Characterization of Collagenous Peptides Bound to Lysyl Hydroxylase Isoforms*

Received for publication, May 20, 2004, and in revised form, June 17, 2004
Published, JBC Papers in Press, June 18, 2004, DOI 10.1074/jbc.M405638200

Maija Risteli‡, Olli Niemitalo‡, Hilika Lankinen§, André H. Juffer‡, and Raili Myllylä‡

From the ‡Department of Biochemistry, Biocenter Oulu, University of Oulu, P. O. Box 3000, Finn-90014, University of Oulu, Finland and §Peptide and Protein Laboratory, Department of Virology and Helsinki University Central Hospital Laboratory Diagnostics, Haartman Institute, P. O. Box 21 (Haartmaninkatu 3), Fin-00014, University of Helsinki, Finland

Lysyl hydroxylase (LH, EC 1.14.11.4) is the enzyme catalyzing the formation of hydroxylysyl residues in collagens and other proteins with collagenous domains. Although lower species, such as Caenorhabditis elegans, have only one LH orthologue, LH activity in higher species, such as human, rat, and mouse, is present in three molecules, LH1, LH2, and LH3, encoded by three different genes. In addition, LH2 is present in two alternatively spliced forms (LH2a, LH2b). To understand the functions of the four molecular forms of LH in vertebrates, we analyzed differences in the binding and hydroxylation of various collagenous peptides by the LH isoforms. Nine-amino acid-long synthetic peptides on Pepspot were used for the binding analysis and an activity assay to measure hydroxylation. Our data with 727 collagenous peptides indicated that a positive charge on the peptide and specific amino acid residues in close proximity to the lysyl residues in the collagenous sequences are the key factors promoting peptide binding to the LH isoforms. The data suggest that the LH binding site is not a deep hydrophobic pocket but is open and hydrophilic where acidic amino acids play an important role in the binding. The data do not indicate strict sequence specificity for the LH isoforms, but the data indicated that there was a clear preference for some sequences to be bound and hydroxylated by a certain isoform.

Hydroxylysyl residues are post-translational modifications of proteins found mainly in collagens, the components of connective tissues (1–5). Hydroxylysyl residues are found in the Y positions of repeating X-Y-Gly sequences of collagenous polypeptide chains. In addition, some other sequences in the non-helical amino-terminal region of fibrillar collagens contain hydroxylysyl residues but lack the repeating collagenous sequence. The amount of hydroxylysyl residues among the different collagen types varies from 6 residues per 1000 amino acids in type III collagen to about 70 residues per 1000 amino acids for type VI collagen. Hydroxylysyl residues are also present in some noncollagenous proteins, such as acetylcholinesterase, complement C1q, macrophage scavenger receptor, etc., all of them, however, having a collagenous domain in their structure (1–5).

Hydroxylysyl residues serve as attachment sites for the carbohydrates galactose and glucosylgalactose (6). The function of these carbohydrate moieties, which are unique to collagenous structures, is unknown, however. Hydroxylysyl residues also have an important function in collagen cross-link formation in vivo, which is to build fibril-forming collagen molecules into stable supramolecular structures (2, 7, 8). Two alternative pathways exist for cross-link formation, one based on lysine residues and the other on hydroxylysyl residues. The lysyl-associated pathway predominates in skin, cornea, sclera, and other soft tissues, whereas more stable cross-links are formed by the hydroxylysyl-derived pathway, which predominates in weight-bearing and mineralized tissues such as bone, cartilage, ligament, and tendons. Embryonic skin and major connective tissues of the body also contain hydroxylysyl linked cross-links (2, 7, 8).

Lysyl hydroxylase (LH, EC 1.14.11.4) is the enzyme that catalyzes the formation of hydroxylysyl residues in collagens and other proteins with collagen-like amino acid sequences (1–5). The reaction occurs as a post-translational event before triple helix formation in the endoplasmic reticulum. The reaction requires ferrous ion, 2-oxoglutarate, molecular oxygen, and ascorbate. The amino acid sequence required for the LH reaction is an X-Lys-Gly sequence, a triplet repeated many times in collagenous proteins. There is evidence that in some cases glycine in the above triplet can be replaced by some other amino acids (1, 2, 9). The interaction of the amino acid sequence with LH is influenced by peptide chain length, the effect of which is seen mainly in the $K_{m}$, which decreases with increasing chain length, whereas the $V_{\text{max}}$ of the reaction is unaffected (1, 10, 11). Furthermore, the triple helical or other type of conformation of the peptide prevents lysine hydroxylation (1, 12).

Lysyl hydroxylase activity in the human, mouse, and rat tissues is present in three different molecules, LH1, LH2, and LH3, originating from three different genes (1, 13–21, 44). Our phylogenetic data suggest that LH1 and LH2 are more closely related and produced by a more recent duplication event than the less closely related LH3 (18). LH2 is present in two alternatively spliced forms (LH2a and LH2b); LH2a excludes exon 13A from the LH2 transcript (22, 23). LH3 differs from LH1 and LH2 in that it is multifunctional and able to catalyze in addition to lysyl hydroxylation, sugar transfer reactions, the consecutive steps in the formation of glucosylgalactosylhydroxylysyl residues (24–25). Only one isoform for LH is present in lower species such as Caenorhabditis elegans (26). This ancestral C. elegans LH is also able to glycosylate hydroxylysyl residues and, thus, is functionally similar to LH3 (25, 27).

