

1989
ID
9

École nationale Supérieure des bibliothèques
Option conception et gestion de systèmes et
réseaux d'informations spécialisées
Université Claude Bernard Lyon I
D.E.S.S. d'informatique documentaire

LES GENES DES COLLAGENES ET LEURS ANOMALIES



Présenté par:
Catherine MAITRE

Sous la direction de:
Roland OUAZANA

25^e Promotion
Année universitaire:
1988-89

1989
ID
9

Ecole Nationale Supérieure des bibliothèques
Option conception et gestion de systèmes et
reseaux d'informati ons spécialisées
Université Claude Bernard Lyon I
D.E.S.S. d'informatique documentaire

LES GENES DES COLLAGENES ET LEURS ANOMALIES



Présenté par:
Catherine MAITRE

Sous la direction de:
Roland OUAZANA

25° Promotion
Année universitaire:
1988-89

1989
ID
9

PREMIERE PARTIE/ METHODE DE RECHERCHE

I- LE SUJET

- A) BUT DE LA RECHERCHE
- B) INTRODUCTION DU SUJET
- C) PRECISION DU SUJET

II- METHODE DE RECHERCHE

- A) CHOIX DES INSTRUMENTS
- B) LA RECHERCHE MANUELLE
 - 1) bibliographie de départ
 - 2) bibliographies courantes
- C) LA RECHERCHE AUTOMATISEE
 - 1) définition des mots-clefs et équation de recherche
 - é) interrogation des bases de données
 - 3) évaluation des résultats

III- OBTENTION DES DOCUMENTS PRIMAIRES

DEUXIEME PARTIE/ SYNTHÈSE

I- BREF RAPPEL DE LA STRUCTURE DU COLLAGÈNE

- A) DIFFÉRENTS TYPES DE COLLAGÈNE
- B) STRUCTURE ET SYNTHÈSE DU COLLAGÈNE

II- LES GÈNES CODANT POUR LES CHAÎNES PRO-ALPHA

- A) LOCALISATION
- B) STRUCTURE
- C) ÉVOLUTION : Quelques hypothèses

III- LES MALADIES HÉRÉDITAIRES DES COLLAGÈNES

- A) LE SYNDROME D'EHLERS-DANLOS
 - 1) expression clinique
 - 2) les anomalies responsables
 - a. SED type IV
 - b. SED type VIIB
- B) L'OSTÉOGENÈSE IMPARFAITE
 - 1) expression clinique
 - 2) les anomalies responsables
 - a. OI type II
 - b. OI type I
 - c. OI type III
 - d. OI type IV
- C) LE SYNDROME DE MARFAN

TROISIEME PARTIE/ BIBLIOGRAPHIE CLASSEE

I-LE SUJET

A) BUT DE LA RECHERCHE

Ce sujet m'a été proposé par Roland Ouazana, maître de conférences à l'université Claude Bernard Lyon 1. C'est celui d'un cours qu'il donne à des médecins en voie de spécialisation dans le cadre du C.E.S. (certificat d'études spéciales) de génétique du Pr J.M.Robert.

Il souhaite le mettre à jour, et c'est dans cette optique que j'ai réalisé pour lui ce travail.

Quant à moi, j'ai accepté avec intérêt, ayant une formation initiale scientifique.

B) INTRODUCTION DU SUJET

Le tissu conjonctif est très répandu dans le corps humain, il forme l'essentiel de la peau et des structures articulaires comme les tendons, les ligaments, le cartilage; c'est aussi un constituant de la cornée de l'oeil, des os, des vaisseaux et de nombreuses enveloppes d'organes.

Or le composant principal de ce tissu est une protéine, ou plutôt une famille de protéines: les collagènes.

Des anomalies à différents niveaux de la synthèse de ces protéines pourront donc avoir des conséquences multiples pour tout l'organisme, qui se traduiront par des maladies très nombreuses et variées.

Dans ce cadre, notre sujet se limite aux études sur la structure, la localisation chromosomique et l'évolution des gènes des différents types de collagène ainsi qu'à l'identification de défauts au niveau de ces gènes, et de leur responsabilité dans certaines maladies héréditaires du tissu conjonctif comme l'ostéogénèse imparfaite.

Ces dix dernières années, les progrès de la recherche et des techniques en biochimie, en biologie cellulaire et moléculaire, ont permis de faire avancer considérablement les connaissances dans ce domaine.

C) PRECISION DU SUJET.

Une première approche m'a été nécessaire pour mieux comprendre le sujet, et en particulier pour distinguer les 4 maladies où, actuellement, un défaut génétique du collagène est en cause. Il s'agit de:

- Le syndrome d'Ehlers danlos type IV et VIIB.
- L'ostéogénèse imparfaite.
- La maladie de Marfan.
- Le collagenome cutané familial.

Ceci a été possible d'abord grace aux explications du Pr Ouazana lors de notre première entrevue, et à la lecture de quelques articles qu'il m'a procuré. J'ai pu aussi assister à son cours.

Ensuite, en salle de lecture de la bibliothèque universitaire de médecine de Lyon Rockefeller, la consultation de quelques dictionnaires :

- MANUILA et collab. Dictionnaire français de médecine et de biologie. Masson, 1970-72. 4 volumes.
- Dictionnaire de médecine. Flammarion, 1987.
- JABLONSKI S. Illustrated dictionary of eponymic diseases and syndromes and their synonyms. Saunders, 1969.

m'a permis:

- De traduire le sujet en anglais.
- D'éliminer les synonymes rencontrés. En effet deux maladies étant connues depuis fort longtemps (le syndrome d'ehlers-Danlos, par exemple, a été décrit pour la première fois en 1682!) de nombreux synonymes ont été utilisés au cours de leurs descriptions successives. N'effectuant pas une recherche rétrospective, j'ai dû éliminer ceux qui ne sont plus usités actuellement. Ainsi j'utiliserai:

Ostéogénèse imparfaite pour:
 Syndrome de Lobstein
 Syndrome de Vrolick
 Ostéopsathyrosis
 Maladie des os de verre

Syndrome d'Ehlers danlos pour:
Cutis laxa
Syndrome de Meekrin
Fibrodysplasie élastique généralisée

II-METHODE DE RECHERCHE

A) CHOIX DES INSTRUMENTS

Plusieurs considérations, liées au sujet, ont guidé mon choix.

Ce sujet, je l'ai dit, est l'objet de nombreux travaux depuis une dizaine d'années. Je m'attends à trouver de nombreux articles récents, actualisant le cours. Certaines caractéristiques communes aux domaines scientifiques font que les périodiques et les congrès sont les supports principaux des informations. L'anglais est la langue la plus couramment utilisée dans ces communications.

Il s'agira donc de rechercher surtout des articles de périodiques ou des rapports de congrès, publiés après 1980 et non seulement en français mais en anglais.

M.Ouazana nous confirme son intérêt pour ce type de documents et nous indique plus particulièrement les publications de certains chercheurs (PROCKOP D.J., CHU M.-L...) sa préférence va à une recherche tendant à être exhaustive quitte à obtenir de nombreuses références. Tout ceci nous oriente vers une recherche documentaire informatisée.

B) LA RECHERCHE MANUELLE

1. BIBLIOGRAPHIE DE DEPART

Nous recherchons en premier lieu quelques documents faisant le point sur certains aspects de la question et contenant une bibliographie de base: en anglais "review". De tels articles se trouvent:

- Dans l'E.M.C. (encyclopédie médico-chirurgicale).
- Dans l'index medicus (forme papier correspondant à la base de données MEDLINE-décrite au par.IIC12a) partie "bibliography of medical review"

2.BIBLIOGRAPHIES COURANTES.

La recherche dans les bibliographies courantes spécialisées (par ex. l'index medicus cf B1) présente l'avantage d'être gratuite mais je n'ai pas persévéré dans leur consultation à cause de certains inconvénients:

- La perte de temps que cela occasionne, en particulier, pour l'année en cours, il faut éplucher chaque volume mensuel.
- L'impossibilité de croiser plusieurs termes (ex. ici collagène et gène)

C) LA RECHERCHE AUTOMATISEE

1 DEFINITION DES MOTS-CLEFS ET EQUATION DE RECHERCHE

Après avoir choisi avec M.Ouazana les mots et concepts principaux du sujet, j'ai défini une première équation de recherche en anglais:

- 1 COLLAGEN? OU PROCOLLAGEN?
- 2 1 ET GENE+
- 3 2 ET >1980
- 4 3 ET EHLERS-DANLOS SYNDROME
- 5 3 ET OSTEOGENESIS IMPERFECTA
- 6 3 OU 4 OU 5

En utilisant:

- ?- comme signe de troncature limitée:remplace une lettre quelconque.
- + comme signe de troncature illimitée:remplace un nombre quelconque de lettres.
- ET-OU-SAUF-

Mais chaque base de données ayant sa structure propre et étant interrogeable par l'intermédiaire de serveurs différents ayant eux aussi leur propre langage d'interrogation, cette équation générale a du être reformulée pour chaque interrogation.

D'autre part, lors d'une première interrogation le nombre de références étant trop important, (de l'ordre de 10 000 pour la question 2) j'ai été amenée à introduire des limitations supplémentaires:

- GENE+ devient GENE?
- Limitation aux résultats concernant le genre humain.
- Limitation aux articles publiés en français ou en anglais.
- Pondération des termes GENE et COLLAGEN qui seront ainsi les sujets principaux des articles retrouvés.

2 INTERROGATION DES BASES DE DONNEES

Trois bases de données ont été sélectionnées d'après leur importance et les domaines couverts.

2A MEDLINE

- Présentation de la base

MEDLINE = MEDical litterature analysis and retrieval system on-LINE.

PRODUCTEUR: National library of medicine (N.L.M.), Bethesda, U.S.A. avec la coopération de l'IMA-INSERM pour la France.

DOMAINES: Biomédical et en particulier biologie, génétique, biochimie qui sont intéressants pour le sujet.

DONNEES: Les articles de 3200 périodiques sont indexés, dont 100 français; 69% des références sont en anglais.

NOMBRE DE DOCUMENT: 5,3 millions+300 000 par an.

DEBUT: 1966.

AIDE A LA RECHERCHE: La recherche des descripteurs se fait dans le thésaurus MeSH (medical subjects headings).

- Interrogation

J'ai interrogé MEDLINE à la B.U. médecine de Lyon avec Melle Seguin, sur le serveur DATA-STAR. Nous avons interrogé une partie seulement de la base: MEYY, qui contient les références depuis 1977.

