Retrieval-independent localization of lysyl hydroxylase in the endoplasmic reticulum via a peptide fold in its iron-binding domain

Marko SUOKAS*, Outi LAMPELA†, André H. JUFFER†, Raili MYLLYLÄ† and Sakari KELLOKUMPU*¹

*Department of Biochemistry, University of Oulu, P.O. Box 3000, FIN-90014, Oulu, Finland, and †Department of Biochemistry and the Biocenter Oulu, University of Oulu, P.O. Box 3000, FIN-90014, Oulu, Finland

Lysyl hydroxylase (LH) is a peripheral membrane protein in the lumen of the endoplasmic reticulum (ER) that catalyses hydroxylation of lysine residues in collagenous sequences. Previously, we have mapped its primary ER localization motif within a 40-amino acid segment at its C-terminus. Here, we have characterized this localization mechanism in more detail, and our results indicate that this segment confers ER residency in a KDEL-receptor-independent manner, and without any apparent recycling of the enzyme between the Golgi apparatus and the ER. In addition, we show that a rather long peptide region, rather than a specific peptide sequence *per se*, is required for efficient retention of a reporter protein in the ER. Accordingly, the minimal retention motif was found to require the last 32

C-terminal amino acids, and sequential substitution of all five charged residues within this critical segment interfered only marginally with the retention or association of the enzyme with the ER membranes. Moreover, our fold-recognition and structure-prediction analyses suggested that this critical peptide segment forms an extended loop within LH's iron-binding domain, and that this loop is exposed and readily accessible for binding. Collectively, our results define a novel retrieval-independent retention mechanism in the ER.

Key words: peripheral membrane protein, procollagen hydroxylation, protein targeting.

INTRODUCTION

In the exocytotic pathway, the endoplasmic reticulum (ER) poses a special problem in terms of protein sorting. It harbours a variety of newly synthesized soluble and membrane proteins destined for later exocytotic organelles and extracellular space, as well as its own structural and functional proteins. These include a number of luminal or membrane-bound enzymes and molecular chaperones that are segregated from their substrates by either direct retention or retrieval mechanisms. To date, two common motifs, both based on a retrieval mechanism, have been characterized. The KDEL motif at the C-terminal end of a number of luminal proteins is recognized by the KDEL receptor (erd2p) in the Golgi and the receptor–ligand complex is returned to the ER via a retrograde transport route [1]. A C-terminal double-lysine motif (KKXX or KKKX) serves a similar purpose for type I membrane proteins of the ER [2], but the flanking sequences around the motif affect the steady-state distribution within the early secretory pathway [3–5]. Retrieval of these proteins has been shown to be mediated by the coatamer complex (COP I) [6].

Recently, it has become clear that some membrane proteins of the ER also have signals for both static retention and retrieval. Cytochrome P450 2C1 has a 28-amino acid signal anchor sequence that has been shown to prevent its incorporation into forward transport vesicles, and to result in its static retention in the ER [7]. However, deletion of the 7-amino acid cytoplasmic linker sequence adjacent to the transmembrane domain, or mutation of the KQS sequence within this domain, resulted in chimaerae that exited the ER, but which were returned from the intermediate compartment by a retrieval mechanism [8]. Thus, two independent systems seem to operate for correct localization of cytochrome P450 2C1 in the ER.

Mammalian oligosaccharyltransferase (OST) complex also uses alternative ER localization mechanisms. It consists of four membrane proteins, ribophorins I and II, OST48 and DAD1. At least three of these proteins contain independent information for their correct localization in the ER. OST48 has been shown to possess a cytoplasmic double-lysine motif [9], whereas the transmembrane and cytoplasmic domains of ribophorin II appear to be crucial for ER residency [10]. Retention of ribophorin I, on the other hand, appears to be mediated solely by its luminal domain, although the exact mechanism is currently not known [9]. Additionally, protein–protein interactions within the complex itself also contribute to the steady-state localization of the OST [9].

Lysyl hydroxylase (LH; EC 1.14.11.4) is an enzyme that catalyses hydroxylation of lysyl residues in collagens [11]. So far, three different isoforms (LH1, LH2 and LH3) with LH activity have been characterized in human and mouse tissues [12–15]. Of these, the LH3 isoform has been shown to be a multifunctional enzyme, possessing also glucosyltransferase and galactosyltransferase activities [16–19]. Thus, LH3 alone appears to be able to catalyse all three consecutive steps in the formation of glycosylated hydroxylysines found in collagens [16,19]. All LH isoforms seem to utilize a novel mechanism for their localization on the luminal side of the ER membrane [20], given that they do not possess any of the known retention or retrieval motifs characterized thus far for other ER-resident proteins.