* This work was supported by grants from the Research Council within the Academy of Finland and by grants from the Sigrid Juselius Foundation and Biocenter Oulu. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

To whom correspondence should be addressed: Dept. of Biochemistry, P. O. Box 3000, 90014 University of Oulu, Finland. Fax: 358-8-5531141; E-mail: raili.myllyla@oulu.fi.

† The abbreviation used is: LH, lysyl hydroxylase.
To better understand the evolutionary development of the lysyl hydroxylase isoforms and to obtain information about the functions of the three isoforms in vivo, we compared the interactions of LH1, LH2, and LH3 with different collagenous peptides. Furthermore, we analyzed the change in activity and peptide binding when LH2 was changed via alternative splicing of one exon, from LH2a (a short form) to LH2b (a long form). We have studied the characteristics of the peptides bound to the LH isoforms and the ability of each isoform to hydroxylate the peptide. Our data revealed that there was no limited requirement of the collagenous peptide substrate for the LH1, LH2, or LH3 isoforms, but there was a clear preference for some sequences to be bound and hydroxylated by a certain isoform.

EXPERIMENTAL PROCEDURES

Production and Purification of LH Isoforms—All recombinant LH isoform expressions were carried out by baculovirus transfer vector pFastBacI in the BAC-TO-BAC™ Expression system (Invitrogen). The vector was modified to contain the human LH1 signal peptide, His tag, and a BamHI site for insertion of the desired cDNA (12, 14). The construct of human LH1 covered the nucleotides from 235 to 2405 (13). The construct of human LH2a and LH2b covered the nucleotides from 76 to 2267 (14, 22), the LH2b containing the alternatively spliced exon sequence between nucleotides 1500 and 1501 (22, 23). The construct of human LH3 covered the nucleotides from 295 to 1455 (15). The vectors were transfected in Sf9 cells. For Pepspots analysis the His-tagged recombinant proteins were purified on a nickel nitrilotriacetic acid-agarose (Qiagen) column as described elsewhere (14, 24, 25). For activity assays the crude supernatant of the Sf9 cells was used as the enzyme source.

Activity Measurements—LH activity was assayed by the hydroxylation-coupled decarboxylation of 2-oxo [1-14C]glutarate (28) using a synthetic peptide as a substrate. The specific activity of 2-oxoglutarate was 74 × 10^4 dpm/μmol. The synthetic peptides used as substrates in the LH reaction were synthesized at the Department of Biochemistry, University of Oulu, using an ABI 433A synthesizer and an Fmoc peptidase-scanning assay. For interaction studies, the membrane was blocked for 2 h at room temperature with 5% dried milk in 20 mM Tris-HCl, pH 7.6, 170 mM NaCl, and 0.1% Tween 20 and subsequently incubated with a baculovirus-expressed and nickel column-purified LH isoform at a concentration 5 μg/ml in the presence of 50 mM Tris-HCl buffer, pH 7.8, 50 mM FeSO4, and 1 mM ascorbate for 1 h at 37 °C. Unbound protein was removed by washing two times with Tris-buffered saline. The bound LH isoform was electroblotted onto a polyvinylidene difluoride membrane (Millipore) using a SemiPhor transfer unit ( Hoefer) as recommended by Jerini Peptide Technologies GmbH. and visualized on x-ray film (Eastman Kodak Co.) according to the protocol of ECL™ Western blotting (Amersham Biosciences) using the monoclonal anti-polyhistidine clone His-1 (Sigma) and horseradish peroxidase-goat anti-mouse IgG conjugate (Zymed Laboratories Inc.).

Bayesian Analysis—A Bayesian approach (31) was taken to analyze the Pepspot data. We wanted to access the qualities of the peptides that would promote or hinder binding. The binding probability of each set of peptides carrying a certain quality was compared against the binding probability of peptides that were not in the set. Because of the limited amount of data, the exact values of binding probabilities were not calculated, but instead, the binding probabilities were treated as random variables. Their probability distributions were calculated based on our prior estimates and the data. The binding probability of a set of peptides was defined as the probability that a Pepspot binding test gives a positive result for a peptide sampled randomly from the set. The probability of sampling a particular peptide sequence from a set of peptide sequences equals the probability that evolution would produce such a sequence in a uniformly random position in the collagen sequence given that the collagen subsequence is known to be in the set. The selection of the sequences for the Pepspots set was modeled as this kind of sampling.

SAMMON MAP OF SEQUENCES—To find more qualities of the peptides that might be related to binding, a visualization tool called Sammon mapping (32, 33) was used to map the Pepspot peptides onto a two-dimensional map in such a way that the distances on the Sammon map approximated calculated pair-wise sequence dissimilarities. Similar peptides were therefore positioned close to each other, and dissimilar peptides were placed farther apart. The sequence dissimilarity (WAC dissimilarity) was calculated as a sum over the WAC amino acid dissimilarities of the corresponding positions in the two sequences. The WAC similarity matrix (34), which is based on physicochemical properties of the microenvironments created by amino acids, was chosen over matrices based on phylogeny because it was thought that the peptide sequences may have evolved at least partially independently. The WAC dissimilarity used here was derived from the WAC similarity as 14 minus WAC similarity, where 14 is the value of the WAC self-similarity. The G3PCX (35) evolutionary algorithm was used to optimize the Sammon map.