2B BIOSIS

-Présentation de la base

BIOSIS=BioSciences Information Service

PRODUCTEUR:BIOSIS, Philadelphie, U.S.A.

DOMAINES:Biologie, médecine, en particulier biologie moléculaire, génétique qui sont intéressants pour le sujet.

DONNEES:9000 périodiques sont indexés ainsi que des actes de congrès, rapports de recherche, ouvrages et brevets américains.

NOMBRE DE DOCUMENTS:5,7 millions+480 000 par an.

DEBUT(sur DATA-STAR):1970.

AIDE A LA RECHERCHE:BIOSIS search guide

-interrogation

J'ai interrogé BIOSIS au cours d'un stage de formation à l'URFIST de Lyon (service interuniversitaire de formation à l'interrogation de banques de données).Une possibilité intéressante: l'utilisation de "concepts codes", qui recouvrent un sujet large, et permettent d'obtenir les références appartenant à ce domaine.Ici j'ai utilisé CC03502: genetics and cytogenetics general,et CC03508:genetics and cytogenetics human, à la place de GENE?.

2C PASCAL

-Présentation de la base.

PRODUCTEUR:INIST-CNRS,Centre de documentation scientifique et technique, Paris, France.

DOMAINES:Sciences et techniques, en particulier sciences de la vie et médecine qui correspondent au sujet.

DONNEES:Articles de périodiques, ainsi que de rapports scientifiques, thèses, comptes rendus de congrès.Interrogation possible en français, anglais et espagnol(à partir de 1987)

NOMBRE DE DOCUMENTS:6,5 millions + 430 000 par an.

DEBUT: 1973.

AIDE A LA RECHERCHE: Thésaurus bilingue français-anglais.

-Interrogation.

J'ai interrogé PASCAL à L'E.N.S.B. avec C.ANDRE, sur le serveur TELESYSTEMES/QUESTEL.

3.EVALUATION DES RESULTATS.

Voici les résultats de ces interrogations:

| | MEDLINE | BIOSIS | PASCAL |
|--------------------------------|---------|--------|--------|
| 1 COLLAGEN? OU PROCOLLAGEN? | | | |
| 2 1 ET GENE? MJ | | | |
| 3 2 ET LIMITATIONS | 25 | 4 | 130 |
| 4 3 ET EHLERS-DANLOS | 5 | 20 | 3 |
| 5 3 ET OSTEOGENESIS IMPERFECTA | 19 | 16 | 17 |
| 6 3 OU 4 OU 5 | 44 | 27 | 33 |

Comme il était à prévoir un certain nombre de recoupements ont eu lieu entre mes différentes sources .

J'ai choisi les références pertinentes en collaboration avec M.OUAZANA. En considérant uniquement les références obtenues par les interrogation de bases de données, le taux de pertinence obtenu est:

| | | |
|------------------------------|-----|-------------|
| documents pertinents retenus | 56 | |
| ----- | = | ----- = 53% |
| documents trouvés | 104 | |

En fait, la plupart des documents retrouvés étaient en rapport avec le sujet, mais nous ne les avons pas retenus car ils n'apportaient que des informations dépassées ou redondantes. Le taux de pertinence est donc en réalité supérieur à 53%, ce qui est normal compte tenu des nombreuses limitations que j'ai été obligée de faire .

III-OBTENTION DES DOCUMENTS PRIMAIRES.

C'est cette dernière étape qui s'est avérée être la plus longue.

Pour la localisation des périodiques:

- J'ai utilisé le CCN sur CDROM.
- J'ai consulté les fichiers de périodiques des B.U. médecine et B.U. sciences de Lyon, la plupart des périodiques s'y trouvaient.

Pour l'obtention des documents:

- La plupart ont été photocopiés dans les bibliothèques et centres de documentation de Lyon suivants:
 - .BU médecine et BU sciences
 - .laboratoire de biologie cellulaire, Lyon I.
 - .Service de génétique de l'Hotel Dieu.
- Les autres ont été demandés par le prêt inter.

CONCLUSION

A l'issue de cette première partie de mon travail, la recherche bibliographique, je peux déjà tirer quelques conclusions.

J'ai obtenu de nombreuses références qui sont pour la plupart en anglais et qui font appel à des connaissances que je maîtrise peu, aussi ai-je choisi, en accord avec M.OUAZANA et compte tenu de l'objectif de ce travail qui est la mise à jour de son cours, de réaliser une synthèse assez courte, et d'autre part de privilégier les informations nouvelles par rapport au cours déjà existant.

La liste de références est ordonnée selon un classement systématique correspondant aux différentes parties du sujet, avec un sous classement alphabétique par auteurs. Ceci pour plus de clarté d'une part, et pour être plus utile, d'autre part. En effet, sous cette forme, il sera plus facile, pour moi lors de la rédaction et pour M.OUAZANA par la suite, de s'y reporter ou de la compléter.

SYNTHESE

I - BREF RAPPEL DE LA STRUCTURE DU COLLAGENE

Le collagène (Co) est la plus abondante des protéines du corps humain. Cette molécule est composée de trois chaînes alpha associées en triple hélice, chacune d'une longueur approximative de 1000 acides aminés (AA).

A / DIFFERENTS TYPES DE COLLAGENES

Actuellement, une vingtaine de chaînes alpha ont été identifiées, qui forment plus de 11 types de Co.

Ces types sont habituellement classés en 3 groupes :

- Le groupe 1, de chaîne $M_r \gg 95000$, avec une triple hélice de 300 nm (types I, II, III, V, K).
- Le groupe 2, de chaîne $M_r \gg 95000$, mais dont la triple hélice est interrompue par des segments non hélicoïdaux (types IV, VI, VII, VIII).
- Le groupe 3, de chaîne $M_r < 95000$ (types IX, X).

B / STRUCTURE ET SYNTHÈSE DU COLLAGENE

Une molécule de collagène est donc constituée de 3 chaînes alpha. Sauf pour les telopeptides, courtes régions des extrémités d'une vingtaine d'AA chacune, la structure polypeptidique de ces chaînes est caractéristique : c'est la répétition d'un triplet Gly-X-Y, Gly étant la glycine, X et Y étant le plus souvent de la proline, de l'hydroxyproline, de l'alanine.

Lors de la synthèse du Co, les gènes donnent d'abord naissance à des chaînes pro-alpha, qui sont de longs polypeptides de 1500 AA environ. Puis ces chaînes pro-alpha vont s'assembler par 3 pour former la molécule de procollagène, celle-ci comprend aux deux extrémités de la triple hélice de 297 nm de long sur 1,5 nm de diamètre, deux propeptides :

- Le propeptide N-terminal comporte successivement un domaine globulaire, une triple hélice et un autre domaine globulaire.

- Le propeptide C-terminal, plus court, comporte seulement un domaine globulaire.

Pour finir, le procollagène a aussi un peptide signal.

La molécule de Co est obtenue par perte du peptide signal

Les différents types de Collagènes

TERMINOLOGY USED FOR COLLAGENS*

| Current terminology | Older terminology | Number of unique chains |
|---------------------|---|-------------------------|
| Type I | Collagen | 2 |
| Type II | Cartilage collagen | 1 |
| Type III | Embryonic collagen | 1 |
| Type IV | Basement membrane collagen | 2 |
| Type V | A-B collagen | 3 |
| Type VI | SC collagen | 3 |
| | Intima collagen | |
| Type VII | LC collagen | 1 |
| Type VIII | EC collagen | 1 |
| Type IX | Type M collagen | 3 |
| | HMW-LMW collagen | |
| Type X | G collagen | 1 |
| | SC cartilage collagen | |
| Type K (XI) | 1 α , 2 α , 3 α collagen | 3 |

* SC, short chain; LC, long chain; EC, endothelial cell; HMW, high molecular weight; LMW, low molecular weight.

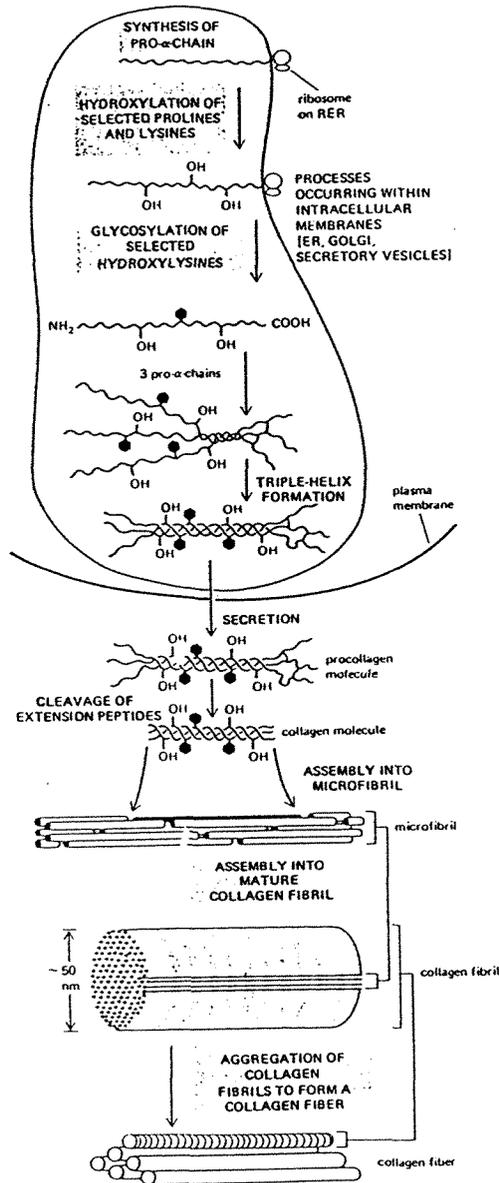
GROUP 2 COLLAGEN MOLECULES

| Collagen | Chains | Chain M_r | | Molecular species |
|-----------|-----------------|-------------|-----------|--|
| | | Procollagen | Collagen | |
| Type IV | $\alpha 1(IV)$ | 185,000 | 185,000 | $[\alpha 1(IV)]_2\alpha 2(IV)$ |
| | $\alpha 2(IV)$ | 170,000 | 170,000 | $[\alpha 1(IV)]_3$ $[\alpha 2(IV)]_3$ |
| Type VI | $\alpha 1(VI)$ | | 140,000 | $[\alpha 1(VI)\alpha 2(VI)\alpha 3(VI)]$ |
| | $\alpha 2(VI)$ | (240,000) | 140,000 | $[\alpha 1(VI)]_3$ |
| | $\alpha 3(VI)$ | | 140,000 | |
| Type VII | $\alpha 1(VII)$ | ? | >170,000 | $[\alpha 1(VII)]_3$ |
| Type VIII | ? | (180,000) | (180,000) | ? |

