

ered common in real food webs (6–9). The cascade model assumes a hierarchical relationship between species that constrains the trophic role as consumer or resource: hierarchically higher species are always feeding on species with a lower ranking. This model cannot generate any loops, but explains some important properties of natural food webs such as the species richness among trophic levels (18).

20. E. A. Bernays, D. J. Funk, *Proc. R. Soc. London Ser. B* **266**, 151 (1999).

21. H. Matsuda, M. Hori, P. A. Abrams, *Evol. Ecol.* **10**, 13 (1996).

22. W. F. Humphreys, *J. Anim. Ecol.* **48**, 427 (1979).

23. C. T. Robbins, *Wildlife Feeding and Nutrition* (Academic Press, London, 1983).

24. Iterated simulation runs with randomly chosen parameter sets ( $r_i = 0.0$  to  $0.1$ ,  $f_{ij} = 0.0$  to  $1.0$ ,  $s_i = 1.0$ ), initial abundances ( $X_i = 0.0$  to  $0.1$ ) and linkage patterns, and measured the frequency of all species co-occurring ( $X_i > 10^{-13}$  for all  $i$ ) after sufficiently long periods ( $t = 10^5$  by which time community

persistence reaches an asymptote) in  $10^4$  runs. As in other models (1, 3–5), every species persists (with positive equilibrium biomass,  $r_i / s_i$ ) in the absence of interspecific interaction (food web being always persistent when  $C = 0$ ), they are either autotrophs or use external resources (35). This assumption avoids a confounding effect (36) of an increase in interspecific links decreasing the number of heterotrophic species with no potential diet present in the web.

25. R. Dukas, E. A. Bernays, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2637 (2000).

26. M. Egas, M. W. Sabelis, *Ecol. Lett.* **4**, 190 (2001).

27. M. Hori, *Science* **260**, 216 (1993).

28. G. P. Closs, P. S. Lake, *Ecol. Monogr.* **64**, 1 (1994).

29. A. F. Tavares-Cromar, D. D. Williams, *Ecol. Monogr.* **66**, 91 (1996).

30. R. T. Paine, *Nature* **355**, 73 (1992).

31. D. G. Raffaelli, S. J. Hall, in *Food Webs: Integration of Pattern and Dynamics*, G. A. Polis, K. O. Winemiller, Eds. (Kluwer Academic, Norwell, MA, 1996), pp. 185–191.

32. J. T. Wootton, *Ecol. Monogr.* **67**, 45 (1997).

33. N. G. Hairston Jr. et al. *Nature* **401**, 446 (1999).

34. K. Schoenly, J. E. Cohen, *Ecol. Monogr.* **61**, 267 (1991).

35. G. A. Polis, S. D. Hurd, in *Food Webs: Integration of Pattern and Dynamics*, G. A. Polis, K. O. Winemiller, Eds. (Kluwer Academic, Norwell, MA, 1996), pp. 275–285.

36. D. L. DeAngelis, *Ecology* **56**, 238 (1975).

37. I thank H. Jones, N. Yamamura, and G. Takimoto for comments. Study partly supported by Japan Society for the Promotion of Science Research Fellowship for Young Scientists.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/299/5611/1388/DC1  
SOM Text  
Figs. S1 and S2  
Reference  
Movie S1

7 October 2002; accepted 23 January 2003

## Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome

Dario Boffelli,<sup>1,2</sup> Jon McAuliffe,<sup>3</sup> Dmitriy Ovcharenko,<sup>2</sup> Keith D. Lewis,<sup>2</sup> Ivan Ovcharenko,<sup>1,2</sup> Lior Pachter,<sup>4</sup> Edward M. Rubin<sup>1,2\*</sup>

Nonhuman primates represent the most relevant model organisms to understand the biology of *Homo sapiens*. The recent divergence and associated overall sequence conservation between individual members of this taxon have nonetheless largely precluded the use of primates in comparative sequence studies. We used sequence comparisons of an extensive set of Old World and New World monkeys and hominoids to identify functional regions in the human genome. Analysis of these data enabled the discovery of primate-specific gene regulatory elements and the demarcation of the exons of multiple genes. Much of the information content of the comprehensive primate sequence comparisons could be captured with a small subset of phylogenetically close primates. These results demonstrate the utility of intraprimate sequence comparisons to discover common mammalian as well as primate-specific functional elements in the human genome, which are unattainable through the evaluation of more evolutionarily distant species.