CD Spectra—The secondary structure of the peptides that were employed in the activity measurements was studied with CD spectroscopy. The CD spectra were recorded with a Jasco J-715 spectropolarimeter equipped with a microprocessor for spectral accumulation and data manipulation. All measurements were made in 1-mm path length quartz cells. The peptide concentration was 0.5 μg/ml, and the measurements were performed at room temperature (+22 °C). The ellipticity was normalized to the mean residue molar ellipticity. Secondary structure percentages of peptides were calculated from the circular dichroism spectra using the CONTIN/LML method (36) with the SP3TA reference set of 37 soluble proteins (37), as provided by the CDPro software package (38).

Theoretical Peptide Structures—It is suspected that peptides binding to LH have a similar fold (11). This was investigated by solving theoretical structures of the Pepspot peptides in water and by comparing the structural data to the Pepspot binding data. The structures were solved using the biased probability Monte Carlo global optimization procedure with distance dependent electrostatics, as provided by the ICM software package (39). The carboxyl end negative charge was protonated to account for the effects of the peptides being bound at their carboxyl end to the membrane via a polyethylene glycol spacer molecule.

Theoretical Peptide Structures and Binding Data—Various approaches were used to find correlations between the peptide structures and the Pepspot binding data. As a structural dissimilarity measure, the mean square deviation of the coordinates of the backbone and Lys-5 heavy atoms between all Pepspot peptide pairs were calculated. A Sammon map of the peptides was created from the calculated root mean square deviation data and colored according to the Pepspot binding. The dependence of binding probability on peptide solvation free energy was analyzed statistically. A large negative solvation free energy suggests that the peptide prefer an aqueous environment over a hydrophobic one. The solvation free energy was calculated using ICM software. Previous research (11) suggested a bent structure with lysine in the middle as a criterion for hydroxylation. To assess this, a secondary structure analysis (40) was carried out on the three-dimensional structures.

RESULTS

We have synthesized nine-amino acid collagenous peptides having lysyl residues at the Y position of repeating X-Y-Gly sequences. The sequences were obtained from polypeptide chains of type I-XIX collagens (29) and comprised altogether 727 different sequences. The peptides represent potential in vivo substrates for the lysyl hydroxylase isoforms. As shown in Table I, most of the peptide sequences were obtained from collagen types IV, V, VI, VII, IX, and XI. The synthetic peptides were immobilized onto a membrane, forming Pepspots, and the binding of the LH isoforms to the Pepspots was recognized via a His-tag signal at the amino terminus of the bound enzyme. Subsequently, we performed a statistical analysis of our Pepspots data and studied the characteristics of peptides bound to
Pepspots and the peptides that failed to bind. We analyzed the effects of the net charge, the physicochemical properties, and the position of individual amino acids in the sequence on the binding of the peptide to the enzyme. Some of the peptides were also chosen for in vitro activity assays.

**Binding of Collagenous Peptides to LH1**—LH1 bound to 196 peptides derived from various collagen types, the majority from types XI, XII, XV, and XVI (Table I). As shown in Fig. 1, the net charge of the peptide strongly influenced the binding. A net charge of 0 or less was associated with a low binding probability. A net charge of the peptide was again an important determinant for binding to LH3 (Fig. 1). Visualization of the peptide binding is shown in Fig. 2d. Our data indicated that arginine and lysine in positions 1, alanine and threonine in position 4, lysine, serine, tyrosine, and histidine in position 7, and arginine and lysine in position 8 promote peptide binding, whereas glutamate in position 1, aspartate in position 2, glutamate and aspartate in position 7, and glutamate, glutamine, and proline in position 8 inhibited the binding (Table II).

**Some of the Peptides Bound to One Isoform**—There was some overlapping in the binding of peptides to the LH isoforms; 46 of the peptides bound to all isoforms (LH1, LH2a, LH2b, LH3), 6 bound to LH1 and LH3, 22 bound to LH2a and LH3, 33 bound to LH2b and LH3, 6 bound to LH2a and LH2b, 19 bound to LH1 and LH2a, and 30 bound to LH1 and LH2b. There were some peptides, however, that bound to only one LH isoform. We found 21 sequences, which only bound to LH1, 45 bound only to LH2a, 32 bound only to LH2b only, and 40 bound only to LH3.