GROUP 1 COLLAGEN MOLECULES

| Collagen | Chains | Chain M_r | | Molecular species |
|----------|-----------------|-------------|----------------|---------------------------------------|
| | | Procollagen | Collagen | |
| Type I | $\alpha 1(I)$ | 140,000 | 95,000 | $[\alpha 1(I)]_2\alpha 2(I)$ |
| | $\alpha 2(I)$ | 125,000 | 95,000 | $[\alpha 1(I)]_3$ |
| Type II | $\alpha 1(II)$ | 140,000 | 95,000 | $[\alpha 1(II)]_3$ |
| Type III | $\alpha 1(III)$ | 140,000 | 95,000-110,000 | $[\alpha 1(III)]_3$ |
| Type V | $\alpha 1(V)$ | 240,000 | 115,000 | $[\alpha 1(V)]_2\alpha 2(V)$ |
| | $\alpha 2(V)$ | 160,000 | 125,000 | $[\alpha 1(V)]_3$ |
| | $\alpha 3(V)$ | ? | ? | $[\alpha 1(V)\alpha 2(V)\alpha 3(V)]$ |
| Type K | 1 α | ? | ? | ? |
| | 2 α | ? | ? | |
| | 3 α | ? | ? | |

Biosynthèse du Collagène



Biosynthèse du Collagène

[La Recherche - n° 120 - mars 1984]

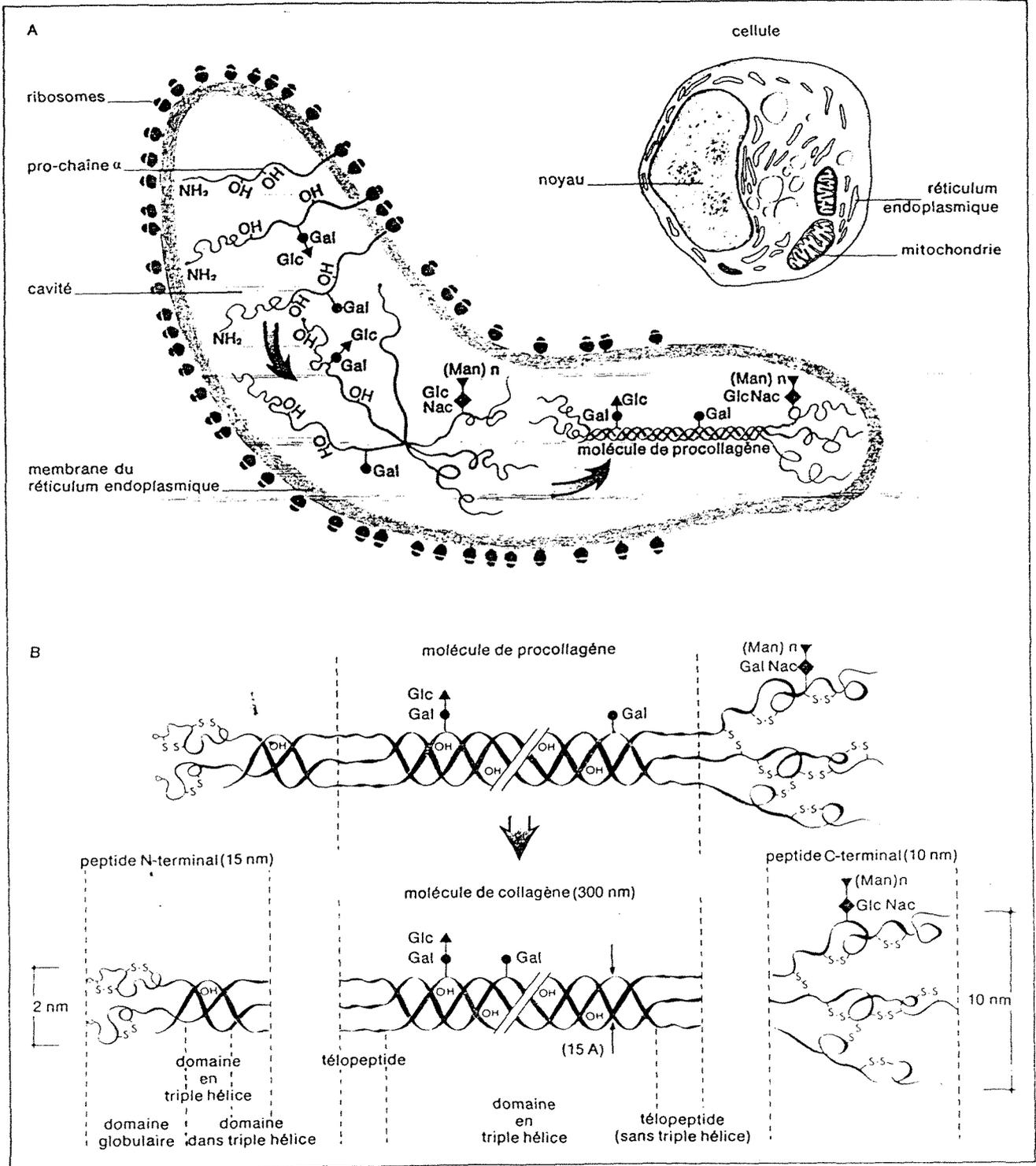


Figure 5. Comme toutes les protéines, le collagène est fabriqué à l'intérieur des cellules par traduction du message génétique sur les ribosomes.

(A). Ce schéma montre une cellule avec ses différents organites: le collagène étant destiné à l'exportation est sécrété dans des cavités de réticulum endoplasmique qui vont donner des vacuoles. Celles-ci se sont ensuite déversées à la surface de la cellule. A l'intérieur d'une cavité du réticulum endoplasmique, chacune des chaînes polypeptidiques de la triple hélice est synthétisée sur un ribosome, puis s'associe pour former la molécule définitive. En réalité, la chaîne polypeptidique qui sort du ribosome est une pro-chaîne α , plus longue que la chaîne définitive, et qui sera coupée. L'assemblage dans la cavité des pro-chaînes α donne une molécule de procollagène.

(B). Ces schémas détaillent la structure des molécules de procollagène et de collagène, organisée en différentes portions. On remarquera que les peptides terminaux contiennent des petits domaines organisés en triple hélice. Inversement, la molécule de collagène a ses parties terminales libres: ce sont les télopeptides. Glc et Gal représentent les sucres glucose et galactose, qui sont liés aux hydroxylsines de la triple hélice. Man et GlcNac représentent les sucres mannose et N-acétylglucosamine, qui ne figurent que sur les peptides terminaux. (D'après Prockop et al.)

et clivage des deux propeptides.

A l'extérieur du fibroblaste, les molécules de Co vont s'organiser en fibrilles puis en fibres et faisceaux de fibres.

On comprend qu'au cours de cette biosynthèse complexe, de nombreuses anomalies puissent arriver.

II - LES GENES CODANT POUR LES CHAINES PRO-ALPHA

A / LOCALISATION

On connaît maintenant la situation d'un certain nombre de ces gènes : sept d'entre eux sont situés sur cinq chromosomes. Ils sont donc relativement dispersés dans le génome.

Pour les Co de type I, les gènes codant pour les chaînes pro-alpha1(I) et pro-alpha2(I) sont situés respectivement sur le bras long du chromosome 17 et du chromosome 7.

Pour le Co de type II, les gènes codant pour la chaîne pro-alpha1(II) est situé sur le chromosome 12.

Pour les Co type III et type V, les gènes codant pour les chaînes pro-alpha1(III) et pro-alpha2(V) sont situés sur le long bras du chromosome 2.

Pour le Co type IV, les gènes codant pour les chaînes pro-alpha1(IV) et pro-alpha2(IV) siègent sur le chromosome 13.

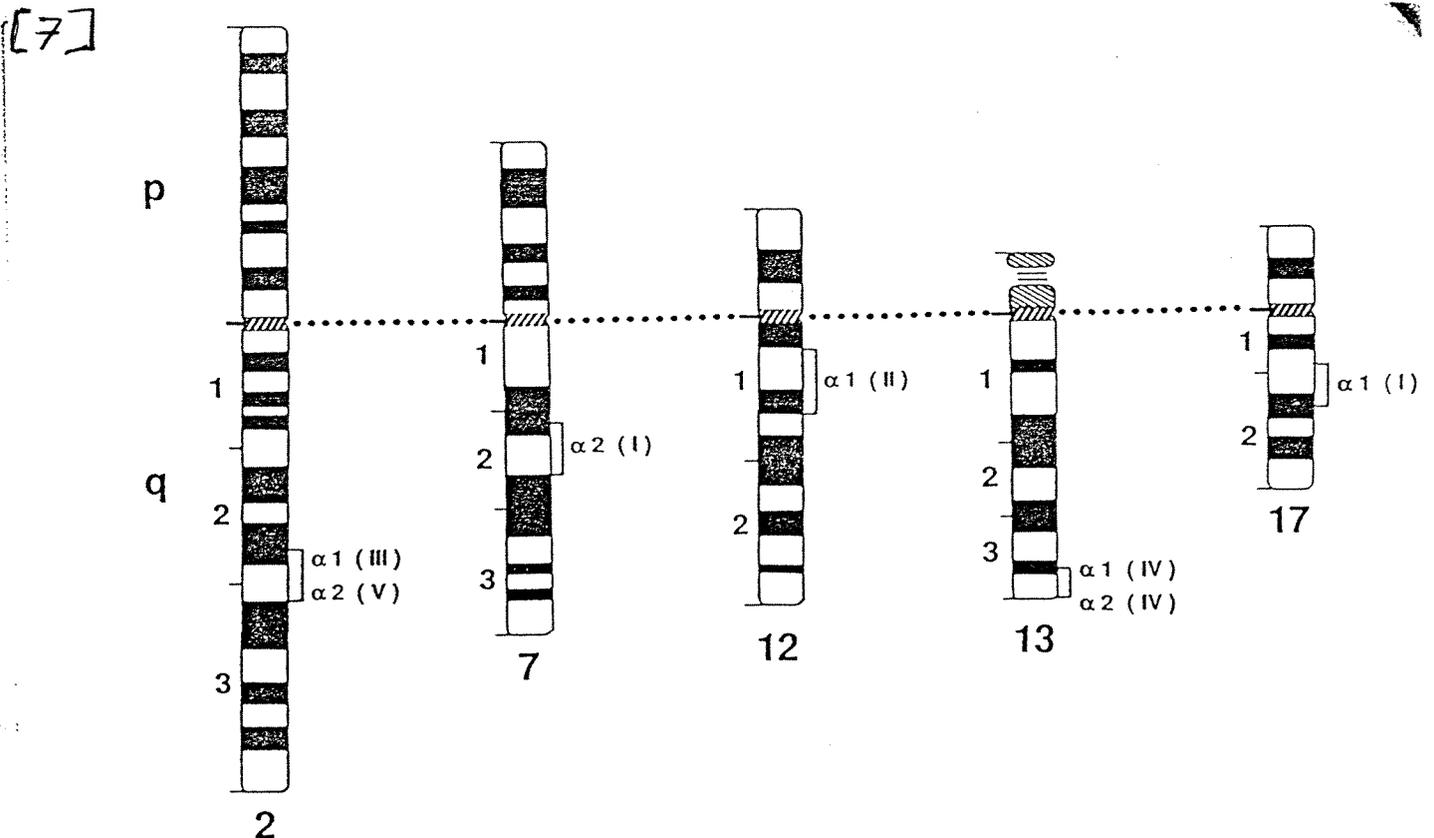
B / STRUCTURE

Chez l'homme certains syndromes semblent être liés à des dysfonctionnements des gènes du Co ou de leurs produits. L'analyse de cette famille de gènes au niveau moléculaire est donc importante pour comprendre le mécanisme de ces états pathologiques.