Genomic sequence comparisons between distant species have been extensively used to identify genes and determine their intron-exon boundaries, as well as to identify regulatory elements present in the large noncoding fraction of the genome (1–3). This strategy has been successful in human-mouse comparisons, because the ~75 million years (My) of separation from their last common ancestor have provided sufficient time for a large fraction of nucleotides to have been exposed to considerable mutation and selection pressure. Although such comparisons readily identify regions of the human genome performing general biological functions

shared with evolutionarily distant mammals, they will invariably miss recent changes in DNA sequence that account for uniquely primate biological traits.

As a consequence of their short evolutionary separation (apes 6 to 14 My, Old World monkeys 25 My, New World monkeys 40 My) (4), there is a paucity of sequence variation between humans and each of their nearest primate relatives. This lack makes it difficult to distinguish functional from passive conservation on the basis of pairwise comparisons, thus limiting the usefulness of such comparisons. However, the additive collective divergence of higher primates as a group (fig. S1) is comparable to that of humans and mice. This suggests that deep sequence comparisons of numerous primate species should be sufficient to identify important regions of conservation that encode functional elements. Phylogenetic footprinting (5, 6) has been used to identify highly conserved putative

regulatory elements, exploiting alignments across numerous evolutionarily distant species. We developed a variant of phylogenetic footprinting, which we termed phylogenetic shadowing. In contrast to footprinting, phylogenetic shadowing examines sequences of closely related species and takes into account the phylogenetic relationship of the set of species analyzed. This approach enabled the localization of regions of collective variation and complementary regions of conservation, facilitating the identification of coding as well as noncoding functional regions.

We first examined the ability of this strategy to identify functional regions with precise locations within the human genome, such as intron-exon boundaries. The lack of clone-based libraries for multiple primate species limited us to sequencing orthologous regions from a large set of primates, using genomic DNA as template (7). The sole criterion used in the selection of the four different regions that we studied was that each should contain at least one annotated exon. The sequences were generated (8) for a set of 13 to 17 primate species that included those evolutionarily closest to humans, such as Old World and New World monkeys and hominoids, but not distant primates such as prosimians. The resulting sequences were analyzed to determine the likelihood ratio under a fast-versus slow-mutation regime for each aligned nucleotide site across all four regions analyzed (supporting online text) (8). This ratio represents the relative likelihood that any given nucleotide site was subjected to a faster or slower rate of accumulation of variation and is related to functional constraints imposed on each site. The corresponding likelihood ratio curves were used to describe the variation profile of the four genomic intervals analyzed.

In all regions examined, the exon-containing sequences displayed the least amount of cross-species variation, in agreement with the constraint imposed by their functional role (Fig. 1, A to D). A limited number of short regions of minimal variation similar to the exon-containing sequences appeared in the likelihood plots.

<sup>1</sup>U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA. <sup>2</sup>Department of Genome Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. <sup>3</sup>Department of Statistics, <sup>4</sup>Department of Mathematics, University of California, Berkeley, CA 94720, USA.