Previously, we have shown that the C-terminal end of LH1 is able to convert a cathepsin D-based reporter protein into an ER-resident protein [21]. The same peptide segment was also able to convert cathepsin D from a soluble protein into a membrane-associated protein, suggesting that this 40 amino acid segment possesses most of the targeting information needed to localize LH1 in the ER lumen. In this work, we set out to dissect the exact

Abbreviations used: COP I, coatamer complex; DAOCS, deacetoxycephalosporin C synthase; ER, endoplasmic reticulum; IPNS, isopenicillin synthase; LH, lysyl hydroxylase; 2OG-Fell, iron-dependent 2-oxoglutarate dioxygenase; OST, oligosaccharyltransferase; PDI, protein disulphide isomerase; P-3-H, prolyl-3-hydroxylase.

¹ To whom correspondence should be addressed (e-mail Sakari.Kellokumpu@oulu.fi).

localization motif within this peptide segment, and to assess the overall mechanism that is responsible for the localization of LH in the ER lumen. Based on both biochemical and protein structural approaches, we show that the ER residency of LH is mediated by a rather long peptide segment (fold), and that the mechanism itself seems to represent a novel, true ER retention system that does not involve any apparent recycling or retrieval of the enzyme from post-ER compartments.

EXPERIMENTAL

Cell culture and transfections

COS-7 cells were cultured in Dulbecco's modified Eagle's medium (Life Technologies) supplemented with glutamax, 10% fetal bovine serum and penicillin/streptomycin. Cells were plated one day before transfections with the FUGENE6 transfection reagent (Roche Molecular Biochemicals). Cells were typically used for the experiments 24–40 h after the transfection. Transfected cells were treated with nocodazole (Sigma-Aldrich) and cycloheximide (Sigma-Aldrich) using final concentrations of 20 and 25 µg/ml, respectively.

Plasmid preparations

Generation of plasmid CDLH40, which encodes human cathepsin D tagged with a c-Myc epitope and the last 40 amino acids from the C-terminal end of human LH isoform 1 (LH1), has been described elsewhere [21]. The KDEL-tagged version of cathepsin D, the CDMK plasmid, was a gift from Dr Hugh Pelham (Laboratory of Molecular Biology, Medical Research Council, Cambridge, U.K.). Deletion mutants of the CDLH40 plasmid (CDLH24 and CDLH32) were prepared first by amplifying the C-terminal coding sequence of LH1 by PCR using oligonucleotide pairs 5'-CCGGAATTCGCTACGCATTAC-CATGAGGGG-3' and 5'-CTAGTCTAGATTAGGGATCGA-CGAAGGAGA-3', and 5'-CCGGAATTCGTGGACCCTCA-TGCACCCTG-GA-3' and 5'-CTAGTCTAGATTAGGGATC-GACGAAGGAGA-3', respectively. Amplification products were then cloned into the plasmid CDLH40 using flanking *EcoRI* and *XbaI* restriction sites, thus replacing the sequence encoding the 40-amino acid segment with a sequence covering either 32 or 24 amino acids from the C-terminal end of LH1 (Figure 3, LH32 and LH24; see below). The CDLH40QG construct was generated from the CDLH40 plasmid using specific complementary oligonucleotides 5'-TCCATCCGAGC-CCCACAGGGGGGCTGGACCCTCATG-3' and 5'-CATG-AGGGTCCAGCCCCCTGTGGGGCTCGGATGGA-3' and the QuikChange site-directed mutagenesis kit (Stratagene; nucleotide changes are underlined). As a result, two amino acid changes, Arg⁶⁹³ → Gln and Lys⁶⁹⁴ → Gly, were introduced into the LH1 part of the construct (Figure 3, LH40QG; see below).

Point mutations into the human LH coding region were generated sequentially using the QuikChange site-directed mutagenesis kit and the LHMyC-plasmid expressing full-length LH1 tagged with the c-Myc epitope (see [21]) as a template. The following complementary oligonucleotide pairs were used for amplification reactions: 5'-CCCACCACCAGCGGCCA GCTACATCGCA-3' and 5'-TGCGATGTAGCTGGTGCCG CTGGTGGTGGG-3' (LHMyCMut1), 5'-ATGCACCCTGG-ACAACCTACGATTAC-3' and 5'-GTAATGCGTGAGTTG TCCAGGGTGCAT-3' (LHMyCMut2), 5'-ACGCATTACCA TTCGGGGCTCCCCACC-3' and 5'-GGTGGGGAGCCCC- GAATGGTAATGCGT-3' (LHMyCMut3) and finally 5'-GTC-

TCCTTCGTCTCTCCGAATTCGAG-3' and 5'-CTCGAAT-TCGGGAGAGACGAAGGAGAC-3' (LHMyCMut4; nucleotide changes are underlined). Altogether this resulted in four new constructs (LHMyCMut1–4; shown in Figure 3, see below) in which charged residues within the 32-amino acid segment were replaced by uncharged residues. Finally, the *BsmBI*–*EcoRI* DNA fragments (≈ 500 nt) from the C-terminal end of these clones were subcloned into the original LHMyC plasmid. Amino acid replacements in LHMyCMut2 and LHMyCMut3 plasmids were transferred also into the cathepsin D reporter protein vector using the same strategy and oligonucleotides as with the original CDLH40 construct (see [21]). As a result, plasmids CDLH40Mut2 and CDLH40Mut3 were generated.