Des portions de plusieurs gènes du Co dans différentes espèces ont été identifiées. 90 % du Co des vertébrés est du type I, ce qui explique que les gènes codant pour les chaînes pro-alpha1(I) et pro-alpha2(I) furent les premiers à être identifiés, d'abord chez le poulet, puis chez l'homme.

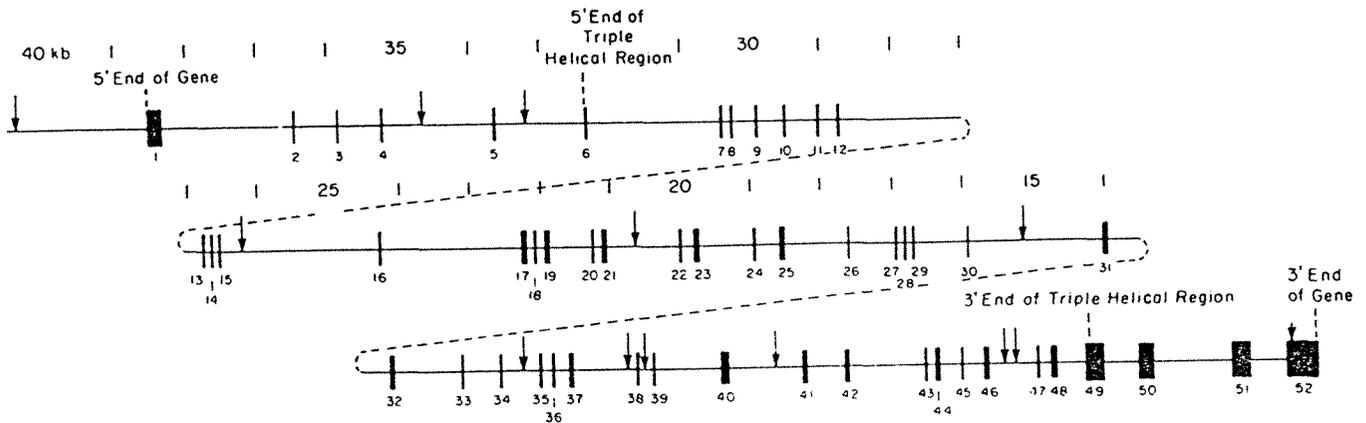
LOCALISATION DES GENES

| CHROMOSOME | REGION | GENE |
|------------|-------------------------------|--|
| 17 | q21.3 → q22 | alpha1(I) (COL1A1) |
| 7 | q21.3 → q22 | alpha2(I) (COL1A2) |
| 12 | q13.1 → q13.2 q (ou q14.3) | alpha1(II) (COL2A1) |
| 2 | q24.3 → q31 (ou q23) | alpha1(III) (COL3A1 et COL5A2) et alpha2(V) |
| 13 | q33 → q34 | alpha1(IV) (COL4A1 et COL4A2) et alpha2(IV) |

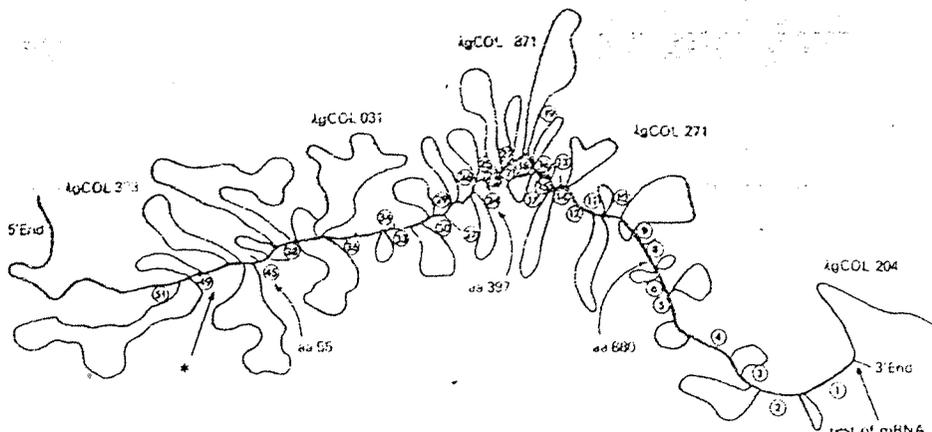
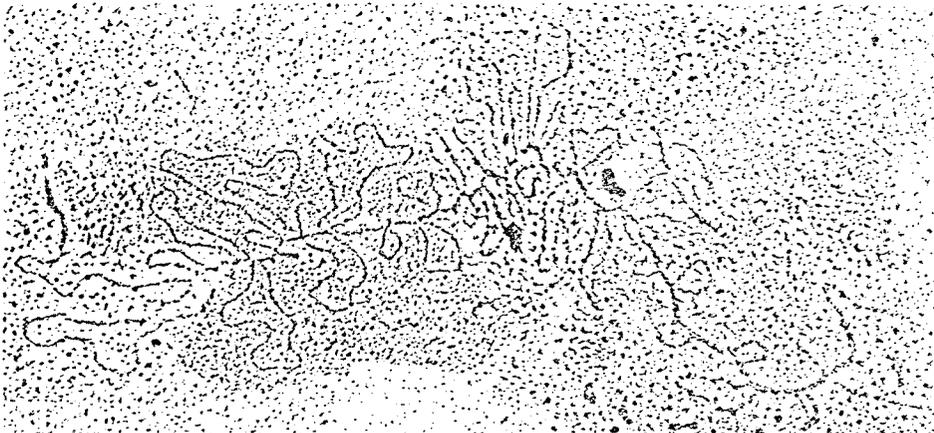


REPARTITION DES INTRONS ET DES EXONS : SCHEMA

[12] Gène du Co pro- $\alpha_2(I)$



[12] Em microscopie electronique



Les exons sont numérotés, les introns sont les boucles.

Ces gènes sont caractérisés par leur longueur (18 000 paires de bases pour la chaîne pro-alpha1(I) humaine et 38 000 pour la chaîne pro-alpha2(I)) et surtout, l'existence d'un très grand nombre d'introns, alternant avec les régions codantes: les exons.

Nous allons décrire la structure du gène codant pour la chaîne pro-alpha2(I) du poulet:

Ce gène est long de 38 kb, et composé d'au moins 51 exons, qui représentent seulement 10 % de la taille du gène. La taille des exons varie de 48 paires de bases à 327 paires de bases, la taille des introns varie de 81 à 2200 paires de bases. Si on divise ce gène en régions codant pour les 6 domaines de la molécule de proCo voici sa structure:

* Région codant pour l'extrémité N-terminale

Six exons codent pour cette région, d'une longueur de 400bp. Elle contient les séquences (TATAAATA box et CCATT box) essentielles à l'initiation de la transcription. De courtes séquences ont été remarquées, dont la structure pourrait expliquer la formation de structures en épingle à cheveux.

Plus de 75 % du propeptide consiste en triplets de Gly-X-Y formant une triple hélice, cette région étant entourée de 4 et 8 résidus non hélicoïdaux, le peptide-signal étant formé de 26 résidus.

* Région codant pour La triple hélice

41 exons codent pour cette partie. Leur structure est très homogène, chacun étant un multiple des 9 paires de bases qui codent pour le triplet Gly-X-Y. La moitié au moins de ces exons est long de 54 paires de bases, soit 6 fois les 9 paires de bases correspondant à 18 AA.

* Région codant pour l'extrémité C-terminale

4 exons codent pour cette partie. L'exon 1 est le plus long, l'exon 4 code aussi pour les 15 derniers résidus de la triple hélice, les 15 résidus du telopeptide, et le début du propeptide C-terminal. Ils correspondent à 4 domaines fonctionnels du propeptide: les résidus de cystéine impliqués dans les liaisons entre et intra-chaines, le site de clivage par la carboxypeptidase, le site de fixation du carbohydrate.

12] GENE DU Co pzo-2(I)

Les AA correspondant à la séquence d'ADN sont en majuscules.

PROMOTER

-406
 -286
 -166
 -46

EXON 49

1
 121
 1
 1
 121
 241

(-0.56 KB INCLUDING EXON 48 ?)

(-0.45 KB INCLUDING EXON 48 ?)

EXON 48

204
 AAGTGAGTGAG
 1:ValSerGlu

(-0.45 KB INCLUDING EXON 48 ?)

Detailed description: This block contains a DNA sequence with annotations. It starts with a 'PROMOTER' region with positions -406, -286, -166, and -46. Below this is 'EXON 49' with positions 1, 121, 1, 1, 121, and 241. The sequence is presented in several lines, with some regions enclosed in boxes. A protein sequence is shown below the DNA: MetLeuSerPheValAspThrArgIleLeuLeuLeuLeuAlaValThrSerTyrLeuAlaThrSerGlnHis. There are two distance markers: '(-0.56 KB INCLUDING EXON 48 ?)' and '(-0.45 KB INCLUDING EXON 48 ?)'. At the bottom, 'EXON 48' is indicated with position 204 and the sequence AAGTGAGTGAG, which corresponds to the amino acids ValSerGlu.

