even in the face of extreme and unusual nucleotide bias.

Materials and Methods

Plasmid Construction and In Vitro Mutagenesis

Full coding sequences for mouse and sand rat *Pdx1* genes were synthesized by GenScript, with alterations in sand rat

codon usage to lower GC content ([Supplementary Material online](#)). Mouse *Pdx1* was inserted into pSNAP_f vector (New England BioLabs) using *Bam*HI and *Xho*I restriction sites, and sand rat *Pdx1* inserted using *Not*I and *Xho*I restriction sites. Stop codons were removed to allow expression of Pdx1-SNAP fusion protein. Successful insertion was verified by Sanger sequencing. In vitro mutagenesis was carried out using polymerase chain reaction.

Cell Culture and Transfection

INS-1 cells (below 30 passages) were maintained in RPMI-1640 medium, supplemented with 10% fetal bovine serum, 50- μ M beta-mercaptoethanol, 1-mM sodium pyruvate, 10-mM HEPES, and 100-U/ml penicillin/streptomycin, at 37 °C/5% CO₂ and passaged every 4 days. For transfection, cells were seeded into 96-well plates at 60,000 cells/well and transfected 5 h later using Lipofectamine 3000 (ThermoFisher Scientific). For each well, 40- μ l DNA-lipid mix containing 400-ng plasmid DNA, 0.8- μ l P3000 Reagent, and 1.2- μ l Lipofectamine was added and incubated overnight.

SNAP-Staining and Fluorescence Recording

SNAP-Cell TMR-Star dye (New England BioLabs) was diluted 1:400 in culture medium. Fifty microliters was added to each well and incubated at room temperature for 5 min before three washes. Stained cells were incubated in fresh culture medium for 1 h at 37 °C, twice, to ensure dissociation of unbound SNAP-Cell TMR-Star dye prior to analysis. At each timepoint, pictures were taken using a ZOE Fluorescent Cell Imager (Bio-Rad) with fixed settings: brightfield gain = 18, exposure = 300, LED = 40, contrast = 0; red channel gain = 40, exposure = 500, LED = 50, contrast = 0. Image analysis was performed using ImageJ ([Schneider et al. 2012](#)).

MG132 Treatment

Ten-millimolar MG132 stock solution in dimethyl sulfoxide (DMSO, Sigma-Aldrich) was diluted to 10- μ M in treatment wells at hour -2. Culture medium containing 10- μ M MG132 was used for all medium changes until hour 4, giving a total of 6-h treatment. Controls were treated in parallel using DMSO.

Statistical Analysis

We used nlme 3.1-137 ([Pinheiro et al. 2018](#)) in R 3.5.1 ([R Core Team 2018](#)) to model the pattern of protein degradation based on data collected for all protein and treatment types. Square-root transformed data were used to ensure no obvious deviations from homoscedasticity or normality ([Supplementary Material online](#)). The linear mixed effects model was constructed using the interaction between sqrt(Time) and type of species as the fixed effect, and the random effect was cell identity. Different experiments were analyzed individually using the same fixed and random effects. Parameters reported are given in [Supplementary Material online](#). Half-life predictions were made for each individual cell using the model generated and the generic predict function in R ([R Core Team 2018](#)); half-lives of mouse and sand rat *Pdx1* proteins are reported as mean \pm standard deviation.

Supplementary Material

[Supplementary data](#) are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Sonia Trigueros for advice on microscopy, Anna Gloyn and Natasha Ng for INS-1 cell culture advice, Daniel Lunn (University of Oxford Statistical Consultancy Services) for advice on data analysis, Rodrigo Pracana, Amy Royall, Adam Hargreaves, John Mulley, Thomas Brekke, Sebastian Shimeld, and Gil McVean and for helpful discussions, Ofuase Dirisu for assistance, and reviewers for useful suggestions. This work was supported by the Elizabeth Hannah Jenkinson Fund (to Y.D.), the Rhodes Trust (to Y.D.), and a Leverhulme Trust Research Project Grant (RPG-2017-321 to P.W.H.H.).