All amplified sequences were confirmed by automatic sequencing using the DYEnamic™ ET terminator cycle sequencing premix kit (Amersham Biosciences) and an ABI Prism 377 DNA sequencer (Applied Biosystems).

Indirect immunofluorescence microscopy

Cells were fixed using 4% paraformaldehyde and permeabilized with saponin, as described earlier [22]. Expressed proteins were visualized with monoclonal anti-c-Myc (9E10) antibody (Santa Cruz Biotechnology) and Alexa® Fluor 488-conjugated goat anti-mouse antibodies (Molecular Probes). In double-staining experiments, expressed proteins were stained with polyclonal goat anti-cathepsin D antibodies (Santa Cruz Biotechnology) and Alexa® Fluor 488-conjugated donkey anti-goat secondary antibodies (Molecular Probes), and with a monoclonal anti-KDEL receptor antibody (Stressgen) and Alexa® Fluor 594-conjugated rabbit anti-mouse secondary antibodies (Molecular Probes). Stained specimens were examined using the Olympus BX-60 fluorescence microscope with appropriate filters.

Metabolic labelling and immunoprecipitation

Pulse-chase experiments and immunoprecipitations were carried out as described earlier [21]. Briefly, cells were first starved for 15 min in methionine/cysteine-free medium (Sigma-Aldrich), followed by a 2 h labelling with 200 µCi/ml Pro-mix™ cell-labelling reagent (Amersham Biosciences). Labelled cells were washed and then chased for 3 h in the presence of unlabelled methionine, cycloheximide and 10 mM NH₄Cl (to ensure maximal secretion of cathepsin D [21]). Immunoprecipitations from cell culture media and cell lysates were performed using agarose-conjugated 9E10 antibody (Santa Cruz Biotechnology). Bound proteins were resolved by SDS/PAGE and the gel was treated with EN³HANCE (Dupont) before exposure to X-ray film. Cell lysate/medium ratios were calculated after quantification of the corresponding signal intensities with the Imagequant 5.2 analysis program (Amersham Biosciences).

Fractionation of soluble and membrane proteins and immunoblotting

Membrane-association experiments were carried out as described earlier [21] using either TN buffer (50 mM Tris/HCl, pH 7.4/150 mM NaCl) or TKM buffer (50 mM Tris/HCl, pH 7.4/10 mM KCl/1 mM MgCl₂). Briefly, cells were disrupted by sonication and cell membranes and supernatants were separated by centrifugation. Membrane fraction was solubilized directly into SDS sample buffer, while soluble proteins were concentrated using trichloroacetic acid precipitation. Equivalent amounts of both fractions were subjected to SDS/PAGE and transferred to a nitrocellulose membrane. Immunoblotting was performed with

monoclonal antibodies against protein disulphide isomerase (PDI) [23] and the c-Myc epitope (9E10 anti-c-Myc antibody), followed by peroxidase-conjugated anti-mouse antibodies (P.A.R.I.S., Compiègne, France). Immunoreactive proteins were visualized using the ECL detection system (Amersham Biosciences) and quantified as above.

Fold-recognition and structure-prediction analyses of the C-terminal of LH1

Sequence homology searches were performed with the SWISS-model program (server version 3.5 at <http://www.expasy.ch/swissmod> [24]). The domain organization of LH1 was examined using both the Prodom [25] and Pfam [26] databases. Secondary structure elements and solvent accessibility of the LH1 C-terminal domain were predicted using programs PHD (version 5.94_317 at <http://www.embl-heidelberg.de/predictprotein/predictprotein.html> [27]) and PsiPred (version 2.2 at <http://bioinf.cs.ucl.ac.uk/psipred> [28,29]).

Structure alignment for the iron-dependent 2-oxoglutarate dioxygenase (2OG-FeII) domains of model proteins were carried out with the DALI structural search server (<http://www2.embl-ebl.ac.uk/dali/fssp> [30]). The initial sequence alignments were carried out with ClustalW (version 1.82 at <http://www.ebi.ac.uk/clustalw/> [31]) and refined using the JalView multiple alignment editor program (<http://circinus.ebi.ac.uk:6543/jalview/>). Threading programs 3D-pssm (<http://www.sbg.bio.ic.ac.uk/~3dpssm/> [32]) and GenTHREADER (<http://bioinf.cs.ucl.ac.uk/psipred> [28,29]) with default parameter values were used for fold recognition of the 2OG-FeII domain of LH1.