**Amino Acids Present in Collagenous Sequences in Vivo**—We constructed a nine-amino acid sequence figure from amino acids present in the collagenous sequences (Fig. 3) according to the following constraints. Gly was constrained in position 6 to the maximal score and, the size of the other amino acids represents their abundance in the other positions. The sequences were collected by computer from collagen types I to XIX. All collagen α-chain sequences were examined for the occurrence of short sequence motifs such as XYG (Fig. 3a), XRG (Fig. 3a), and XKG (Fig. 3a). For each motif found, a nine-residue subsequence was determined around that position. Finally, the relative frequencies of the amino acids in each position were calculated. As seen in Fig. 3a, proline is most abundant amino acid in all triplets in the X and Y positions. The other common amino acids in the X positions are leucine, glutamate, and alanine, whereas in the Y position the most common amino acids are lysine, arginine, alanine, and glutamine. The most abundant amino acids preceding the lysyl residue in the Y position are proline, glutamate, leucine, alanine, aspartate, and glutamine (Fig. 3a).

**Analysis of Peptides Bound to Different LH Isoforms Based on Amino Acid Sequences**—As seen in Fig. 2, the WAC distance-based Sammon maps showed no absolute sequence specificity for any of the four LH isoforms. However, a detailed analysis of the Sammon maps (Fig. 2, areas marked by 1–5) revealed some differences between the isoforms. As seen in Fig. 4, in certain areas of the Sammon map the binding probability was not identical to all the isoforms. An analysis of sequences in these areas (Fig. 3b, 1–5) showed that area 1, where LH1 has

| Collagen type | Number of sequences | Percentage of the peptides bound to LH isoforms |
|--------------|---------------------|-----------------------------------------------|
|              | LH1 | LH2a | LH2b | LH3 |
| I            | 34  | 21   | 9    | 29  |
| II           | 19  | 36   | 21   | 33  |
| III          | 35  | 18   | 11   | 11  |
| IV           | 263 | 25   | 27   | 33  |
| V            | 57  | 23   | 25   | 35  |
| VI           | 42  | 33   | 33   | 45  |
| VII          | 46  | 37   | 37   | 22  |
| VIII         | 20  | 20   | 35   | 35  |
| IX           | 19  | 22   | 23   | 14  |
| X            | 49  | 39   | 27   | 49  |
| XI           | 13  | 38   | 23   | 54  |
| XII          | 19  | 32   | 21   | 26  |
| XIII         | 8   | 36   | 25   | 50  |
| XIV          | 15  | 47   | 27   | 40  |
| XV           | 35  | 26   | 40   | 26  |
| XVI          | 13  | 38   | 0    | 46  |
| XVII         | 21  | 33   | 33   | 24  |
| XVIII        | 20  | 0    | 20   | 40  |
| Synthetic    | 2   | 0    | 50   | 0   |

*35 of the sequences are present in two collagen types.
a low binding probability, has an increased frequency of proline in position 2, leucine in position 4, and glutamine in position 7. Area 2, in which LH2a has a low binding probability, has glutamate at position 7 in most peptides and an increased frequency of proline in position 4. Area 3 gives a moderate binding probability for all isoforms, somewhat higher for LH1 and LH2a, and has a positively charged amino acid in position 2, mostly glutamate in position 4, and at a high probability a negatively charged amino acid at position 7. Area 4, in which all isoforms except LH2a have a high binding probability, also has mostly glutamate in position 4, but position 7 is rarely occupied by an acidic residue; proline has a somewhat increased frequency in position 8. Area 5 has a somewhat high binding probability for LH3, and its sequences are marked by an increase in alanine in positions 2, 7, and 8 and proline in position 4, replacing the charged residues normally occupying these positions.

**Theoretical Peptide Structures and Binding Data**—It was found that the free energy of solvation of a peptide into water correlates well with binding (Fig. 5). With all isoforms, a low solvation free energy, favorable for an aqueous environment, is
associated with a high binding probability, and a higher solvation free energy is associated with a low binding probability. Because it was suspected that the solvation free energy might be directly explained by the peptide net charge, a scatter plot of the two properties (Fig. 6) was created. The solvation free energy was found to be strongly dependent on peptide net charge (Fig. 6), with both positively and negatively charged peptides preferring an aqueous environment. No increase in binding was found (not shown) for peptides with a hydrogen-bonded turn over Lys-5, as identified in the secondary structure analysis. The Sammon map based on structural root mean square deviation dissimilarities (not shown) showed no significant aggregation of binding or non-binding peptides.

Activity Assay to Test Bound Peptides as Substrates for LH Isoforms—An activity assay was used to determine whether the peptides bound to the LH isoforms can be hydroxylated by the enzymes, i.e. serve as substrates in the reaction. The conditions for the activity assay differ from those used on the Pepspots analysis, however. One of the cosubstrates, 2-oxoglutarate, was not present in the solution in the Pepspot analysis so as to prevent hydroxylation and subsequent release of the enzyme. Furthermore, peptides on Pepspots were immobilized while they were in solution in the activity assay. For these reasons, the binding of individual peptides in the activity assay (K_m values in Table III) was not correlated with how they bound on the Pepspots. This was analyzed by comparing the corresponding K_m values between binding and non-binding peptides in Pepspots data. The result is understandable because, for individual peptides, even a small e.g. structural variation can cause a different result in either assay. With enough data a statistical analysis can overcome this variation, but for practical reasons, the number of activity assays could not be increased to such a high level. We were able to determine, however, in our activity assay whether the peptides were bound to the active site of the enzyme and served as substrates for the LH isoforms. We selected 28 peptides from Pepspots representing all types of sequences for in vitro measurements, and these peptides were tested with all four LH isoforms; altogether there were 112 measurements. The data revealed that in 57 measurements the result of the activity assay agreed with the Pepspots data. Furthermore, in 50 measurements, the isoform was able to hydroxylate the peptide, although the Pepspot analysis was negative. As seen from these data, the activity assay was more sensitive than the Pepspots binding analysis. This was as expected and indicated that our screen by Pepspots had picked about half of the peptides compared with the data obtained by the activity assay. There were only a few peptides, 5 of 112 analyses, that showed binding in the Pepspots analysis but did not serve as substrates in the activity assay. These were probably peptides that bound to the enzyme at a site other than in the active site, or they may represent peptide inhibitors.