\*To whom correspondence should be addressed. E-mail: emrubin@lbl.gov

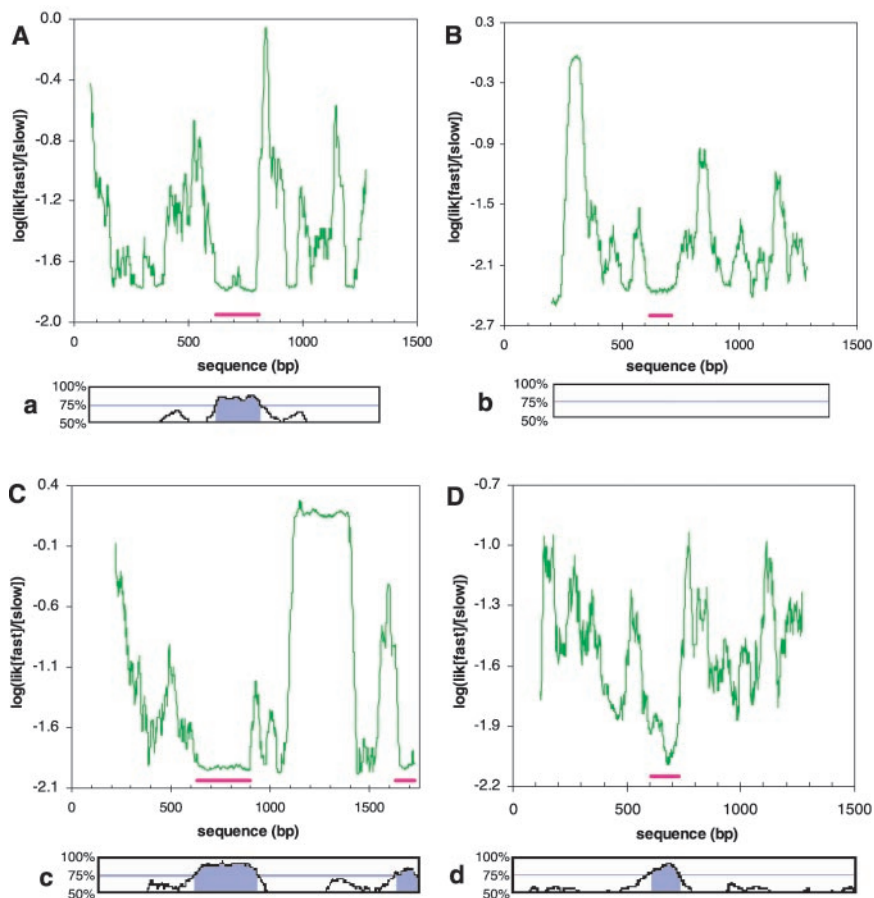
## REPORTS

**Table 1.** The discriminative power of phylogenetic shadowing at increasing species subset sizes. Each row is an exonic region under study. Each column shows, by region, the percentage of total phylogenetic divergence in a complete species set of that region that is captured by the most-divergent subset of the indicated size. The minimum number of species required to capture at least 75% of the full power of the approach is highlighted in bold for each gene.

Gene	Subset size (number of species)							
	1	2	3	4	5	6	7	8
LXR- $\alpha$	0	38	54	65	72	<b>77</b>	81	88
APO-B	0	39	54	64	73	<b>79</b>	83	88
CETP	0	45	69	<b>77</b>	83	88	90	93
PLG	0	35	47	58	64	72	<b>79</b>	84

These regions, however, were all less than 50 base pairs (bp) long, making them unlikely to be candidate exons for alternative splicing (the vast majority of internal exons predominantly range between 50 and 200 bp in length) (9). These regions may represent previously unidentified regulatory elements. In agreement with observations from other parts of the genome (10), the four sequenced regions have evolved at different rates, as indicated by their differing absolute likelihoods. Comparison of human-mouse versus the multiple-primate visualization plots (Fig. 1, a to d) yields similar results, with exons showing the highest level of conservation in all regions studied. The primate sequence comparisons illustrate the effectiveness of phylogenetic shadowing in yielding a precise identification of the exon boundaries. In the case of the cholesterol ester transfer protein gene (Fig. 1, B and b), the human-mouse comparison is unavailable, because this gene has been inactivated in the mouse. This inactivation represents an extreme example of the frequently encountered situation where the mouse genome, owing to the lack of meaningful alignments between human and mouse sequences, fails to localize functional elements in the human genome (11).

We next analyzed the sequence data for the four regions studied to assess the contribution of additional species to the discriminative power of phylogenetic shadowing and to identify the most informative minimum subset of species (8). The sequences from the most informative subset of only four to seven species, depending on the genomic locus, were calculated to be sufficient to capture ~75% of the total available discriminative power of this approach (Table 1). We were able to unequivocally identify the position of exon 3 of liver X receptor- $\alpha$  from the five most informative species (*Homo*, *Saguinis*, *Colobus*, *Callicebus*, *Allenopithecus*), as indicated by the likelihood curve (fig. S2). Comparison with Fig. 1C shows that additional species only marginally improve this plot. As would be predicted, the species included in the set maximizing the discriminative power