An *ab initio*-prediction of the 17-amino acid peptide structure was based upon a molecular dynamics simulation. A simulation of the free peptide was carried out with the Gromacs molecular dynamics program 2.1/3.0 [33,34] with periodic boundary conditions and particle mesh Ewald summation for long-range electrostatics. To ensure constant temperature and pressure, weak coupling to a temperature (300 K) and pressure (1.0 bar) bath was employed [35]. The LINCS algorithm [36] and SETTLE [37] were applied to reset bond lengths in the peptide and water, respectively. The initial co-ordinates for the peptide were generated from the buildpolymer module of InsightII (<http://www.accelrys.com/insight/Insight2.html>). Initially, the extended molecule was solvated into 9350 SPC (simple point charge) water molecules [38], and a 1 ns simulation with a time step of 2 fs was performed, after which the molecule was solvated into a truncated dodecahedron box with 4600 SPC water molecules. Dummy atoms [39] were created to decrease high-frequency movements of hydrogen atoms in aromatic rings. The peptide was simulated for 78 ns using a time step of 5 fs. Conformations of the peptide were extracted from the trajectory by clustering analysis using the gromos method [40] with a backbone atom root-mean-square-deviation difference of 0.1 nm, and analysing frames every tenth ps. Secondary structure content of the peptide was analysed using the DSSP method [41]. Figures were processed using the Rasmol program version 2.7.1.

RESULTS

LH1 is localized permanently in the ER

To differentiate whether ER localization of LH involves permanent retention of the enzyme in the ER or retrieval from post-ER compartments, we utilized nocodazole, a microtubule-depolymerizing drug. Drug treatment has been shown to alter the steady-state distribution of proteins that continuously recycle between the Golgi apparatus and the ER, but not of those that

Figure 1 Localization of LH does not involve recycling between the ER and the Golgi

Cells expressing either KDEL-tagged cathepsin D (**A, B**) or the c-Myc-tagged full-length LH (**C, D**) were treated with nocodazole for 4 h and fixed for immunofluorescence staining. Note that depolymerization of microtubules causes redistribution of CDMK into small dispersed dot-like structures close to or at the ER exit sites. Similar treatment of cells expressing LH did not cause such a redistribution, indicating that LH is a permanent resident of the ER.

are permanent residents of the ER [8,9]. Redistribution is thought to involve trapping of proteins in transport vesicles at or near the ER exit sites, whereas permanent residents remain in the ER [42,43].

Therefore, we treated cells expressing either the full-length LH1 or the KDEL-tagged cathepsin D reporter protein (CDMK) for 4 h with nocodazole, and stained expressed proteins with the anti-c-Myc antibody. We found that nocodazole induced the appearance of vesicular structures in cells expressing the KDEL-tagged cathepsin D (Figure 1B; compare with Figure 1A which shows an untreated control cell). In contrast, nocodazole-treatment of cells expressing the full-length LH (LHMyC; Figures 1C and 1D) did not cause such a redistribution, and the enzyme remained localized in the reticular ER (including the nuclear envelope).

Similarly to the α -subunit of prolyl-4-hydroxylase [44], localization of LH in the ER could also depend on its potential interactions with some KDEL-containing proteins, such as PDI. This possibility was assessed by following the observed redistribution of the KDEL receptor from the Golgi to the ER upon overexpression of the KDEL-containing proteins [1]. As expected, expression of cathepsin D tagged with the KDEL motif (CDMK; Figure 2A) induced a clear shift in KDEL-receptor localization from the Golgi to the ER (Figure 2B). In contrast, expression of cathepsin D tagged with the 40-amino acid C-terminal segment of human LH1 (CDLH40; Figure 2C) did not cause such a change in KDEL-receptor localization, even in highly expressing cells (Figure 2D). Collectively, these results suggest clearly that localization of LH in the ER is based on true retention, and not on its retrieval from post-ER compartments.

A 32-amino acid-long peptide segment is required for efficient retention

To map the critical localization signal within the C-terminal 40-amino acid peptide segment of LH, we created two new constructs

Figure 4 Retention efficiency of the different cathepsin D constructs

(A) Immunoprecipitation of the reporter proteins with agarose-conjugated anti-c-Myc antibody from either the total cell extracts (C) or the medium (M) after a 3 h chase period. Bound proteins were separated by SDS/PAGE (10% gels) and analysed by autoradiography. (B) The proportions of intracellular and secreted proteins in transfected cells. After quantification of the bands in (A) the retention efficiencies were calculated and expressed as percentages of the total protein. Similar retention efficiencies were obtained for CDLH32, CDLH40 or mutated CDLH40QG reporter proteins. However, CDLH24 secretion was higher than in other constructs. Retention efficiency of the full-length LH (LHMyC) is used as a control. (C) Membrane association of the different protein constructs. Cells expressing CDLH24, CDLH32 or CDLH40 protein were fractionated into membrane (M) and soluble (S) fractions. The distribution of the reporter proteins and PDI in the two fractions was analysed by Western blotting using specific antibodies. After normalization against PDI (a soluble protein in the ER), no marked changes were found in membrane association of the reporter proteins. C.D., cathepsin D.

residues (Arg⁷⁰³, Glu⁷⁰⁹, Arg⁷¹⁵, Arg⁷¹⁷ and Asp⁷²⁶ in Figure 3) within the C-terminal 32-amino acid peptide segment, it was necessary to assess whether any of these play a role in retention and membrane association of LH1 in the ER. Therefore, we created new constructs (LHMyCMut1–4 in Figure 3), in which each of the charged residues were substituted with a neutral amino acid. Mutated LH proteins were then tested for their retention efficiency and membrane association (Figure 5). The results revealed only minor differences in either the secretion rate (Figure 5A) or their association with the membranes (Figure 5B). Similarly to the intact enzyme itself (LHMyC), only about 5–15 % of the expressed mutant proteins were secreted during the chase period and all proteins were detected almost exclusively in the membrane fractions.