The K_m and V_max values of the different peptides for the LH isoforms are listed in Table III. The V_max data indicated differences in the capacity of the various isoforms to hydroxylate different peptides. Interestingly, none of the negatively charged peptides tested were hydroxylated by LH1, whereas most of them were substrates for the other isoforms. The peptides with a positive charge all served as substrates in the reaction, which agrees with the Pepspot binding data. The presence of one extra exon in LH2 (LH2b in comparison with LH2a) caused a change in the behavior of LH2, as also seen in our binding assay. LH2a had a different ability to hydroxylate collagenous peptides compared with LH2b. Furthermore, changing one of the amino acids or the order of two amino acids in the peptide (compare L26 with L32 and L22 with L23) affected the ability of the peptide to be hydroxylated by the LH isoforms. Furthermore, increasing the amount of NaCl in the activity assay caused a remarkable reduction of the activity in most cases, in agreement with our finding that the net charge of the peptide is a key determinant for binding (Table IV).

Circular Dichroism Spectra—The secondary structure of peptides used in the activity assays were obtained from CD spectra. Three typical CD spectra are shown in Fig. 7. The secondary structure percentages (Table V) as interpreted from the CD spectra showed that all the analyzed peptides had a largely unordered structure, 28–50%, with the β-sheet being the most prevalent, 24–48%, of the ordered structural motifs. Polyproline II was present only in minor amounts, 11–15%, and similar for turns, 10–16%. None of the secondary structure percentages was found to correlate with K_m or V_max values.

DISCUSSION

Collagens are a large family of extracellular proteins having more than 40 genes in vertebrates, the number of collagen genes in nematodes, like C. elegans, exceeding 150 (1–5, 41). A large number of post-translational modifications are associated with collagen biosynthesis. Lysyl hydroxylase is the enzyme catalyzing one of the post-translational modifications, hydroxylation of lysyl residues (1–5). It is not known why vertebrates, like human, mouse, and rat (13–21, 44), have three LH isoforms, whereas lower species like C. elegans (26) have only one orthologue for lysyl hydroxylase. Our earlier data indicate that all three isoforms are expressed in the same cells, although at different levels, when analyzed in various cultured human cell
mRNA analysis data suggest that the expression of LH1, LH2, and the \( \alpha \) subunit of prolyl 4-hydroxylase, another post-translational enzyme of collagen biosynthesis (1–5), are associated with collagen biosynthesis, showing similar co-regulation and statistically significant correlations with each other and with total collagen synthesis (42). The data argue against collagen-type specificity in LH1 and LH2 reactions. A different behavior was found for LH3. LH3 did not show any correlation with the other LH isoforms or with specific collagen types or total collagens (39). We have recently found that LH3 has additional functions, i.e. galactosylation of hydroxylysyl residues and glucosylation of galactosylhydroxylysyl residues (24–25, 27), although the extent of these reactions in vivo is still under investigation (43). Recent reports suggest LH2b to be a telopeptide lysyl hydroxylase (21, 45). There are no data so far to indicate the substrate requirements for the different LH isoforms. It is not known if the amino acid sequences surrounding the individual lysine residues determine which LH isoform...
Lysyl Hydroxylase Isoforms

**Table III**

K<sub>m</sub> (μM) and V<sub>max</sub> values (dpm/min) of different collagenous peptides measured in an LH activity assay

Equal amounts of human recombinant LH isoforms were used in the activity measurements and the equality was verified by immunoblotting the His-tagged proteins. The values are the mean values of three measurements. Dashes (——) indicate nonmeasurable values.

| Peptide | Sequence | Collagen polypeptide | Net charge |
|---------|----------|----------------------|------------|
| L12     | ERGLKHG R | αI(II)               | +3         |
| L22     | PREGK G E R | αIII(IV)             | +1         |
| L25     | QGKKGMDAG | αI(XII)              | 0          |
| L26     | DPGKGD K R | αI(IV)               | 0          |
| L27     | EAEMKG D KG | αI(XVII)             | -1         |
| L28     | ERGKEGSG | α2(VI)               | -1         |
| L29     | YPGKGD E G | α2(VI)               | -1         |
| L30     | IPGEKGDE G | α1(XIX)              | -2         |
| L31     | EQEKPEG | α1(XV)               | -3         |
| L32     | DPGLKGDGP | Synthetic            | -1         |
| L33     | SRPGKYKG | α1(V)                | +3         |
| L34     | TPGKGMWG | α2(IV)               | +1         |
| L35     | PVMGKGGS | α2(IV)               | +1         |
| L36     | SRGKGSRG | α1(VI)               | +2         |
| L37     | RSGDGPRB | α3(VII)              | +3         |
| L38     | ERGKEGVDG | α1(VII)             | -1         |
| L39     | PQQKGKEKG | α1(XIII)            | +1         |