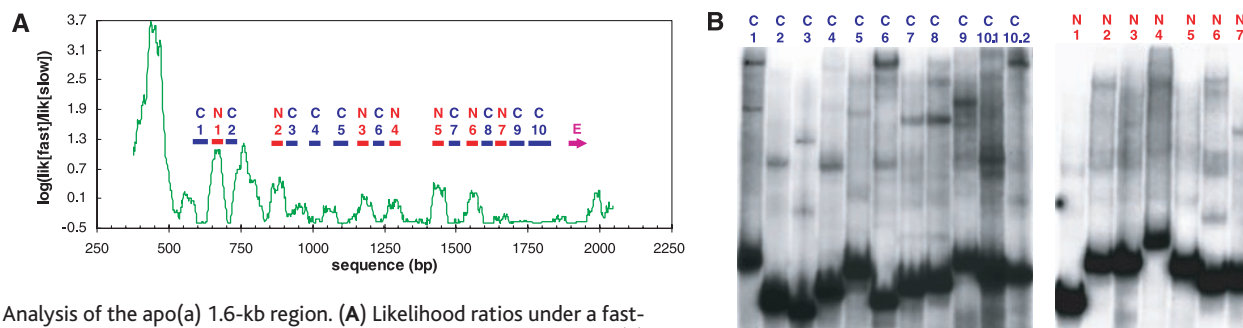
**Fig. 1.** Likelihood ratios under a fast- versus slow-mutation regime for genomic intervals containing apo-B exon 19 (A), CETP exon 8 (B), LXR- $\alpha$  exon 3 (C), and plasminogen exon 6 (D). The x axis represents the position in the multiple alignment consensus sequence, and the y axis the log likelihood ratio at that position. The plot is smoothed by means of a 20%-trimmed mean over the 50-base window centered at each aligned site. A lower ratio indicates a higher degree of constraint on mutability of that site. The position of the exon in each sequence is shown by the magenta line under the green ratio curve. The LXR- $\alpha$  plot (C) contains a fragment of an additional exon at the right end of the plot. Panels (a) to (d) show the VISTA conservation plots for the corresponding orthologous regions in human and mouse. The y axis represents the percentage of sequence conservation and the blue area the position of the exon.

of phylogenetic shadowing include representatives of the different clades that compose the primate phylogenetic tree and therefore constitute the least related species in the set studied (fig. S1).

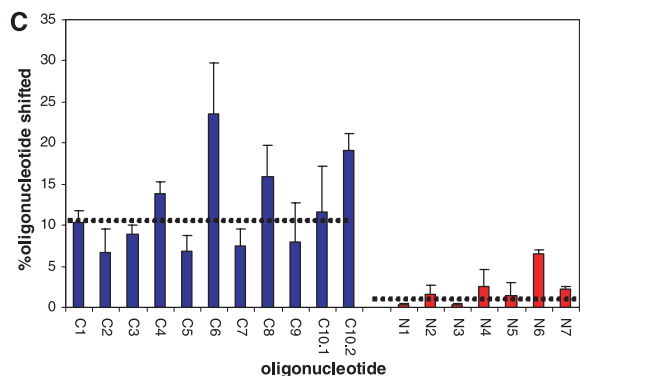
To investigate the ability of phylogenetic shadowing to discover primate-specific gene-regulatory elements, we studied apolipoprotein (a) [apo(a)], a recently evolved primate gene whose orthologous distribution is limited to Old World monkeys and hominoids (12). Defining the regulatory sequences determining apo(a) expression levels is of considerable biomedical relevance, because high plasma levels of this protein serve as an important cardiovascular disease risk predictor (13, 14). The sparse distribution of this gene among mammals precludes classical comparative genomic approaches. Sequence comparison of the apo(a) locus in humans, chimps, and baboons revealed a region of extreme conservation of ~1.6 kb adjacent to the transcription start site, surrounded by ~8 kb

of reduced conservation on either side, owing to the different patterns of repetitive-element insertion in the three species. We sequenced and analyzed this 1.6-kb region in 18 Old World monkeys and hominoids (8). Phylogenetic shadowing revealed regions of varying conservation similar to that observed for the intron-exon regions described previously (Fig. 2A). In addition to the region containing the first exon of apo(a) (region E) (15), the remaining regions with the lowest variation include the apo(a) promoter's TATA box (region C9) and a critical hepatocyte nuclear factor (HNF)-1 $\alpha$  transcription factor binding site (region C10), both of which have been functionally described (16). Eight additional regions, varying in size between 40 and 70 bp, show levels of conservation comparable to that of these three functional regions, suggesting they are also biologically important.