We next inserted the 40 amino acid segment from two of these mutated forms (LHMyCMut2 and LHMyCMut3) into the C-terminus of cathepsin D. Pulse-chase analysis of these mutated

Figure 5 Effect of charged residues on retention efficiency and membrane association

(A) Retention efficiency of different full-length LH constructs in which 2–5 charged amino acids were substituted with non-charged ones (Figure 3). Relative retention efficiency was studied by pulse-chase analysis and immunoprecipitation with the anti-c-Myc antibody (C, cells; M, medium). Note that removal of the positive and negative charges did not cause any marked increase in secretion of the LHMyC. (B) Cells were fractionated into membrane (M) and soluble (S) fractions and expressed proteins and PDI were detected in fractions using specific antibodies. All mutated forms of LH predominantly associated with the membranes, similarly to full-length LH.

Figure 6 Retention efficiency and membrane association of the mutated reporter protein constructs

(A) Pulse-chase analysis of CDLH40 proteins containing mutations introduced in LHMyCMut2 and LHMyCMut3 constructs. Quantification of the protein bands from the cell lysates (C) and the medium (M) showed that point mutations caused a 5% increase in secretion, when compared with the normal CDLH40. (B) Fractionation of the cells into membrane (M) and soluble (S) fractions were done as above. Quantification of the protein bands and normalization using PDI showed that amino acid changes caused a 15% decrease in membrane association. C.D., cathepsin D.

constructs revealed that secretion of both the CDLH40Mut2 and CDLH40Mut3 proteins was increased from 25 to 30 % in comparison with the non-mutated CDLH40 (Figure 6A). Similarly, the mutated proteins also showed lowered affinity towards membranes. Typically, 75–80 % of the original CDLH40 protein was detected in the membrane fraction, while only 60 % of the mutated forms were membrane-associated (Figure 6B). In each sample, we used soluble PDI as an internal control for our fractionation procedure. Thus charged residues within this

Figure 7 Three-dimensional structure of the 2OG-Fe(II) domain in model proteins

RasMol molecular visualization program was used for visualization. (A–D) Residues representing the last 40 C-terminal amino acids of LH1 are coloured red. The iron(II)-binding residues are represented as sticks. Strands are marked with yellow, turns with blue and helices with pink. (A) DAOCS (PDB code 1DCS 219–268). (B) IPNS (PDB code 1HB1 246–288). (C) P-3-H (PDB code 1E5R 147–177). (D) Canavalin (CAN; PDB code 1DGW 84–167). (E, F) The Figure shows the two representative model proteins, in which the different deletion segments (eight amino acids) are highlighted red (LH32), green (LH24) and blue (LH15) [21]. (G, H) Secondary structure

C-terminal 32-amino acid peptide segment indeed affect the retention efficiency and membrane association, indicating that electrostatic interactions have some role in the retention of LH1 in the ER.

Accessibility of the 40 amino acid peptide segment for binding

Domain analysis of the LH1 sequence revealed that it can be separated into four domains. The most C-terminal domain connects LH to the Pfam protein superfamily (2OG-Fe(II)s). Other enzymes that belong to this superfamily are functionally diverse and include prolyl-4-hydroxylases, isopenicillin synthases (IPNSs), alkylated DNA-repair protein, calavaminic acid synthase and deacetoxycephalosporin C synthase (DAOCS) [45]. Prolyl-3-hydroxylase (P-3-H) is not included in the Pfam superfamily listing, but it has been characterized as belonging to the group of 2-oxoglutarate-dependent oxygenases requiring iron(II) [46]. In addition to DAOCS, GenTHREADER identified canavalin (PDB code 1DGW), a seed storage protein from jack bean, as an additional protein with sequence similarity to the LH1 C-terminus. The DALI comparison of the three-dimensional structures revealed that canavalin also has a 2OG-Fe(II)-resembling fold as part of its structure.

The 2OG-Fe(II) domains of all these proteins (P-3-H, IPNS, calavaminic acid synthase, DAOCS and canavalin) have clear structural similarity despite their sequence diversity. A common motif detected in these proteins includes a ‘jelly roll’ β -strand core [47]. The corresponding 2OG-Fe(II) domain of LH1 after refined sequence alignment showed that it has several common features with the active-site centres of DAOCS, IPNS, P-3-H and canavalin. All model proteins contain an extended loop that partly co-aligns with the 40-amino acid segment of LH1 (Figures 7A–7D). Within the loop, there are distinct regions on either side of the iron-binding domains of these proteins that are surface-exposed and thus solvent-accessible. Two of these proteins (DAOCS and IPNS) also showed a short α -helical region in the middle of this surface-exposed peptide segment. Precise mapping of the most critical eight amino acid segments in terms of retention and membrane association (CDLH24 and CDLH32 constructs) showed that they cover roughly half of the loop and are surface-exposed on both sides of the domain (Figures 7E and 7F). The last 15 amino acids, instead, appear to be buried within the iron-binding domain itself.