| Peptide | Sequence | Collagen polypeptide | Net charge |
|---------|----------|----------------------|------------|
| L12     | ERGLKHG R | αI(II)               | +3         |
| L22     | PREGK G E R | αIII(IV)             | +1         |
| L25     | QGKKGMDAG | αI(XII)              | 0          |
| L26     | DPGKGD K R | αI(IV)               | 0          |
| L27     | EAEMKG D KG | αI(XVII)             | -1         |
| L28     | ERGKEGSG | α2(VI)               | -1         |
| L29     | YPGKGD E G | α2(VI)               | -1         |
| L30     | IPGEKGDE G | α1(XIX)              | -2         |
| L31     | EQEKPEG | α1(XV)               | -3         |
| L32     | DPGLKGDGP | Synthetic            | -1         |
| L33     | SRPGKYKG | α1(V)                | +3         |
| L34     | TPGKGMWG | α2(IV)               | +1         |
| L35     | PVMGKGGS | α2(IV)               | +1         |
| L36     | SRGKGSRG | α1(VI)               | +2         |
| L37     | RSGDGPRB | α3(VII)              | +3         |
| L38     | ERGKEGVDG | α1(VII)             | -1         |
| L39     | PQQKGKEKG | α1(XIII)            | +1         |

**Table IV**

The influence of NaCl on the hydroxylation capacity of LH isoforms

Equal amounts of human recombinant LH isoforms were used in the activity measurements and the equality was verified by immunoblotting the His-tagged proteins. The values are the mean values of three measurements. Dashes (——) indicate nonmeasurable values.

| Peptide | Net charge | Concentration of NaCl | Activity compared without NaCl |
|---------|------------|-----------------------|--------------------------------|
| L2      | +3         | 0.15                  | L1H 50 75 90 40                |
| L25     | 0          | 0.15                  | L1H 50 30 45 5                |
| L30     | -2         | 0.15                  | L1H 0 30 30 10                |
| L36     | +2         | 0.15                  | L1H 55 75 120 40              |

* L2, IKGIKIKIG; for the other peptide sequences, see Table III.
* Dashes indicate nonmeasurable values.

catalyzes the hydroxylation reaction.

We have used peptide scanning to screen the binding of in vivo occurring collagenous sequences to the LH isoforms. Our aim was to screen a large number of peptides, to access the roles of the LH isoforms in hydroxylation different collagen types, and to see if the amino acid sequences in close proximity to the lysyl residues in the Y position of X-Y-Gly sequences influence binding to the enzyme. This is the first study to analyze how the amino acid sequence of the peptide affects the binding of the peptide to the LH isoforms. For practical reasons we have limited the peptide size to nine amino acids, although it is probable that other amino acids farther from the lysine residue may also affect peptide binding in vivo. Our data revealed Pepspots as a useful assay to screen a large number of peptide candidates. Our activity assays suggest that many peptides bound to the active site of the enzyme. As seen from our data, there was no absolute sequence specificity for the binding of peptides to LH1, LH2a, LH2b, and LH3. There was an overlap of the binding of many peptides to the isoforms, but there were some peptides that preferentially bound to certain LH isoforms. These data confirm our earlier finding (42) that LH1, LH2a, LH2b, and LH3 lack collagen-type specificity. There was a tendency, however, for the isoforms to favor some collagen types; LH1 for instance binds peptides of type XI and XV collagen, and LH3 binds peptides of type VI, XI, and XII collagens.

We analyzed the frequencies of amino acids in sequences of collagenous peptides in vivo. Our data revealed proline as the most frequent amino acid in the X and Y positions of repeating X-Y-Gly sequences. If lysine was in the Y position of the general X-Y-Gly triplet (see Fig. 3a, I and 3), the frequency of charged amino acids was higher in the triplet after the lysine-containing triplet. We also compared the lysine-containing sub-sequence with the sub-sequence, in which another basic amino acid, arginine, was substituted for the lysine (Fig. 3a2) to determine whether the sub-sequence charge was the main determinant for other amino acids in the sequence evolution. As seen in the figures, the sequences remained quite similar. The comparison revealed, however, that evolution had to some extent changed the amino acid sequence surrounding the lysyl residues; one possible purpose of this might be to regulate the hydroxylation of the lysyl residues in the sequence.