Because these elements were located immediately upstream of the apo(a) promoter, we



**Fig. 2.** Analysis of the apo(a) 1.6-kb region. **(A)** Likelihood ratios under a fast-versus slow-mutation regime for the genomic interval containing the apo(a) exon 1 and 5'-flanking sequence. The x axis represents the position in the multiple alignment consensus sequence, and the y axis the log likelihood ratio at that position. The plot is smoothed using a 20%-trimmed mean over the 50-base window centered at each aligned site. The position of the exon (E) is shown by the magenta arrow. The blue (C1 to C10) and red (N1 to N7) rectangles show the sequence intervals that were selected as the representatives of conserved and nonconserved regions, respectively. Delineation of regions C9 and C10 was based on previous knowledge of their functional role. These regions were investigated by gel-shift and transfection analysis. **(B)** Electrophoretic mobility-shift patterns of the conserved (C1 to C10) and nonconserved (N1 to N7) intervals. Region C10, being too long (97 bp) for gel-shift analysis, was split into two 60-bp overlapping oligonucleotides (C10.1 and C10.2). **(C)** Densitometric analysis of the electrophoretic mobility-shift patterns. The y axis represents the amount of shifted oligonucleotide normalized to the total amount of oligonucleotide in the lane. Columns and error bars represent the mean and 1 SD, respectively, of five (C1 to C10.2) and three (N1 to N7) independent gel-shift experiments. The dotted line indicates the median electrophoretic mobility-shift of the conserved and nonconserved regions.



**Fig. 3.** Transfection analysis of conserved and nonconserved regions in the 1.6-kb apo(a) region. The colored bars represent expression values of luciferase reporter vectors carrying individual deletions from the whole 1.6-kb apo(a) region of one of the conserved (blue bars) or nonconserved (red bars) regions. Luciferase values are normalized by  $\beta$ -galactosidase expression and reported as percentages of the expression of the whole 1.6-kb apo(a) region. Each bar represents the average of three independent transfection experiments, each determined in quadruplicate. Error bars indicate 1 SD. The dotted line indicates the median expression of the conserved and nonconserved regions.

cell line. The 10 most conserved regions, as well as the 7 least conserved regions, were individually deleted from the whole 1.6-kb region (8). These constructs were compared with the intact 1.6-kb region construct for their ability to drive the expression of a luciferase reporter gene. Three independent experimental determinations reproducibly indicated that conserved regions had a much larger functional impact than did nonconserved regions in these reporter-gene expression assays (8). The deletions of the conserved sequence elements reduced the expression of the full reporter by 25 to 55% (17), with the exception of the construct carrying the deletion of region 6 (Fig. 3). In contrast, deletions of all but one of the seven nonconserved regions had a minimal effect on the expression of the full reporter construct, reducing its levels by a median value of only 4%.

tested these sequences for their ability to be recognized by DNA binding proteins. The highly conserved regions were predicted to interact with such proteins more efficiently than were regions characterized by a low degree of conservation. We performed electrophoretic mobility-shift assays on 10 oligonucleotides spanning the most-conserved regions and, as putative negative controls, seven similarly sized oligonucleotides representing regions with the least conservation (Fig. 2A, regions N1 to N7). In this assay, nuclear extracts from a liver cell line were mixed with radiolabeled oligonucleotides and separated by electrophoresis (8). All oligonucleotides from the conserved regions interacted with one or more DNA binding proteins, as reflected by the slower electrophoretic mobility

of the protein-bound oligonucleotides relative to the mobility of the pure oligonucleotides (Fig. 2B, lower bands). Conversely, oligonucleotides from nonconserved regions showed only weak or no protein binding. Overall, binding of oligonucleotides from conserved regions with hepatic nuclear proteins was more than six times as strong as binding of oligonucleotides from regions with lower conservation (Fig. 2C, regions C1 to C10.2 versus N1 to N7).

Consistent with the prediction that conserved regions interact preferentially with DNA binding proteins is the prediction that these regions are involved in transcriptional regulation activity. To explore this possibility, we developed an assay for *in vitro* enhancer activity based on transient transfection of a human liver

The analysis of closely related primates facilitates the discovery of functional elements specific to primates as well as elements shared with evolutionarily distant mammals. The high absolute degree of similarity minimizes ambiguity in the computation of the multiple alignment, which in turn greatly facilitates the subsequent construction of the phylogenetic tree (18). Furthermore, the generalization of gene-finding algorithms to multiple organism annotation is simplified, which results in more accurate predictions (supporting online text). The facility of sequence alignments and the possibility of comparative assembly for nonhuman primates, with human as the reference sequence, have important

practical relevance, because they considerably diminish the depth of sequence coverage required for an organism to be informative in annotating the human genome.