(Suite pages suivantes) ->

1
8885tggatggccagtcgcccgttaantttcattgacggaggattatagagactcaggcgtgatttccattatgctatattttaggctataataagtalaaaattgccac
121
ctttcaacagagggtcfaattgttgacataatattcctataataaacctcttaacccttgctagtaatttanaagacncabggcatgcttaagttatgggtgclgacagaattgtag
241
ggttsaaaggagatgacgaanttaagtttgacattgtatttttttaaaaactgataaatttgacccttttttccctantctctgcttattctctccag

EXON 47 215
GCATCTGCAGGGCGGAAG
AlaSerAlaGlyArgLys

1
gtaagcagcigcgtttttctgtgfgggatgtagatgacatctaaatattcctgancagagtaatttttgggaattgttggatgncasagtggaclaaagtgttatttcccagtgct
121
cactgagatgagtaagaatgataaggagagacaaatgggtgcatlagcttagcaaaacggtaaatgcaagccttaggattttagctgclttatccnagagcagactesgcatactgaga
241
accacactgcatacagaagaattcttcttcaacaggagtttcttggagggtttcatgagttgacccgtacagcngtaccatcacagatctgaaggataangtagn!accactttat
361
tttgaagaagctctctgtitaantaacagacagttatcatggagacaaaacaaactgccagagtaacttctgtgacagatattcagctgagtcagtggaacngtaactantttattc
481
tagtagcatatcgggctaaagatacaccgtggaaaaatagtgtagaacctatccttgacgtctcttcttcttctcagtaaccasatpntcttctgttttccctgatacctagtt
601
tcigtggtaactaagttctgttttcttcccaacag

EXON 46 233
GGCCCTAGAGGAGACAAAGGCCACAGGGAGAAAGG
GlyProArgGlyAspLysGlyProGlnGlyGluArg

1
gtaaggagtagattgacatttatttttgggtcttgaacagttgttggaaatcagtagganaagctaaatttagctatttaaatagtttaattgactttgt
(-0.52 KB)

1
cattgtcttactcgtgtagtgtgtgtttgtattttttgggtattgtctcttgaanagtaaaatccctgtctctgtaggaaagctattgaaagtctctttaggtcttagtatgacaa
121
acacaaatgagaactactaggagagtgatctataatggtcagclggatatttgccttggactaccctgcttaaatctgttttautgcagtagagatccgtactttgttctgaaagacgaa
241
gcgtaattctgcgcttttgaanctctaaagattcnaatgtggttttctctngatcaagctcactcgaanactgtgtagcaaaccttgaccnzcctttgtcattgaggtgacagagc

361
gtaaggcaaatattggcc (-0.45 KB)

1
accggcagcagggcaaggcagtgtaaacctcaggctatgctcattttctgacattagcctataatgtatattccgcatatgacagtgctacagttatgtgctgtagtaacctgattgtctca
121
tcactgcagctttcctaaagctacttttaagttcactttatnaatcctcgttttgtctctag

EXON 45 269
GGTCCACCAGGTCCACAGGCACAGATGGTGAAGACGGCCACACAGGTCTCCAGGCCCCCTGGTCTCCAGGTCTTGGCGGA
GlyProProGlyProProGlyArgAspGlyGluAspGlyProProGlyProProGlyProProGlyLeuGlyGly

1
gtaagacatgcaaacctttttatfacigaaatgataaattttgclttgtgtagtaactaaagactaantgtatatttactggcacgctgacattcaactggagtagtagcgaacactcttg
121
tgtcaattggcactcagattagtggtctcagttccagtgcaactctccttaggatctgtatcttccacagcaactttggat

(-0.48 KB)

1
attcactcaccctcagtgtaaatagctnaasattttttgtattagcaaaccccaactgtttttattagttttcttatttcaactgctantaaatnaaacagtgagaaneagatt
121
ttcaattgtgtgcatgtaactctttccatttttagaggagagggctegatgtgtgttagtctataaggctttcattgattaaagggtgaaaaatccagtgcaatgacctgcatgctctac
241
agtccatttg

(-0.25 KB)

1
acccttccactgaatcaagctcttctgaaaccagatgaatgaaatgtatttggtaaaagggtcagtgtaactgagttttagtattgtgtcttcaacactctgtctagaaatagggtta
121
tggatcactgtagataggggttatcagtgctttctgaaatgaagattgtagtgaactttatctcaantttttgtctctcag

EXON 44 (JUNCTION AA# -16/3) 353
AATTTGCTGCTCAGTATGATCCATCTAAAGCGGCTGACTTTGGCCCGGACCTATG
AsnPheAlaAlaGlnTyrAspProSerLysAlaAlaAspPheGlyProGlyProMet

1
gtaagataatgatttaacacttggtaacttgcataaacatgtattttaagtgtactccag (-1.35 KB)

1
attcaaaaggcttctgacagctagatgatttaaacctgtcaggaaacagttgaaatagtttattgactataatgtgtatacttctatagctatagccttttaccgcallagctttcc

Gene mod 2(I) SUITE ET FIN.

EXON 4 (JUNCTION AA# 1000/1014 +)

3398 *
 GGCCCTCCTGGCCCTCCTGGTCCCCCTGGCCCCCTGGTCCAATGGTGGCGGATATGAAGTTGGCTTTGATGCAGAATACTACCGGGCTGATCAGCCTTCTCTCAGACCCAAGGATTAT
 GlyProProGlyProProGlyProProGlyProProGlyProProGlyProAsnGlyGlyGlyTyrGluValGlyPheAspAlaGluTyrTyrArgAlaAspGlnProSerLeuArgProLysAspTyr

3518 *
 GAAGTTGATGCCACTGAAAAATTGAAACAACCAAAATTGAGACCTGCTGAUCCCAAGGCTCCAAAAAGAACCGCTCGCACCTGCGTGCACCTCAGACTTAGCCACCAGAAATGG
 GluValAspAlaThrLeuLysThrLeuAsnAsnGlnIleGluThrIleLeuThrProGluGlySerLysLysAsnProAlaArgThrCysArgAspLeuArgLeuSerHisProGluTyr

3638 *
 AGCAGCG
 SerSerGly 1 gcaagtcctgcccagctgtttccctctttctggctcagtcagtcattttca
 (-0.3 KB)

1
 gatattttatgaactcgtgagtttctatatcacaattcttgaatataactgaacgcagtgctttctttccctcaag

EXON 3

3645 *
 GTTCTACTGGATTGATCCCAACCAAGGCTGCCTG CAGATGCCATTAGAGCTACTGTGACTTTGCTACTGGTGAGACTTGATCCATGCTAGCCCTTGAAGATATTCGGACTAAGACCT
 lyPheTyrTrpIleAspProAsnGluGlyCysThrAlaAspAlaIleArgAlaTyrCysAspPheAlaThrGlyGluThrCysIleHisAlaSerLeuGluAspIleProThrLysThrT

3765 *
 GGTATGTCAGCAAGAACCCCAAGGACAAAAAGCACATATGGTTGCGTGAACCTATCAATGGTGGTACTCAG
 rpTyrValSerLysAsnProLysAspLysLysHisIleTrpPheGlyGluThrIleAsnGlyGlyThrGln

1
 gatatgtatgcatggaggatgatggttccctcaggtgctttcttaagaatttgcctaccgctgactctcagtgagcgaactcgttttatataactctttagaggagaaatatacaaggaggga
 121
 gagc (-1.0 KB) 1 tctttttcag

EXON 2

3836 *
 TTTGAATACAATGGTGAAGGTGTGACCACAAAGGACATGGCCACCCCACTGCTTTTCATGCGTCTGCTGGCCACCATGCTCTCAGAACATCACCTACCACTGCAAGAACAGCATTGCC
 PheGluTyrAsnGlyGluGlyValThrThrLysAspMetAlaThrGlnLeuAlaPheMetArgLeuLeuAlaAsnHisAlaSerGlnAsnIleThrTyrHisCysLysAsnSerIleAla

3956 *
 TACATGGATGAGSAGACTGGAAACCTTAAAAAGGCTGTATATACTCCAGGGATCCAATGATGTTGAACCTACGAGCTGAAGGCCAACAGCAGATTCACTTTCAGTGTCTCTGGATGGCTGC
 TyrMetAspGluGluThrGlyAsnLeuLysLysAlaValIleLeuGlnGlySerAsnAspValGluLeuArgAlaGluGlyAsnSerArgPheThrPheSerValLeuValAspGlyCys

4076 *
 TCT
 Sor 1 gtaagttaacacgcaaacgcataaggcagctatggttggatattatcacaagttctactttgattcagaactagattatckggcagaccacaaacattttatccag
 101
 :gaggataaantagcctgt (-0.32 KB) 1 tgaaccattctcagttctatgagcaatacnaatacaantctataactgctctctcttgcag

EXON 1

4079 *
 AAAAAGAACAACAAATGGGGCAAAAGCATCAITGAGTACAGAACAAATAAGCCRTCTCRCTGCCCATCCTTGA.CATTGCACTTTGGACATTGGTGGCGCTGACCAAGAAATCGGTTTG
 LysLysAsnAsuLysTrpGlyLysThrIleIleGluTyrArgIhrAsnLysProSerArgLeuProLleLeuAspIleAlaProLeuAspIleGlyGlyAlaAspGlnGluPheGlyLeu

4199 *
 CACATTGGCCCAGTCTGTTCAAATgeatgaactaannattaacttaaggccccccctcagaattattctttgtcatttcttttttgtautgagagctgactccttccatttttttctg
 HisIleGlyProValCysPheLys

4319 *
 ttcactcacttgcttaactctggcgaagagagaggagaaattgattgagcattgcaatgaatttaatacagcccccaaggacttggaagtctttcaagatttaacaccttg

4439 *
 ctttgggaattgtcaacctttgaangaaaaaaaccccaaaaaaaaccccaaaaaaaatacaaatcaaaatgatgaagtttg

* Région codant pour les zones de jonction intron / exon

On a pu établir que presque tous les introns commencent par GT et se terminent par AG. Ici pour le Co, tous les exons commencent bien par G qui codent pour la triple hélice, et des 10 restants, 4 commencent par A et un cinquième par T. Par contre seulement 12 des exons se terminent par G.

* Les introns

Sur les 38,5 kb de ce gène, 34 kb sont des introns ! Ils sont en majorité dans la moitié 5' du gène. Ces introns contiennent 35 % de GC en moyenne.

* Comparaison avec d'autres gènes:

-Taille des exons

La taille des exons dans la région en triple hélice est relativement conservée, par contre elle varie dans la région C-terminale.

-Séquence nucléotidique

Les 2/3 des AA de la région en triple hélice sont de la Glycine, de la proline ou de l'alanine, ils sont codés par GGX, CCX, GCX, ce qui donne 67 % de GC pour le gène.