The data in this study indicated that the net charge of the peptide is a very important determinant for the binding to all isoforms, a net charge of +1 or higher giving better binding than a net charge of 0 or less. The alternative splicing, which occurs in LH2, changed the peptide binding properties of LH2, suggesting that the alternatively spliced exon participates in the peptide binding to the active site, or it changed the three-dimensional structure of the active site and altered the binding properties. Individual amino acids in our peptide sequences also influenced the binding to the different isoforms. These included lysine in position 2, histidine in position 7, and arginine and lysine in position 8, which promoted the binding to all isoforms. Furthermore, glutamate in position 7 and proline in position 8 prevent peptide binding to all isoforms. Some amino acids in the sequence facilitated the binding to certain LH isoforms. For example, alanine in position 4, promoted binding to LH2b and LH3 but not to LH2a. The data clearly indicated that the amino acids surrounding the lysyl residue have an effect on the binding of the peptide to LH.

We calculated theoretical structures for the peptides in an aqueous environment. Comparison of these data to data from the binding assay revealed no structural motifs that would have affected the binding. Similar conclusions were reached by comparing the CD data with the values obtained from the activity assays. However, we found that peptides with a stronger preference for an aqueous environment bound LH more
often, with similar results for all isoforms. These peptides typically had a high net charge, which most probably is what caused the high binding. The finding also suggests that the LH binding site might be hydrophilic, i.e. not be a deep hydrophobic pocket.

It was found earlier (1, 10) that a 99-amino acid cyanogen bromide fragment from the α1 chain of rat skin collagen containing two Ala-Lys-Gly sequences was a particularly poor substrate for LH (LH1). We used 11 Ala-Lys-Gly-containing peptides originating from human α1 chain of type I collagen in our Pepspots screening (Table I). Our Pepspot data revealed that only two of the peptides bound to LH1, whereas nine remained unbound, in agreement with the earlier data (1, 10). None of the Ala-Lys-Gly peptides bound to LH2a (the binding to LH2b was negative for six peptides), whereas five of the peptides bound to LH2b. LH3 showed the same binding pattern as LH2b.

LH1 is known to be involved in the heritable connective tissue disorder, Ehlers-Danlos syndrome type VI (OMIM 225400 (46, 47)). Numerous mutations of LH1 have been characterized in patients with this syndrome. Muscular hypotonia, kyphoscoliosis, marfanoid habitus, fragile eyeballs, osteopenia, and occasionally vascular fragility are symptoms associated with the disorder. A recent study suggests that mutations of LH2 cause Bruck syndrome, and an increased expression level of LH2b is associated with systemic sclerosis (45). No mutations or disorders associated with LH3 have been reported so far. It is interesting to note that the reports of the LH1 mutations (48) indicate differences in under-hydroxylation between bone type I and cartilage type II collagens, type I being more dramatically affected than type II collagen. These data suggest that LH1 preferentially hydroxylates lysyl residues of type I collagen in bone. Similar results were obtained when the cross-linked peptides originating from type I or type II collagens, which were secreted to the urine of the patient with the Ehlers-Danlos syndrome VI, were analyzed (49). Our observations from mouse tissues (18) revealed that there are clear differences in the tissue-specific expression of the LH isoforms. LH1 is extremely highly expressed for instance in heart, liver, lung, skeletal muscle, and kidney, LH2 in the heart, lung, and kid-

**FIG. 7.** CD spectra of some synthetic peptide substrates of lysyl hydroxylase in water. a, L22; b, L27; c, L36.
ney, and LH3 in the heart, lung, liver, and testis. The tissue-specific expression of the LH isoforms partially explains the localization of enzyme defects to certain tissues, as seen in the specific expression of the LH isoforms. Spectra measurements showed that the collagenous sequence were key determinants in the binding peptides bound to each LH isoform. Our data with a two-dimensional Sammon map was used to visualize the different types of sensitive as the activity assay in the screening. A two-dimensional Sammon map was used to visualize the different types of peptides for possible interaction with the LH isoforms. Peptide sequences were key determinants in the binding of the peptides to LH1, LH2a, LH2b, and LH3. These enzymes were able to hydroxylate many sequences from different collagen types, but there was a preference of some peptides to be hydroxylated by a specific isoform.

### CONCLUSIONS

This is the first study to screen a large number of collagenous peptides for possible interaction with the LH isoforms. Peptide sequences were key determinants in the binding of the peptides to LH1, LH2a, LH2b, and LH3. These enzymes were able to hydroxylate many sequences from different collagen types, but there was a preference of some peptides to be hydroxylated by a specific isoform.

### Acknowledgments

We gratefully acknowledge Anna-Maija Koisti for expert technical assistance, Rabah Solyaman and Jussi Hepojoki for help in the Pepspot preparations, and Paivi Pirila for helping in the CD spectra measurements.

