

Modules et interactions moléculaires au sein des matrices extracellulaires

Claire Giry-Lozinguez
Jean-Philippe Kleman
Michel van der Rest

De nombreuses protéines comportent des motifs homologues, appelés modules. Les macromolécules sécrétées, malgré leur grande diversité, sont construites à partir d'un nombre relativement restreint de modules différents. La modularité des protéines de la matrice extracellulaire est importante, tant par les assemblages supramoléculaires que par son rôle physiologique. Les modules peuvent, en effet, être impliqués à la fois dans la structure de ces protéines et dans les interactions qu'elles ménagent entre elles ou avec des récepteurs cellulaires qui leur sont propres, ce qui contribue à la cohérence des matrices extracellulaires. Les mutations affectant la structure d'un module et/ou la fonction qui peut lui être associée sont à l'origine de plusieurs maladies héréditaires des matrices extracellulaires.

La matrice extracellulaire comprend un grand nombre d'assemblages supramoléculaires différents, spécifiques de tissus donnés. Ces assemblages complexes se construisent et se spécialisent au cours du développement, et leur composition peut évoluer en fonction de différentes situations, physiologiques (croissance ou remodelage du tissu lors de la cicatrisation par exemple), ou pathologiques (fibrose, ostéoarthrose, etc.). La grande diversité des matrices extracellulaires est liée au grand nombre de molécules qui la composent et à leurs multiples capacités d'interaction. A l'origine de cette diversité moléculaire se trouve un nombre limité d'unités homologues constitutives : les modules. Les modules re-

présentent des séquences homologues rencontrées plusieurs fois à l'intérieur d'une même protéine ou dans plusieurs protéines différentes, sécrétées ou non.

Ces modules diffèrent cependant les uns des autres à l'intérieur d'une protéine par le contexte dans lequel ils sont présents, et parfois par les fonctions qu'ils prodiguent. Ils sont souvent disposés en groupes de plusieurs modules adjacents (*clusters*), formant ainsi des domaines entiers de la protéine. On retrouve cependant aussi des domaines composés d'un module unique ou de plusieurs modules différents adjacents.

Ces modules représentent généralement une capacité de repliement indépendant du reste de la protéine, ce repliement étant supposé conservé

ADRESSE

C. Giry-Lozinguez : doctorante, allocataire Moniteur Normalien (université Lyon I). M. Van der Rest : professeur à l'ENS de Lyon, directeur du laboratoire de structure et ingénierie des collagènes de l'IBCP. Institut de biologie et chimie des protéines (IBCP), Cnrs UPR 412, 7, passage du Vercors, 69367 Lyon Cedex 07, France. J.P. Kleman : docteur en biologie. Institut de biologie structurale (IBS) 41, avenue des Martyrs, 38027 Grenoble Cedex 1, France.