C'est dans la région C-terminale qu'elle est le mieux conservée et en particulier, on rencontre dans l'exon 2, la plus longue séquence nucléotidique conservée (56 paires de bases). Il se peut que cette région joue un rôle dans l'alignement des trois chaînes en triple hélice.

La structure détaillée des différents gènes codant pour les autres types de collagène se trouve décrite dans les articles donnés en référence, et la structure décrite ci-dessus est illustrée par les schémas joints.

C / EVOLUTION : QUELQUES HYPOTHESES ...

A partir des connaissances sur la structure et la localisation de ces gènes, certaines hypothèses concernant leur évolution ont été émises:

Nous avons vu que plus de la moitié des exons ont une

taille de 54 bp. Ceci a donné à penser que le gène ancestral avait cette longueur de 54 bp, et que les gènes actuels seraient apparus par des recombinaisons multiples à partir de ce gène ancestral.

Nous avons vu aussi, à partir des études comparatives entre les différents gènes appartenant à la famille des Co interstiels (types I, II, III) dans différentes espèces, que la distribution et la taille des exons est bien conservée. Ceci a donné à penser que les gènes de cette famille de Co ont une origine commune, et seraient apparus par duplications successives dans l'ordre suivant: pro-alpha1(III), pro-alpha2(I), pro-alpha1(II) et pro-alpha1(I). En effet, c'est le gène du pro-alpha1(III) qui a le moins d'homologie avec celui du pro-alpha1(I) , alors que c'est le gène du pro-alpha1(II) qui en est le plus proche. La taille des exons aurait donc été conservée depuis la divergence, il y a de cela plus de 600 millions d'années. Ceci a servi de base pour définir le groupe I des Co et l'élargir récemment aux types V et K .

III - LES MALADIES HEREDITAIRES DES COLLAGÈNES

Ces maladies sont toutes transmises par des gènes. Des troubles au cours des étapes de la biosynthèse du Co ou des déficits en enzymes sont aussi à la source de certaines maladies, mais ne seront pas étudiés ici.

Ces mutations restent assez rares.

A / LE SYNDROME D'EHLERS-DANLOS

Ce syndrome est connu depuis très longtemps car ses symptômes ont suscité la curiosité et ont souvent été utilisés comme phénomènes de cirque. Sept types de S.E.D. sont actuellement connus, correspondant à des défauts enzymatiques ou moléculaires différents. Ainsi l'intérêt suscité par cette maladie s'est-il déplacé de l'échelon clinique à l'échelon moléculaire.

1) EXPRESSION CLINIQUE

Les 4 symptômes caractéristiques sont:

- L'hyperélasticité cutanée
- La fragilité cutanée
- L'hyperlaxité articulaire

- La fragilité vasculaire

Il s'y ajoute des atteintes viscérales multiples qui s'expliquent par l'ubiquité du tissu conjonctif dans l'organisme. La fragilité vasculaire est surtout redoutée quand elle atteint les gros vaisseaux comme dans le type IV, on rencontre aussi des manifestations oculaires, digestives, cardiaques (prolapsus mitral), pulmonaires, génitales.

C'est la biopsie cutanée qui permet de confirmer le diagnostic en montrant l'hypoplasie du tissu collagène.

Le S.E.D. s'exprime selon des degrés de gravité très différents, ce qui fait que parfois diagnostiqué dès la naissance il peut aussi passer inaperçu et cette affection est donc en réalité plus fréquente qu'on ne le croit.

2) LES ANOMALIES RESPONSABLES

Parmi les types de S.E.D. pour lesquels le défaut biochimique a pu être mis en évidence, ceux qui sont d'origine génétique sont les suivants:

a) S.E.D. type IV

C'est la forme la plus sévère de S.E.D.

On y a démontré la diminution du collagène de type III, ce défaut pouvant résulter de plusieurs mécanismes:

- Défaut de synthèse, par délétion d'un gène codant pour le pro Co type III

- Défaut de sécrétion: à cause d'une modification de la structure des chaînes pro- α_1 (III) qui empêcherait la formation de la triple hélice.

- Dégradation trop rapide: certaines chaînes auraient des AA supplémentaires, rendant le pro Co plus sensible aux protéinases.

b) S.E.D. type VIIB

La transmission est récessive autosomique.

Ce type est caractérisé par un défaut de clivage du pro-peptide N-terminal du pro Co (Rappelons que lors de la biosynthèse du Co, les chaînes alpha sont d'abord synthétisées sous forme de très

longues chaînes pro-alpha. Elles perdent d'abord leur peptide-signal, puis sont assemblées en triple hélice après hydroxylation et glycosylation de certains AA. Après la sécrétion, les propeptides amino- et carboxyl-terminal sont coupés par l'action de protéinases spécifiques) C'est une altération structurale de la chaîne pro-alpha qui empêche ce clivage du propeptide N-terminal par l'enzyme " procollagène N-protéinase ": on a trouvé une délétion de 18 résidus dans la moitié des chaînes pro-alpha2(I) du Co type I, dans la région du N-telopeptide; elle supprime donc le site normal de clivage par la N-protéinase ainsi qu'un résidu de lysine important pour le " cross-linking ".

Ces 18 AA correspondent à la zone codée par l'exon 6 du gène.

On a trouvé aussi une délétion de 24 AA dans la chaîne pro-alpha1(I) du Co type I. Ils correspondraient à la zone codée par l'exon 46.

B / L'OSTEOGENESE IMPARFAITE OI

1) EXPRESSION CLINIQUE

L'incidence de l'OI est de 1/20 000. Ces principaux symptômes sont :

- fragilité osseuse avec fractures multiples
- sclérotiques bleues
- surdit 

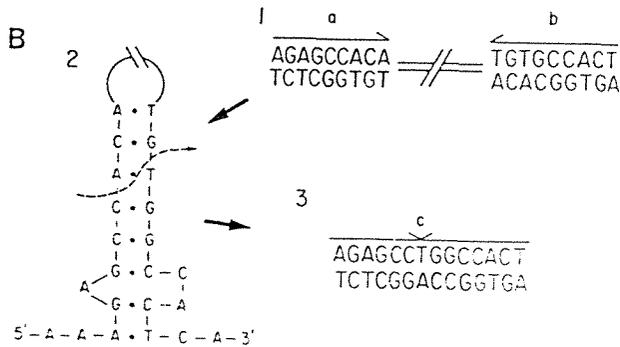
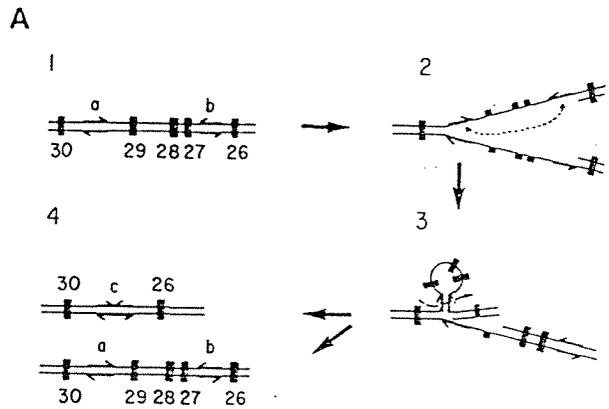
Cette affection est tr s h t rog ne (voir tableau pour classification)

a) OI type II

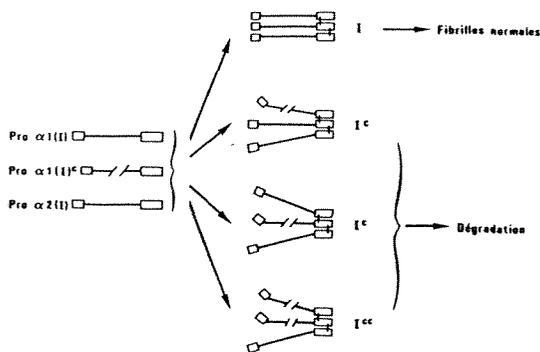
Il y a une d l tion de 500 bases d'un des all les du g ne codant pour la cha ne pro-alpha1(I),  liminant les exons 27-28-29, les points de rupture pourraient avoir  t  favoris s par la formation de structures en  pingle   cheveux.

La moiti  des cha nes pro-alpha1(I) synth tis es sont raccourcies de 80 AA, l'autre moiti   tant normale. Ces deux types de cha nes  tant incorpor es   la triple h lice, un quart seulement des mol cules de collag ne form es sont normales, les autres comportant

[39] Oi : Mécanisme possible d'une délétion favorisée par la formation d'une "épingle à cheveux" dans l'O.i. type II



"Protéine-suicide" dans l'O.i. type IV



1 ou 2 chaînes pro-alpha1(I) anormales, sont instables et se dégradent rapidement avant de pouvoir former des fibrilles de Co. Ce phénomène a été surnommé phénomène de " protéine-suicide ".

On a trouvé d'autres anomalies:

- Substitution de la glycine en position 391 de la chaîne alpha1(I) par de l'arginine.
- Substitution d'une glycine en position 988 de la chaîne alpha1(I) par de la cystéine

b) OI type I

Diminution de synthèse de la chaîne pro-alpha1(I), peut-être due à une substitution d'un AA en position X ou Y par une cystéine.

c) OI type III

Il existe une délétion de 4 paires de bases de l'exon 1.

Ceci a pour résultat une mutation au niveau du propeptide C-terminal de la chaîne pro-alpha2(I) d'où synthèse de trimères pro-alpha1(I), car les chaînes pro-alpha2(I) ne peuvent pas s'intégrer dans la triple hélice.

d) OI type IV

La chaîne pro-alpha2(I) est trop courte par délétion d'une dizaine de résidus, ce qui donne des fibrilles de Co anormales, de diamètre réduit. cette délétion est à l'extrémité N-terminale.

Substitution de la dernière glycine de la région en triple hélice de la chaîne pro-alpha2(I) par une arginine.

C / LE SYNDROME DE MARFAN

Le syndrome de Marfan comporte une atteinte des divers éléments du tissu conjonctif : collagène, mais aussi élastine et fibroblastes.

En ce qui concerne le collagène, on a pu montrer la présence de 2 types de chaînes pro- $\alpha_2(I)$, l'une normale, l'autre ayant subi une insertion d'un peptide supplémentaire d'une vingtaine d'AA dans sa partie C-terminale. Cette anomalie conduit à la synthèse d'un collagène trop soluble.

CONCLUSION

Nous avons que les anomalies génétiques pouvant être à l'origine de maladies du collagène sont très nombreuses et variées, mais sans doute sont elles bien plus nombreuses encore à n'avoir pas été identifiées.