Evaluation of our data suggests that sequence from as few as four to six primate species in addition to humans is sufficient for the identification of a large fraction of functional elements in the human genome, many of which are likely to be missed by human-mouse comparisons. With the goal of annotating the human genome, these studies indicate that the generation of a limited amount of sequence from a few additional properly selected primate genomes will provide information previously unavailable from comparisons of humans with more evolutionarily distant species.

References and Notes

1. G. G. Loots *et al.*, *Science* **288**, 136 (2000).
2. L. A. Pennacchio *et al.*, *Science* **294**, 169 (2001).
3. L. A. Pennacchio, E. M. Rubin, *Nature Rev. Genet.* **2**, 100 (2001).
4. M. Goodman, *Am. J. Hum. Genet.* **64**, 31 (1999).
5. D. A. Tagle *et al.*, *J. Mol. Biol.* **203**, 439 (1988).
6. D. L. Gumucio *et al.*, *Mol. Cell. Biol.* **12**, 4919 (1992).
7. Genomic DNA from the following species was analyzed in this study: *Pan troglodytes* (\*), *Pan paniscus* (\*), *Gorilla gorilla* (\*), *Pongo pygmaeus* (\*), *Hylobates lar lar* (#), *Hylobates syndactylus* (#), *Papio hamadryas*, *Macaca nemestrina* (\*), *Macaca mulatta* (\*), *Mandrillus leucophaeus* (#), *Cercopithecus hamlyni* (#), *Lophocebus albigena* (#), *Allenopithecus nigroviridis* (#), *Miopithecus talapoin* (#), *Cercopithecus aethiops* (#), *Erythrocebus patas* (\*), *Colobus guereza kikuyuensis* (#), *Pygathrix nemaeus* (#), *Trachypithecus francoisi* (#), *Presbytis entellus* (#), *Nasalis larvatus* (#), *Ateles geoffroy* (#), *Callicebus moloch* (#), *Alouatta seniculus* (#), *Saimiri sciureus* (#), *Aotus trivirgatus* (#), *Callithrix* (#), and *Saguinus labiatus* (#). DNA for *Papio hamadryas* was from the RPCI-41 Baboon BAC Library ([www.chori.org](http://www.chori.org)). DNA for species marked with \* was from the Coriell Cell Repositories (<http://locus.umdnj.edu>). DNA for species marked with # was provided by the San Diego Zoo/Center for the Reproduction of Endangered Species.
8. Materials and Methods are available as supporting material on Science Online.
9. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
10. I. Ebersberger, D. Metzler, C. Schwarz, S. Paabo, *Am. J. Hum. Genet.* **70**, 1490 (2002).
11. Mouse Genome Sequencing Consortium, *Nature* **420**, 520 (2002).
12. R. M. Lawn *et al.*, *J. Biol. Chem.* **270**, 24004 (1995).
13. J. Danesh, R. Collins, R. Peto, *Circulation* **102**, 1082 (2000).
14. G. Luc *et al.*, *Atherosclerosis* **163**, 377 (2002).
15. This region was used to learn the mutation rates for "conserved" and "nonconserved" regimes used in this paper.
16. D. P. Wade, G. E. Lindahl, R. M. Lawn, *J. Biol. Chem.* **269**, 19757 (1994).
17. Deletion of region 9, containing the promoter's TATA box, reduces expression by only 30%, likely attributable to the presence of one of the several cryptic TATA boxes further upstream.
18. C. Notredame, *Bioinformatics* **17**, 373 (2001).
19. We thank J.-F. Cheng for support with the sequencing infrastructure, the Zoological Society of San Diego for providing primate DNA samples, and I. Udalova and L. Pennacchio for useful discussions. M. Jordan contributed useful suggestions concerning statistical methods. This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research; by the University of California, Lawrence Berkeley National Laboratory under Contract No. DEAC0376SF00098; supported by Grant #HL66728, Berkeley-PGA, under the Programs for Genomic Ap-

plication, funded by National Heart, Lung, and Blood Institute, USA. L.P. was partially supported by a grant from NIH (R01-HG02362-01).