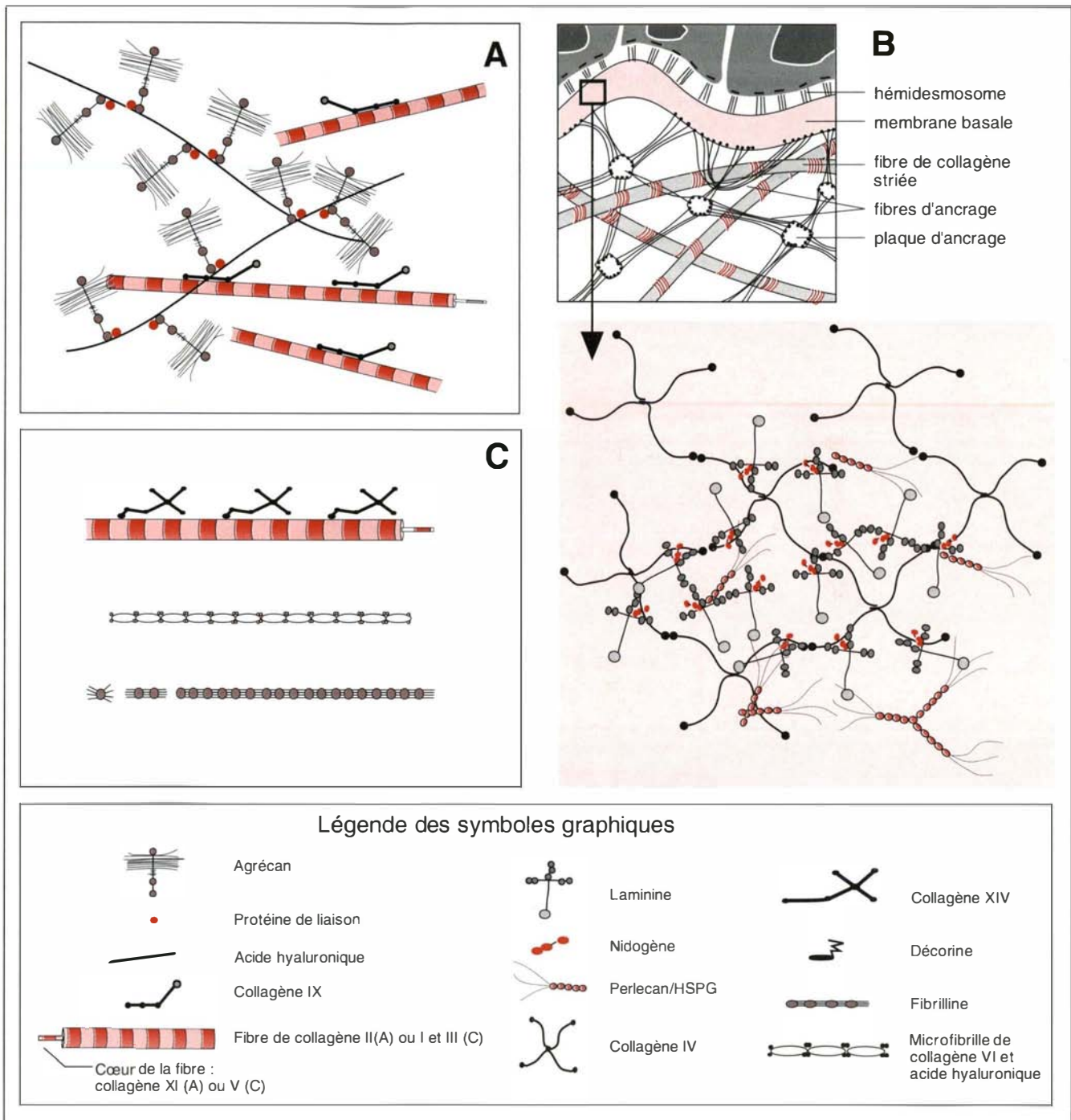


Figure 1. **Exemples de chaînes d'interaction.** Les macromolécules ne sont pas dessinées à l'échelle. **A.** Interactions des composants de la matrice cartilagineuse. **B.** Interactions des composants de la jonction dermo-épidermique. Pour des raisons de simplicité, la complexité du lacs de collagène IV n'a pas été représentée. **C.** Assemblage de macromolécules rencontrées dans le derme et d'autres tissus.