Les progrès actuels en ce domaine sont rapides.

Pour l'instant, ces connaissances n'ont pas encore de conséquences thérapeutiques appréciables, mais l'identification de ces anomalies pourrait dans l'avenir permettre un diagnostic prénatal plus fréquent et le conseil génétique.



BIBLIOGRAPHIE

- (1) BOYD C.D., WELIKY K., TOTH-FEJEL S., et al.
The single copy gene coding for human alpha1(IV) procollagen is located at the terminal end of the long arm of the chromosome 13.
Hum. Genet., 1986, vol.74, pp:121-125.
- (2) EMANUEL B.S., CANNIZARO L.A., SEYER J.M., et al.
Human alpha1(III) and alpha2(V) procollagen genes are located on the long arm of chromosome 2.
Proc. Natl. Acad. Sci. U.S.A., 1985, vol.82, pp:3385-3389.
- (3) GRIFFIN C.A., EMANUEL B.S., HANSEN J.R., et al.
Human collagen genes encoding basement membrane alpha1(IV) and alpha2(IV) chains map to the distal long arm of chromosome 13.
Proc. Natl. Acad. Sci. U.S.A., 1987, vol.84, pp:512-516.
- (4) HUERRE-JEANPIERRE C., MATTEI M.G., WEIL D., et al.
Further evidence for the dispersion of the human fibrillar collagen genes.
Am. J. Hum. Genet., 1986, vol.38, pp:26-27.
- (5) HUERRE-JEANPIERRE C., HENRY I., BERNARD M., et al.
The pro-alpha2(V) collagen gene (COL5A2) map to 2q14-2q32 syntenic to the pro-alpha1(III) collagen locus (COL3A1).
Hum. Genet., 1986, vol.73, pp:64-67.
- (6) LAW M.L., TUNG L., MORSE H.G., et al.
The human type II collagen gene (COL2A1) assigned to 12q14.3
Ann. Hum. Genet., 1986, vol.50, pp:131-137.
- (7) MYERS J.C., EMANUEL B.S.
Chromosomal localization of human collagen genes.
Col. Relat. Res., 1987 06, vol.7(2), pp:149-159.
- (8) POSCHL E., POLLNER R., KUHN K.
The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure.
EMBO J., 1988 09, vol.7(9), pp:2687-2695.
- (9) RETIEF E., PARKER M.I., RETIEF A.E.
Regional chromosome mapping of human collagen genes alpha2(I) and alpha1(I) (COL1A2 and COL1A1).
Hum. Genet., 1985, vol.69, pp:304-308.
- (10) SOLOMON E., HIORNS L.R., SPURR N., et al.
Chromosomal assignments of the genes coding for human types II, III, and IV collagen: a dispersed gene family.
Proc. Natl. Acad. Sci. U.S.A., 1985, vol.82(10), pp:3330-3334.

LES GENES, structure.généralités

- (11) BOEDTKER H., AHO S.
Collagen gene structure: the paradox may be resolved.
Biochem. Soc. Symp., 1984, vol.49, pp:67-84.
- (12) BOEDTKER H., FULLER F., TATE V.
The structure of collagen genes.
Int. Rev. Connect. Tissue. Res., 1983, vol.10, pp:1-63.
- (13) DE CROMBRUGGHE B., LIAU G., SETOYAMA C., et al.
Structural and functional studies on the interstitial collagen genes.
Ciba Found. Symp., 1985, vol.114, pp:20-33.
- (14) MILLER E.J., GAY S.
The collagens: an overview and update.
Methods. Enzymol., 1987, vol.144, pp:3-41.
- (15) RAMIREZ F., BERNARD M., CHU M.-L., et al.
Isolation and characterization of the human fibrillar collagen genes.
Ann. N. Y. Acad. Sci., 1985, vol.460, pp:117-129.

TYPE ALPHA1(I)

- (16) BARSH G.S., ROUSH C.L., GELINAS R.E.
DNA and chromatin structure of the human alpha1(I) collagen gene.
J. Biol. Chem., 1984 12 10, vol.259(23), pp:14906-14913.
- (17) CHU M.-L., DE WET W., BERNARD M., et al.
Fine structural analysis of the human pro-alpha1(I) collagen gene: promoter structure, AluI repeats, and polymorphic transcripts.
J. Biol. Chem., 1985, vol.260(4), pp:2315-2320.
- (18) D'ALESSIO M., BERNARD M., PRETORIUS P.J., et al.
Complete nucleotide sequence of the region encompassing the first twenty-five exons of the human pro-alpha1(I) collagen gene (COL1A1).
Gene, 1988 07 15, vol.67(1), pp:105-115.

TYPE ALPHA2(I)

- (19) DE CROMBRUGGHE B., SCHMIDT A., LIAU G., et al.
Structural and functional analysis of the genes for alpha²(I) and alpha1(III) collagens.
Ann. N. Y. Acad. Sci., 1985, pp:154-162.

- (20) KUHN K.
Structural and functional domains of collagen: a comparison of the protein with its gene.
Coll. Relat. Res., 1984 08, vol.4(4), pp:309-322.
- (21) KUIVANIEMI H., TROMP G., CHU M.-L., et al.
Structure of a full-length cDNA clone for the prepro-alpha2(I) chain of human type I procollagen. Comparison with the chicken gene confirms unusual patterns of gene conservation.
Biochem. J., 1988 06 15, vol.252(3), pp:633-640.

TYPE ALPHA1(II)

- (22) CHEAH K.S.E., STOKER N.G., GRIFFIN J.R., et al.
Identification and characterization of the human type II collagen gene (COL2A1).
Proc. Natl. Acad. Sci. U.S.A., 1985, vol.82(9), pp:2555-2559.
- (23) LOVELL-BADGE R.H., BYGRAVE A., BRADLEY A., et al.
Tissue specific expression of the human type II collagen gene in mice.
Proc. Natl. Acad. Sci. U.S.A., 1987, vol.84, pp:2803-2807.
- (24) NUNEZ A.M., KOHNO K., MARTIN G.R., et al.
Promoter region of the human pro-alpha1(II) collagen gene.
Gene, 1986, vol.44(1), pp:11-16.

TYPE ALPHA1(III)

- (25) CHU M.-L., WEIL D., DE WET W., et al.
Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen: partial characterization of the 3'end region of the gene.
J. Biol. Chem., 1985, vol.260(7), pp:4357-4363.
- (26) MISKULIN M., DALGLEISH R., KLUVE-BECKERMAN B., et al.
Human type III collagen gene expression is coordinately modulated with the type I collagen genes during fibroblast growth.
Biochemistry, 1986 03 25, vol.25(6), pp:1408-1413.
- (27) TOMAN P.D., RICCA G.A., DE CROMBRUGGHE B.
Nucleotide sequence of a cDNA coding for the amino-terminal region of human prepro-alpha1(III) collagen.
Nucleic Acids Res., 1988 07 25, vol.16(14b), p:7201.

TYPES ALPHA1(IV) ET ALPHA2(IV)

- (28) KILLEN P.D., BURBELO P.D., MARTIN G.R., et al.
Characterization of the promoter for the alpha1(IV) collagen gene. DNA sequences within the first intron enhance transcription.
J. Biol. Chem., 1988 09 05, vol.263(25), pp:12310-12314.

- (29) KURKINEN M., BARLOW D.P., HELFMAN D.M., et al.
Isolation and cDNA clones for basal lamina components: typeIV
procollagen.
Nucleic Acids Res., 1983, vol.11, pp:6199-6209.
- (30) PIHLAJANIEMI T., TRYGGVASON K., MYERS J.C., et al.
cDNA clones coding for the pro-alpha(IV) chain of human
type IV procollagen reveal an unusual homology of amino-
acid sequences in two halves of the carboxyl-terminal domain.
J. Biol. Chem., 1985, vol.260, pp:7681-7687.

TYPE ALPHA2(V)

- (31) SCHWARTZ R.C., LIDELL A., RAMIREZ F., et al.
An RFLP in the gene for the human pro-alpha2 chain of typeV
collagen (COL5A2).
Nucleic Acids Res., 1988 06 10, vol.16(11), pp:5225.
- (32) WEIL W., BERNARD M., GARGANO S., et al.
The pro-alpha2(V) collagen gene is evolutionarily related
to the major fibrillar-forming collagens.
Nucleic Acids Res., 1987 01 12, vol.15(1), pp:181-198.

TYPE IX

- (33) LOZANO G., NINOMIYA Y., THOMPSON H., et al.
A distinct class of vertebrate collagen genes encodes chicken
type IX polypeptides.
Proc. Natl. Acad. Sci. U.S.A., 1985, vol.82, pp.4050-4054.
- (34) NINOMIYA Y., OLSEN B.R.
Synthesis and characterization of cDNA encoding a cartilage
specific short collagen.
Proc. Natl. Acad. Sci. U.S.A., 1984, vol.81, pp:3014-3018.

TYPE XII

- (35) MARION K.G., GERECKE D.R., OLSEN B.R., et al.
Type XII collagen, distinct extracellular matrix component
discovered by cDNA cloning.
Proc. Natl. Acad. Sci. U.S.A., 1987 09, vol.84, pp:6040-
6044.

TYPE XIII

- (36) TIKKA L., PIHLAJANIEMI T., HENTTU P., et al.
Gene structure for the alpha1 chain of a human short-chain
collagen type XIII with alternatively spliced transcripts
and translation termination codon at the 5'end of the last
exon.
Proc. Natl. Acad. Sci. U.S.A., 1988, vol.85(15), pp:7491-
7495.

LES GENES, anomalies.GENERALITES

- (37) BYERS P.H., WENSTRUP R.J., BONADIO J.F., et al.
Molecular basis of inherited disorders of collagen biosynthesis: implications for prenatal diagnosis.
Curr. Probl. Dermatol., 1987, vol.16, pp:158-174.
- (38) PROCKOP D.J.
Mutations in collagen genes: consequences for rare and common diseases.
J. Clin. Invest., 1985, vol.75(3), pp:783-787.
- (39) RAMIREZ F.
Human collagens: biochemical, molecular, and genetics features in normal and diseased states.
Horiz. Biochem. Biophys., 1986, vol.8, pp:341-375.
- (40) SYKES B., SMITH R.
Collagen and collagen gene disorders.
Q. J. Med., 1985, vol.56(221), pp:533-548.
- (41) TSIPOURAS P., RAMIREZ F.
Genetics disorders of collagen.
J. Med. Genet., 1987, vol.24(1), pp:2-8.
- (42) UITTO.
UCLA conference. Biochemistry of collagen in diseases.
Ann. Intern. Med., 1986 11, vol.105(5), pp:740-756.
- (43) VUORIO E.
Connective tissue diseases: mutations of collagen genes.
Ann. Clin. Res., 1986, vol.18(5-6), pp:234-241.