A molecular-dynamics simulation was used to assess the most probable conformations of the LH1 C-terminal 17-amino acid peptide present in CDLH32 (amino acids 696–712 in LH1). The results (Figures 7G and 7H) showed two distinct conformations that were most abundant for this peptide during the 78 ns-long simulation. Both conformations adopted a loop-like structure. In one-quarter of the simulation time, the amino acids LTHY (amino acids 704–707 in the LH1 sequence, Figure 3) formed an α -helical structure. These simulated peptide structures are thus in good agreement with the known structures of the model proteins.

DISCUSSION

In this report, we have characterized the overall mechanism and the critical peptide motif that is sufficient for localization of LH1 in the ER lumen and its association with the ER membranes. Our results indicate that the ER residency of this collagen-processing

prediction of the 17-amino acid peptide (amino acids 696–712). Two most abundant structures are shown. The N-terminal end is marked with white, turns with blue and the helix with pink. The iron-binding histidine is shown as a stick.

enzyme does not rely on the known KDEL-receptor- and COP I-mediated retrieval/recycling mechanisms. Rather, LH appears to be localized in the ER via a mechanism that results in its permanent retention in the organelle. Our mutational and structure prediction analyses also indicate that the critical localization motif that was sufficient for efficient retention consists of a rather long peptide segment (17 amino acids) that appears to be part of the enzyme's iron-binding domain, and to be surface-exposed and thus readily accessible for binding.

Together with our previous findings [20,21], these results define a novel and rather unique localization mechanism in the ER. Homology searches with the 40-amino acid peptide segment identified only different LH isoforms (LH1–3) (identity over 90%), but no other protein, even with moderate sequence similarity. Therefore, all known LH isoforms appear to utilize the same mechanism for their localization in the ER. Compatible with this suggestion, we have found that the green-fluorescent-protein-tagged LH3 isoform indeed behaves as an ER-resident protein in nocodazole-treated cells (M. Suokas, R. Myllylä and S. Kellokumpu, unpublished work). Other more prevalent luminal chaperones and folding enzymes have been shown to utilize predominantly the KDEL-receptor- or COP I-mediated retrieval mechanisms.

Based on our mutational analyses, the overall length and structure of the C-terminal peptide segment appeared to be most crucial for efficient retention. Thus, substitution of all charged residues with neutral (and polar) amino acids within the 40-amino acid segment only moderately reduced the retention efficiency and membrane association, when the mutated segment was transferred into the C-terminal end of cathepsin D reporter protein. In good agreement with this, our fold-recognition and structure-prediction analyses indicate that the peptide segment is part of the iron-binding domain of the structurally homologous model proteins (Figure 7), and forms an extended loop within this domain. Most importantly, the critical peptide region appears to be exposed on either side of this iron-binding domain. Detailed mapping of the co-aligned residues of the deletion mutants showed that the most critical part for ER localization reside around the α -helical region found in the middle of the loop in two of the model proteins (Figures 7A and 7B), and also in the simulated peptide structure (Figures 7G and 7H). On the other hand, residues that were dispensable in terms of retention (eight residues from the beginning) of the 40-amino acid peptide segment are not represented in this structural domain, being present only in the LH sequence. These data are thus in good agreement with our experimental data, showing that the amino acids in the middle of this 40-amino acid segment are most crucial for ER localization. Validity of these observations was also estimated by using a molecular-dynamics simulation, showing a similar structure for the free peptide (amino acids 696–712, Figure 3). It is also interesting to note that this 17-amino acid segment consists of 10 non-polar amino acids (of which six are hydrophobic, Figure 3), suggesting that hydrophobic interactions may be involved in the association of the enzyme with the ER membranes.

According to our structure-prediction analyses, this critical segment is part of the enzyme's iron-binding domain [48] and may therefore also be involved in the binding of the enzyme with its substrates, i.e. the collagens. However, our preliminary data with collagen-affinity matrix indicate that the intact enzyme binds with much higher affinity to citrate soluble collagen than cathepsin D tagged only with the C-terminal 40 amino acids of LH (M. Suokas, R. Myllylä and S. Kellokumpu, unpublished work). Therefore, differences either in substrate binding or oligomerization as well as conformational changes may explain

not only the slightly lower retention efficiency of the cathepsin D constructs than that of the intact enzyme, but also the overall insensitivity of the full-length enzyme to point mutations.

Being a peripheral membrane protein itself [20], it is likely that the ER localization of LH1 relies on its binding to some specific, as yet unidentified, ER membrane constituent that is also localized in the ER via a retrieval-independent mechanism. The identity of this membrane counterpart remains to be explored. Functionally, the association of LH with the ER membranes is relevant, as hydroxylation of lysines has been reported to be a co- and post-translational event occurring before triple-helix formation [11]. It is also notable that hydroxylation of prolyl residues by the luminal prolyl-4-hydroxylase enzyme is a critical quality-control step during collagen biosynthesis [49]. Thus it would be logical that other collagen-processing steps, such as those mediated by LH, are carried out earlier, perhaps even co-translationally [50].