Supporting Online Material

[www.sciencemag.org/cgi/content/full/299/5611/1391/DC1](http://www.sciencemag.org/cgi/content/full/299/5611/1391/DC1)

Materials and Methods  
SOM Text  
Figs. S1 and S2  
References

9 December 2002; accepted 14 January 2003

# EDEM As an Acceptor of Terminally Misfolded Glycoproteins Released from Calnexin

Yukako Oda,<sup>1</sup> Nobuko Hosokawa,<sup>1,2</sup> Ikuo Wada,<sup>2,3\*</sup> Kazuhiro Nagata<sup>1,2†</sup>

Terminally misfolded proteins in the endoplasmic reticulum (ER) are retrotranslocated to the cytoplasm and degraded by proteasomes through a mechanism known as ER-associated degradation (ERAD). EDEM, a postulated Man8B-binding protein, accelerates the degradation of misfolded proteins in the ER. Here, EDEM was shown to interact with calnexin, but not with calreticulin, through its transmembrane region. Both binding of substrates to calnexin and their release from calnexin were required for ERAD to occur. Overexpression of EDEM accelerated ERAD by promoting the release of terminally misfolded proteins from calnexin. Thus, EDEM appeared to function in the ERAD pathway by accepting substrates from calnexin.

The endoplasmic reticulum (ER) possesses a quality-control mechanism that discriminates correctly folded proteins from misfolded proteins (1). The misfolded proteins are dislocated to the cytosol through the translocon and degraded by the ubiquitin-proteasome system known as ER-associated degradation (ERAD) (1–4). It has been established that formation of the mannose-N-acetylglucosamine Man<sub>8</sub>GlcNAc<sub>2</sub> isomer B of N-linked oligosaccharides is the critical luminal event as a signal for ERAD (5–9). EDEM is an ER stress-inducible membrane protein with homology to  $\alpha$ -mannosidase but lacking mannosidase activity. Overexpression of EDEM greatly accelerates ERAD, most likely through recognition of the substrate-bound Man8B structure, because the mannosidase inhibitor kifnensine abolished the effects (5). However, the precise role of EDEM in the quality-control machinery is unknown.

In an attempt to search for proteins that work in concert with EDEM, we focused on calnexin (CNX), an ER protein with a potential role in the ERAD pathway (10–12). Although CNX is an abundant transmembrane component with an

affinity to monoglucosylated oligosaccharides (13, 14), its role in ERAD remains controversial (15–17). We first assessed whether CNX can physically interact with EDEM. Immunoprecipitation of cell lysates with antibodies to CNX and hemagglutinin (HA) revealed that pulse-labeled EDEM-HA was coprecipitated by the antibody to CNX (Fig. 1A). Similarly, immunoprecipitation followed by immunoblotting of cells expressing EDEM-myc also showed that EDEM-myc bound to CNX (Fig. 1B). The observed association was not a result of the chaperone function of CNX, because (i) EDEM remained in the CNX complex even after 1 hour of chase (Fig. 1A), and (ii) EDEM was still coprecipitated with CNX even after cells were treated with castanospermine, an inhibitor of glucosidase, which is known to prevent the binding of substrate proteins to CNX.

Next, we tested whether EDEM interacts with calreticulin (CRT), a soluble homolog of CNX that is present in the ER lumen (13, 14) (Fig. 2A). EDEM-HA was not coprecipitated with the antibody to CRT (Fig. 2B), suggesting that there was no interaction between EDEM and CRT. Although CNX and CRT appear to share similar functions, CNX contains a transmembrane region that CRT lacks (Fig. 2A), suggesting that the transmembrane region might be responsible for the binding of EDEM to CNX. CRT that anchored to the ER membrane by fusing it with the membrane region of CNX (CRT-CNXtm) was revealed to coprecipitate with EDEM-myc (Fig. 2C). On the contrary, soluble CNX without its cytoplasmic and transmembrane segments (FS-CNX) was not coprecipitated with EDEM-myc (Fig. 2D).

<sup>1</sup>Department of Molecular and Cellular Biology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606–8397, Japan. <sup>2</sup>Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Japan. <sup>3</sup>Department of Biochemistry, Sapporo Medical University School of Medicine, Sapporo 060–8556, Japan.

\*Present address: Institute of Biomedical Sciences, Fukushima Medical University, Fukushima 960–1295, Japan.

†To whom correspondence should be addressed. E-mail: nagata@frontier.kyoto-u.ac.jp