SYNDROME D'EHLERS DANLOS

- (44) AUMAILLEY M., POSCHL E., MARTIN G.R., et al.
Low product of procollagen III by skin fibroblasts from patients with Ehlers-Danlos syndrome type IV is not caused by decreased levels of procollagen III mRNA.
Eur. J. Clin. Invest., 1988 04, vol.18(2), pp:207-212.
- (45) COLE W.G., CHAN D., CHAMBERS G.W., et al.
Deletion of 24 amino-acids from the pro-alpha1(I) chain of type I procollagen in a patient with the Ehlers-Danlos syndrome type VII.
J. Biol. Chem., 1986, vol.261(12), pp:5496-5550.
- (46) EYRE D., SHAPIRO F.D., ALDRIDGE J.F.
A heterozygous collagen defect in a variant of the Ehlers-Danlos syndrome type VII evidence for a deleted amino telopeptide domain in the pro-alpha2(I) chain.
J. Biol. Chem., 1985, vol.260(20), pp:11322-11329.

- (47) HATA R., KURATA S., SHINKAI H.
Existence of malfunctioning pro-alpha2(I) collagen genes in a patient with a pro-alpha2(I)-chain-defective variant of Ehlers-Danlos syndrome.
Eur. J. Biochem., 1988 06 01, vol.174(2), pp:231-237.
- (48) SULH H.M., STEINMANN B., RAO V.H., et al.
Ehlers-Danlos syndrome type IV D: an autosomal recessive disorder.
Clin. Genet., 1984 03, vol.25(3), pp:278-287.
- (49) TSIPOURAS P., BYERS P.H., SCHWARTZ R.C., et al.
Ehlers-Danlos type IV: cosegregation of the phenotype to a COL3A1 allele of type III procollagen.
Hum. Genet., 1986, vol.74(1), pp:41-46.
- (50) WEIL D., BERNARD M., COMBATES M., et al.
Identification of a mutation that causes exon skipping during collagen pre-messenger RNA splicing in an Ehlers-Danlos syndrome variant.
J. Biol. Chem., 1987, vol.263(18), pp:8561-8564.
- (51) WIRTZ M.K., GLANVILLE R.W., STEINMANN B., et al.
Ehlers-Danlos syndrome type VIIB: deletion of 18 amino acids comprising the N-telopeptide region of a pro-alpha2(I) chain.
J. Biol. Chem., 1987, vol.262(34), pp:16376-16385.

SYNDROME DE MARFAN

- (52) BYERS P.H., SIEGEL R.C., PETERSON K.E., et al.
Marfan syndrome: abnormal alpha2 chain in type I collagen.
Proc. Natl. Acad. Sci. U.S.A., 1981 12, vol.78(12), pp:7745-7749.

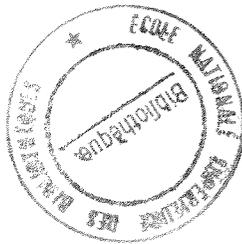
OSTEOGENESIS IMPERFECTA

- (53) AITCHISON K., OGILVIE D., HONEYMAN M., et al.
Homozygous Osteogenesis imperfecta unlinked to collagen I genes.
Hum. Genet., 1988, vol.78(3), pp:233-236.
- (54) BARSH G.S., ROUSH C.L., BONADIO J., et al.
Intron-mediated recombination may cause a deletion in an alpha I type collagen chain in a lethal form of Osteogenesis imperfecta.
Proc. Natl. Acad. Sci. U.S.A., 1985, vol.82(9), pp:2870-2874.
- (55) BATEMAN J.F., CHAN D., WALKER I.D., et al.
Lethal perinatal Osteogenesis imperfecta due to the substitution of arginine for glycine at residue 391 of the alpha1(I) chain of type I collagen.
J. Biol. Chem., 1987, vol.262(15), pp:7021-7027.

- (56) BYERS P.H., COHN D.H., STARMAN B.J., et al.
Molecular basis of Osteogenesis imperfecta OI.
J. Cell. Biochem., 1988, suppl.0(12b), p:165.
- (57) BYERS P.H., STARMAN B.J., COHN D.H., et al.
A novel mutation causes a perinatal lethal form of Osteogenesis imperfecta. An insertion in one alpha(I) collagen allele (COL1A1).
J. Biol. Chem., 1988 06 05, vol.263(16), pp:7855-7861.
- (58) BYERS P.H., TSIPOURAS P., BONADIO J.F., et al.
Perinatal lethal Osteogenesis imperfecta OI type II: a biochemically heterogeneous disorder usually due to new mutations in the genes for type I collagen.
Am. J. Hum. Genet., 1988, vol.42(2), pp:237-248.
- (59) CHU M.-L., GARGIULO V., WILLIAMS C.J., et al.
Multiexon deletion in an Osteogenesis imperfecta variant with increased type III collagen mRNA.
J. Biol. Chem., 1985, vol.260(2), pp:691-694.
- (60) COHN D.H., BYERS P.H., STEINMANN B., et al.
Lethal Osteogenesis imperfecta resulting from a single nucleotide change in one human pro-alpha(I) collagen allele.
Proc. Natl. Acad. Sci. U.S.A., 1986 08, vol.83(16), pp:6045-6047.
- (61) DICKSON L.A., PIHLAJANIEMI T., DEAK S., et al.
Nuclease S1 mapping of a homozygous mutation in the carboxyl-propeptide-coding region of the pro-alpha2(I) collagen gene in a patient with Osteogenesis imperfecta.
Proc. Natl. Acad. Sci. U.S.A., 1984, vol.81(14), pp:4524-4528.
- (62) DOERING J.L., BURKET A.E., VOGEL L.C.
Collagen gene deletions in Osteogenesis imperfecta patients.
J. Cell. Biol., 1987, vol.105(4), p:213A.
- (63) FALK C.T., SCHWARTZ R.C., RAMIREZ F., et al.
Use of molecular haplotypes specific for the human pro-alpha2(I) collagen gene in linkage analysis of the mild autosomal dominant forms of Osteogenesis imperfecta.
Am. J. Hum. Genet., 1986 03, vol.38(3), pp:269-279.
- (64) FROST H.M.
OI: the set point proposal (a possible causative mechanism).
Clin. Orthop. Relat. Res., 1987 03, vol.216, pp:280-297.
- (65) GROBLER-RABIE A.F., WALLIS G., BREBNER D.K., et al.
Detection of a high frequency RsaI polymorphism in the human pro-alpha2(I) collagen gene which is linked to an autosomal dominant form of Osteogenesis imperfecta.
EMBO J., 1985 07, vol.4(7), pp:1745-1748.

- (66) KNISELY A.S., RICHARDSON A., ABUELO D., et al.
Lethal Osteogenesis imperfecta unlinked to collagen I genes.
Hum. Genet., 1988, vol.78(3), pp:233-236.
- (67) KUIVANIEMI H., SABOL C., TROMP G., et al;
A 19-base pair deletion in the pro-alpha2(I) gene of type
I procollagen that causes in-frame RNA splicing from exon
10 to exon 12 in a proband with atypical OI and his asymptomatic mother.
J. Biol. Chem., 1988 08 15, vol.263(23), pp:11407-11413.
- (68) PIHLAJANIEMI T., MYERS J.C.
Characterization of a pro-alpha2(I) collagen gene mutation
by nuclease S1 mapping.
Methods Enzymol., 1987, vol.145, pp:213-222.
- (69) PIHLAJANIEMI T., DICKSON L.A., POPE F.M.
Osteogenesis imperfecta: cloning of a pro-alpha2(I) collagen
gene with a frameshift mutation.
J. Biol. Chem., 1984 11 10, vol.259(21), pp:12941-12944.
- (70) POPE F.M., NICHOLLS A.C., MCPHEAT J., et al.
Collagen genes and proteins in Osteogenesis imperfecta.
J. Med. Genet., 1985, vol.22(6), pp:466-478.
- (71) SYKES B., OGILVIE D., WORDSWORTH P., et al.
Osteogenesis imperfecta is linked to both type I collagen
structural genes.
LANCET, 1986 07 12, vol.2(8498), pp:69-72.
- (72) TROMP G., PROCKOP D.J.
Single base mutation in the pro-alpha2(I) collagen gene that
causes efficient splicing of RNA from exon 27 to exon 29
and synthesis of a shortened but in-frame pro-alpha2(I) chain.
Proc. Natl. Acad. Sci. U.S.A., 1988, vol.85(14), pp:5224-
5258.
- (73) TSIPOURAS P., SCHWARTZ R.C., GOLDBERG J.D., et al.
Prenatal prediction of Osteogenesis imperfecta (OI type IV)
exclusion of inheritance using a collagen gene probe.
J. Med. Genet., 1987 07, vol.24(7), pp:406-409.
- (74) TSIPOURAS P., SCHWARTZ R.C.
Dominantly inherited Osteogenesis imperfecta OI types I and
IV is genetically linked to the COL1A2 genes of type I col-
lagen.
Pediatr. Res., 1987, vol.21(4), p:294A.
- (75) TSIPOURAS P., MYERS J.C., RAMIREZ F., et al.
RFLP associated with the pro-alpha1 gene of human type I
procollagen: application to a family with an autosomal do-
minant form of Osteogenesis imperfecta.
J. Clin. Invest., 1983, vol.72(4), pp:1262-1267.

- (76) WENSTRUP R.J., COHN D.H., COHEN T., et al.
Arginine for glycine substitution in the triple-helical domain of the products of one alpha2(I) collagen allele (COL1A2) produces the Osteogenesis imperfecta type IV phenotype.
J. Biol. Chem., 1988 06 05, vol.263(16), pp:7734-7740.
- (77) WENSTRUP R.J., TSIPOURAS P., BYERS P.H.
Osteogenesis imperfecta type IV: biochemical confirmation of genetic linkage to the pro-alpha2(I) gene of type I collagen.
J. Clin. Invest., 1986, vol.78(6), pp:1449-1455.
- (78) WILLING M.C., COHN D.H., STARMAN B., et al.
Heterozygosity for a large deletion in the alpha2(I) collagen gene has a dramatic effect on type I collagen secretion and produces perinatal lethal Osteogenesis imperfecta.
J. Biol. Chem., 1988, vol.263(17), pp:8398-8404.



BIBLIOTHEQUE DE L'ENSSIB



966040F