### REFERENCES

1. Kivirikko, K. I., Myllyla, R., and Pfahljaniemi, T. (1992) in Post-translational Modifications of Proteins (Harding, J. J., and Crabbe, M. J. C., eds) pp. 1–51, CRC Press, Inc., Boca Raton, FL.
2. Kiely, C. M., Hopkinson, I., and Grant, M.E. (1993) in Connective Tissue and Its Heritable Disorders (Royce, P. M., and Steinmann, P. M., eds) pp. 103–147, Wiley-Liss, Inc., New York.
3. Kadler, K. (1994) Protein Profile 1, 525–534.
4. Prokop, D. J., and Kivirikko, K. I. (1995) Ann. Rev. Biochem. 64, 403–434.
5. Bateman, J. F., Lamande, S. R., and Ramshaw, J. A. M. (1996) in Extracellular Matrix (Comper, W. D., ed) pp. 22–47, Harwood Academic Publishers Gmb, Amsterdam, The Netherlands.
6. Kivirikko, K. I., and Myllyla, R. (1979) Int. Rev. Connect. Tissue Res. 8, 23–72.
7. Reiser, K., McCormick, R. J., and Rucker, R. (1992) PASER J. 6, 2439–2449.
8. Knott, L., and Bailey, A. J. (1998) Bone (NY) 22, 181–187.
9. Rybänen, L. (1975) Biochim. Biophys. Acta 397, 50–57.
10. Kivirikko, K. I., Shudo, K., Sakakibara, S., and Prokop, D. J. (1973) Biochemistry 12, 4966–4971.
11. Jiang, P., and Ananthanarayanan, V. S. (1991) J. Biol. Chem. 266, 22603–22607.
12. Kivirikko, K. I., Ryhänen, L., Saarela, H., and Myllyla, R. (1994) Genomics 24, 464–471.
13. Valtavaara, M., Papponen, H., Pirttilä, A.-M., Hiltunen, K., Hederland, H., and Myllyla, R. (1997) J. Biol. Chem. 272, 6831–6834.
14. Valtavaara, M., Seppä, S., and Myllyla, R. (1998) J. Biol. Chem. 273, 12881–12886.
15. Valtavaara, M., Seppä, S., and Myllyla, R. (2001) Matrix Biol. 20, 137–146.
16. Pajunen, K., Rautavaamaa, M., Kosonen, T., and Kivirikko, K. I. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 10492–10496.
17. Rustalainen, H., Siipila, L., Kerkela, E., Pospiech, H., and Myllyla, R. (1999) Matrix Biol. 18, 325–329.
18. Rautavaamaa, K., Passeo, K., Heikkonen, J., and Kivirikko, K. I. (2000) Matrix Biol. 19, 73–79.
19. Rustalainen, H., Vanhatupa, S., Tampio, M., Siipila, L., Valtavaaara, M., and Myllyla, R. (2001) J. Biol. Chem. 276, 36158–36163.
20. Mercer, D. K., Nicol, P. F., Kemberc, M., and Robins, S. P. (2003) Biochim. Biophys. Acta 1570, 803–809.
21. Yeowell, H. N., and Walker, L. C. (1999) Matrix Biol. 18, 179–187.
22. Valtavaara, M. (1999) Acta Universitatis Ouluensis, Thesis A334, Oulu University Press, Oulu, Finland.
23. Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M., and Myllyla, R. (2000) J. Biol. Chem. 275, 4966–4971.
24. Wang, C., Luosjärvi, H., Heikkinen, J., Risteli, M., Uitto, L., and Myllyla, R. (2002) J. Biol. Chem. 277, 1559–1566.
25. Wang, C., Rautavaamaa, M., Kosonen, T., and Kivirikko, K. I. (1999) Matrix Biol. 18, 179–187.
26. Valtavaara, M. (1999) Acta Universitatis Ouluensis, Thesis A334, Oulu University Press, Oulu, Finland.
27. Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M., and Myllyla, R. (2000) J. Biol. Chem. 275, 36158–36163.
28. Wang, C., Luosjärvi, H., Heikkinen, J., Risteli, M., Uitto, L., and Myllyla, R. (2002) J. Biol. Chem. 277, 22603–22607.
29. Yeowell, H. N., and Walker, L. C. (1999) Matrix Biol. 18, 325–329.
30. Rautavaamaa, K., Passeo, K., Helaakoski, T., and Kivirikko, K. I. (2000) Matrix Biol. 19, 73–79.

### TABLE V

| Peptide* | Secondary structure percentages as calculated from CD spectra |
|----------|-------------------------------------------------------------|
|          | α-Helix | β-Sheet | Turn | Poly-Pro/II | Unordered |
| L12      | 5       | 33      | 13   | 15          | 34        |
| L22      | 1       | 30      | 12   | 13          | 44        |
| L23      | 1       | 29      | 12   | 13          | 45        |
| L24      | 3       | 25      | 12   | 14          | 46        |
| L25      | 3       | 18      | 12   | 14          | 36        |
| L26      | 7       | 34      | 13   | 15          | 32        |
| L27      | 6       | 36      | 13   | 15          | 30        |
| L28      | 0       | 28      | 10   | 11          | 50        |
| L29      | 4       | 30      | 12   | 13          | 41        |
| L30      | 1       | 25      | 12   | 13          | 49        |
| L31      | 2       | 30      | 13   | 12          | 44        |
| L32      | 2       | 27      | 12   | 12          | 47        |
| L33      | 12      | 48      | 16   | 11          | 28        |
| L34      | 1       | 29      | 11   | 12          | 47        |
| L35      | 1       | 29      | 12   | 12          | 47        |
| L36      | 4       | 31      | 12   | 14          | 39        |
| L37      | 6       | 40      | 11   | 15          | 29        |

* For the peptide sequences, see Table III; L26 data are missing due to a shortage of peptide.