References

- Ahlgren U, Jonsson J, Edlund H. 1996. The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. *Development* 122(5):1409–1416.
- Ahlgren U, Jonsson J, Jonsson L, Simu K, Edlund H. 1998. β -Cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the β -cell phenotype and maturity onset diabetes. *Genes Dev.* 12(12):1763–1768.
- Arbeithuber B, Betancourt AJ, Ebner T, Tiemann-Boege I. 2015. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. *Proc Natl Acad Sci U S A.* 112(7):2109–2114.
- Brooke NM, Garcia-Fernández J, Holland PWH. 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* 50(2):61–63.
- Chen Z, Zhou Y, Song J, Zhang Z. 2013. HCKSAAP-UbSite: improved prediction of human ubiquitination sites by exploiting amino acid pattern and properties. *Biochim Biophys Acta Proteins Proteomics* 1834(8):1461–1467.
- Chevret P, Dobigny G. 2005. Systematics and evolution of the subfamily Gerbillinae (Mammalia, Rodentia, Muridae). *Mol Phylogenet Evol.* 35(3):674–688.
- Claiborn KC, Sachdeva MM, Cannon CE, Groff DN, Singer JD, Stoffers DA. 2010. Pcf1 modulates Pdx1 protein stability and pancreatic β cell function and survival in mice. *J Clin Invest.* 120(10):3713–3721.
- Ehrlich M, Wang R-H. 1981. 5-Methylcytosine in eukaryotic DNA. *Science* 212(4501):1350.
- Eyre-Walker A, Hurst LD. 2001. The evolution of isochores. *Nat Rev Genet.* 2(7):549–555.
- Ferrier DEK, Holland P. 2002. *Ciona intestinalis* ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. *Mol Phylogenet Evol.* 24(3):412–417.
- Figuet E, Ballenghien M, Romiguier J, Galtier N. 2015. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biol Evol.* 7(1):240–250.
- Galtier N. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Dev Psychobiol.* 39(4):251–256.
- Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* 23(6):273–277.
- Gao T, Liu Z, Wang Y, Cheng H, Yang Q, Guo A, Ren J, Xue Y. 2013. UUCD: a family-based database of ubiquitin and ubiquitin-like conjugation. *Nucleic Acids Res.* 41(D1):445–451.
- Hargreaves AD, Zhou L, Christensen J, Marlétaz F, Liu S, Li F, Jansen PG, Spiga E, Hansen MT, Pedersen SVH, et al. 2017. Genome sequence of

- a diabetes-prone rodent reveals a mutation hotspot around the ParaHox gene cluster. *Proc Natl Acad Sci U S A* 114(29):7677–7682.
- Jonsson J, Carlsson L, Edlund T, Edlund H. 1994. Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 371(6498):606–609.
- Kerscher O, Felberbaum R, Hochstrasser M. 2006. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol*. 22(1):159–180.
- Leibowitz G, Ferber S, Apelqvist Å, Edlund H, Gross DJ, Cerasi E, Melloul D, Kaiser N. 2001. IPF1/PDX1 deficiency and β -cell dysfunction in *Psammomys obesus*, an animal with type 2 diabetes. *Diabetes* 50(8):1799–1806.
- Liberzon A, Ridner G, Walker MD. 2004. Role of intrinsic DNA binding specificity in defining target genes of the mammalian transcription factor PDX1. *Nucleic Acids Res*. 32(1):54–64.
- Longo A, Guanga GP, Rose RB. 2007. Structural basis for induced fit mechanisms in DNA recognition by the Pdx1 homeodomain. *Biochemistry* 46(11):2948–2957.
- Lu M, Miller C, Habener JF. 1996. Functional regions of the homeodomain protein IDX-1 required for transactivation of the rat somatostatin gene. *Endocrinology* 137(7):2959–2967.
- Macfarlane WM, Frayling TM, Ellard S, Evans JC, Allen LIS, Bulman MP, Ayres S, Shepherd M, Clark P, Millward A. 1999. Missense mutations in the insulin promoter factor-1 gene predispose to type 2 diabetes. *J Clin Invest*. 104: R33–R39.
- Mallon DP. 1985. The mammals of the Mongolian People's Republic. *Mamm Rev*. 15(2):71–102.
- Petrov DA, Hartl DL. 1999. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. *Proc Natl Acad Sci U S A*. 96(4):1475–1479.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2018. nlme: linear and nonlinear mixed effects models. R package version 3.1-137. Available from: <https://CRAN.R-project.org/package=nlme>
- R Core Team. 2018. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.r-project.org/>
- Schaibley VM, Zawistowski M, Wegmann D, Ehm MG, Nelson MR, St. Jean PL, Abecasis GR, Novembre J, Zollner S, Li JZ. 2013. The influence of genomic context on mutation patterns in the human genome inferred from rare variants. *Genome Res*. 23(12):1974–1984.
- Schmidt-Nielsen K, Haines HB, Hackel DB. 1964. Diabetes mellitus in the sand rat induced by standard laboratory diets. *Science* 143(3607):689–690.
- Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH image to ImageJ: 25 years of image analysis. *Nat Methods*. 9(7):671–675.
- Schwitzgebel VM, Mamin A, Brun T, Ritz-Laser B, Zaiko M, Maret A, Jornayvaz FR, Theintz GE, Michielin O, Melloul D, et al. 2003. Agenesis of human pancreas due to decreased half-life of insulin promoter factor 1. *J Clin Endocrinol Metab*. 88(9):4398–4406.
- Shafir E, Gutman A. 1993. *Psammomys obesus* of the Jerusalem colony: a model for nutritionally induced, non-insulin-dependent diabetes. *J Basic Clin Physiol Pharmacol*. 4(1–2):83–100.
- Wang JR, Huang WL, Tsai MJ, Hsu KT, Huang HL, Ho SY. 2017. ESA-UbiSite: accurate prediction of human ubiquitination sites by identifying a set of effective negatives. *Bioinformatics* 33(5):661–668.
- Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.
- Wright CVE, Schnegelsberg P, Robertis EDE. 1989. XlHbox 8: a novel *Xenopus* homeo protein restricted to a narrow band of endoderm. *Development* 105(4):787–794.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13(5):555–556.