This work was supported by grants from the Academy of Finland, Helsinki, Finland. We gratefully thank Mrs Annika Kauppila and Ms Piia Mäkelä for their expert technical assistance. The Center of Scientific Computing, Espoo, Finland, is also greatly acknowledged.

REFERENCES

- Lewis, M. J. and Pelham, H. R. (1992) Ligand-induced redistribution of a human KDEL receptor from the Golgi complex to the endoplasmic reticulum. *Cell* **68**, 353–364.
- Teasdale, R. D. and Jackson, M. R. (1996) Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the Golgi apparatus. *Annu. Rev. Cell Dev. Biol.* **12**, 27–54.
- Jackson, M. R., Nilsson, T. and Peterson, P. A. (1993) Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *J. Cell Biol.* **121**, 317–333.
- Itin, C., Schindler, R. and Hauri, H. P. (1995) Targeting of protein ERGIC-53 to the ER/ERGIC/cis-Golgi recycling pathway. *J. Cell Biol.* **131**, 57–67.
- Andersson, H., Kappeler, F. and Hauri, H. P. (1999) Protein targeting to endoplasmic reticulum by dilysine signals involves direct retention in addition to retrieval. *J. Biol. Chem.* **274**, 15080–15084.
- Letourneur, F., Gaynor, E. C., Hennecke, S., Demolliere, C., Duden, R., Emr, S. D., Riezman, H. and Cosson, P. (1994) Coatamer is essential for retrieval of dilysine-tagged proteins to the endoplasmic reticulum. *Cell* **79**, 1199–1207.
- Szczesna-Skorupa, E., Chen, C. D. and Kemper, B. (2000) Cytochromes P450 2C1/2 and P450 2E1 are retained in the endoplasmic reticulum membrane by different mechanisms. *Arch. Biochem. Biophys.* **374**, 128–136.
- Szczesna-Skorupa, E. and Kemper, B. (2000) Endoplasmic reticulum determinants in the transmembrane and linker domains of cytochrome P450 2C1. *J. Biol. Chem.* **275**, 19409–19415.
- Fu, J. and Kreibich, G. (2000) Retention of subunits of the oligosaccharyltransferase complex in the endoplasmic reticulum. *J. Biol. Chem.* **275**, 3984–3990.
- Fu, J., Pirozzi, G., Sanjay, A., Levy, R., Chen, Y., De Lemos-Chiarandini, C., Sabatini, D. and Kreibich, G. (2000) Localization of ribophorin II to the endoplasmic reticulum involves both its transmembrane and cytoplasmic domains. *Eur. J. Cell Biol.* **79**, 219–228.
- Prockop, D. J. and Kivirikko, K. I. (1995) Collagens: molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* **64**, 403–434.
- Valtavaara, M., Papponen, H., Pirttilä, A.-M., Hiltunen, K., Helander, H. and Myllylä, R. (1997) Cloning and characterization of a novel human lysyl hydroxylase isoform highly expressed in pancreas and muscle. *J. Biol. Chem.* **272**, 6831–6834.
- Valtavaara, M., Szpirer, C., Szpirer, J. and Myllylä, R. (1998) Primary structure, tissue distribution, and chromosomal localization of a novel isoform of lysyl hydroxylase (lysyl hydroxylase 3). *J. Biol. Chem.* **273**, 12881–12886.
- Passoja, K., Rautavuoma, K., Ala-Kokko, L., Kosonen, T. and Kivirikko, K. I. (1998) Cloning and characterization of a third human lysyl hydroxylase isoform. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10482–10486.
- Ruotsalainen, H., Sipilä, L., Kerkelä, E., Pospiech, H. and Myllylä, R. (1999) Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3, their phylogenetic analysis and tissue-specific expression in the mouse. *Matrix Biol.* **18**, 325–329.
- Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M. and Myllylä, R. (2000) Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. *J. Biol. Chem.* **275**, 36158–36163.

- 17 Wang, C., Risteli, M., Heikkinen, J., Hussa, A.-K., Uitto, L. and Myllylä, R. (2002) Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). *J. Biol. Chem.* **277**, 18568–18573
- 18 Rautavuoma, K., Takaluoma, K., Passoja, K., Pirsanen, A., Kvist, A.-P., Kivirikko, K. I. and Myllyharju, J. (2002) Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1–3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. *J. Biol. Chem.* **277**, 23084–23091
- 19 Wang, C., Luosujärvi, H., Heikkinen, J., Risteli, M., Uitto, L. and Myllylä, R. (2002) The third activity for lysyl hydroxylase 3: galactosylation of hydroxyllysyl residues in collagens *in vitro*. *Matrix Biol.* **21**, 559–566
- 20 Kellokumpu, S., Sormunen, R., Heikkinen, J. and Myllylä, R. (1994) Lysyl hydroxylase, a collagen processing enzyme, exemplifies a novel class of luminally-oriented peripheral membrane proteins in the endoplasmic reticulum. *J. Biol. Chem.* **269**, 30524–30529
- 21 Suokas, M., Myllylä, R. and Kellokumpu, S. (2000) A single C-terminal peptide segment mediates both membrane association and localization of lysyl hydroxylase in the endoplasmic reticulum. *J. Biol. Chem.* **275**, 17863–17868
- 22 Kellokumpu, S., Neff, L., Jämsä-Kellokumpu, S., Kopito, R. and Baron, R. (1988) A 115-kD polypeptide immunologically related to erythrocyte band 3 is present in Golgi membranes. *Science* **242**, 1308–1311
- 23 Höyhty, M., Myllylä, R., Piiva, J. and Kivirikko, K. I. (1984) Monoclonal antibodies to human prolyl 4-hydroxylase. *Eur. J. Biochem.* **141**, 477–482
- 24 Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723
- 25 Corpet, F., Gouzy, J. and Kahn, D. (1998) The ProDom database of protein domain families. *Nucleic Acids Res.* **26**, 323–326
- 26 Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. and Sonnhammer, E. L. L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280
- 27 Rost, B. and Sander, C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**, 55–72
- 28 Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202
- 29 Jones, D. T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287**, 797–815
- 30 Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138
- 31 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680
- 32 Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299**, 499–520
- 33 Berendsen, H. J. C., van der Spoel, D. and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.* **91**, 43–56
- 34 Lindahl, E., Hess, B. and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **7**, 306–317
- 35 Berendsen, H. J. C., Postman, J. P. M., DiNola, A., van Gunsteren, W. F. and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **181**, 3684–3690
- 36 Hess, B., Bekke, H., Berendsen, H. J. C. and Fraaije, J. G. E. M. (1997) LINC: a linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472
- 37 Miyamoto, S. and Kollman, P. A. (1992) SETTLE: an analytical version of the shake and rattle algorithms for rigid water models. *J. Comp. Chem.* **13**, 952–962
- 38 Berendsen, H. J. C., Postman, J. P. M., van Gunsteren, W. F. and Hermans, J. (1981) Interaction models for water in relation to proteins hydration. In *Intermolecular Forces* (Pullman, B. D., ed.), pp. 331–342. Reidel Publishing Company, Dordrecht
- 39 Feenstra, K. A., Hess, B. and Berendsen, H. J. C. (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comp. Chem.* **20**, 786–798
- 40 Daura, X., Gademann, K., Jaun, B., Seebach, D., Gunsteren, W. F. and Mark, A. E. (1999) Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed.* **38**, 236–240
- 41 Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637
- 42 Hsu, V. W., Yuan, L. C., Nuchtern, J. G., Lippincott-Schwartz, J., Hammerling, G. J. and Klausner, R. D. (1991) A recycling pathway between the endoplasmic reticulum and the Golgi apparatus for retention of unassembled MHC class I molecules. *Nature (London)* **352**, 441–444
- 43 Cole, N. B., Sciaky, N., Marotta, A., Song, J. and Lippincott-Schwartz, J. (1996) Golgi dispersal during microtubule disruption: regeneration of Golgi stacks at peripheral endoplasmic reticulum exit sites. *Mol. Biol. Cell* **7**, 631–650
- 44 Vuori, K., Pihlajaniemi, T., Myllylä, R. and Kivirikko, K. I. (1992) Site-directed mutagenesis of human protein disulphide isomerase: effect on the assembly, activity and endoplasmic reticulum retention of human prolyl 4-hydroxylase in *Spodoptera frugiperda* insect cells. *EMBO J.* **11**, 4213–4217
- 45 Aravind, L. and Koonin, E. V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**, Research 0007.1–0007.8
- 46 Clifton, I., Hsueh, L., Baldwin, J., Harlos, K. and Schofield, C. (2001) Structure of proline 3-hydroxylase. Evolution of the family of 2-oxoglutarate dependent oxygenases. *Eur. J. Biochem.* **268**, 6625–6636
- 47 Schofield, C. J. and Zhang, Z. (2001) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* **9**, 722–731
- 48 Pirsanen, A., Kaimio, A.-M., Myllylä, R. and Kivirikko, K. I. (1996) Site-directed mutagenesis of human lysyl hydroxylase expressed in insect cells. Identification of histidine residues and an aspartic acid residue critical for catalytic activity. *J. Biol. Chem.* **271**, 9398–9402
- 49 Walmsley, A. R., Batten, M. R., Lad, U. and Bulleid, N. J. (1999) Intracellular retention of procollagen within the endoplasmic reticulum is mediated by prolyl 4-hydroxylase. *J. Biol. Chem.* **274**, 14884–14892
- 50 Pihlajaniemi, T., Myllylä, R., Alitalo, K., Vaheri, A. and Kivirikko, K. I. (1981) Posttranslational modifications in the biosynthesis of type IV collagen by a human tumor cell line. *Biochemistry* **20**, 7409–7415

Received 2 October 2002/19 November 2002; accepted 27 November 2002

Published as BJ Immediate Publication 27 November 2002, DOI 10.1042/BJ20021533