RÉFÉRENCES

1. van der Rest M. Biologie du collagène et maladies héréditaires de la matrice extracellulaire. *médecine/sciences* 1987 ; 3 : 411-20.
2. van der Rest M, Mayne R. Type IX collagen proteoglycan from cartilage is covalently cross-linked to type II collagen. *J Biol Chem* 1988 ; 263 : 1615-8.
3. Vaughan L, Mendler M, Huber S, Bruckner P, Winterhalter HK, Irwin MI, Mayne R. D-periodic distribution of collagen type IX along cartilage fibrils. *J Cell Biol* 1988 ; 106 : 991-7.
4. Mendler M, Eich-Bender SG, Vaughan L, Winterhalter HK. Cartilage contains mixed fibrils of collagen types II, IX, and XI. *J Cell Biol* 1989 ; 108 : 191-7.
5. van der Rest M, Mayne R, Ninomiya Y, Seidah NG, Chrétien M, Olsen BR. The structure of type IX collagen. *J Biol Chem* 1985 ; 260 : 220-5.
6. Ninomiya Y, van der Rest M, Mayne R, Lozano G, Olsen BR. Construction and characterization of cDNA encoding the $\alpha 2$ chain of chicken type IX collagen. *Biochemistry* 1985 ; 24 : 4223-9.
7. Brewton RG, Ouspenskaia MV, van der Rest M, Mayne R. Cloning of the chicken $\alpha 3$ (IX) collagen chain completes the primary structure of type IX collagen. *Eur J Biochem* 1992 ; 205 : 443-9.
8. Zhidkova NI, Brewton RG, Mayne R. Molecular cloning of PARP (proline/arginine rich protein) from human cartilage and subsequent demonstration that PARP is a fragment of the NH₂-terminal domain of the collagen $\alpha 2$ (XI) chain. *FEBS Lett* 1993 ; 326 : 25-8.
9. Nakata K, Ono K, Miyazaki JI, Olsen BR, Muragaki Y, Adachi E, Yamamura KI, Kimura T. Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing $\alpha 1$ (IX) collagen chains with a central deletion. *Proc Natl Acad Sci USA* 1993 ; 90 : 2870-4.
10. Hardingham TE. The role of link protein in the structure of cartilage proteoglycan aggregates. *Biochem J* 1979 ; 177 : 237-47.
11. Doege KJ, Sasaki M, Kimura T, Yamada Y. Complete coding sequence deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. *J Biol Chem* 1991 ; 266 : 894-902.
12. Perkins SJ, Nealis AS, Dudhia J, Hardingham TE. Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal domains and link protein. *J Mol Biol* 1989 ; 206 : 737-53.
13. Meneguzzi G, Aberdam D, Vailly J, Ortonne JP. Vers la compréhension moléculaire des épidermolyses bulleuses héréditaires. *médecine/sciences* 1993 ; 9 : 387-95.
14. Nicolas JF, Michalaki H, Peyron E, Machado P, Cozzani E, Schmitt D. Pathologie acquise de la jonction dermo-épidermique. *médecine/sciences* 1993 ; 9 : 376-86.

d'un module à l'autre, quel que soit le contexte dans lequel il se trouve. On suppose, par exemple, qu'un module EGF (*epidermal growth factor*, ou facteur de croissance épidermique), présent de manière répétée dans diverses protéines de la matrice, aura une structure tridimensionnelle voisine de celle qui est connue pour l'EGF soluble (voir figure 4). On peut souvent faire un parallèle entre la structure modulaire d'une protéine et la structure en exons du gène correspondant. En effet les limites d'un module correspondent souvent aux limites d'un exon. Il existe cependant plusieurs exceptions notoires (modules VWA, homologues du facteur von Willebrand, collagènes, etc.).

Nous allons ici illustrer ces concepts par la description de deux matrices extracellulaires très étudiées actuellement : la matrice cartilagineuse et la jonction dermo-épidermique, en détaillant la structure de leurs composants et les interactions qui les lient. Nous décrirons, de plus, quelques assemblages supramoléculaires retrouvés dans de nombreux tissus. L'utilisation de ces exemples nous conduira à analyser plus en détail certains modules constituant ces protéines, que l'on regroupera ensuite par type de fonctions.

Exemples d'assemblages supramoléculaires

Le cartilage hyalin

Le cartilage hyalin (figure 1A) est un tissu constitué d'une matrice extracellulaire importante riche en collagènes et protéoglycanes, et synthétisée par un nombre relativement petit de cellules, les chondrocytes. Dans ce tissu, la force tensile des fibres de collagène s'oppose à la pression osmotique due à la grande densité de charge des protéoglycanes pour former un système biologique d'amortisseur de chocs. Le composant majeur de cette matrice est le collagène II. Celui-ci est sécrété sous la forme d'homotrimeres (procollagènes), dont les extrémités N et C-terminales sont clivées dans l'espace extracellulaire [1]. Les molécules de collagène mature s'assemblent alors entre elles pour former des fibrilles, dont le cœur est probablement formé de collagène XI. Un col-

lagène non fibrillaire, le collagène IX, s'associe de façon covalente avec les molécules de collagène II de la surface de la fibrille de manière périodique (figure 1A) [1-4]. Ce collagène non fibrillaire appartient à la famille des collagènes associés aux fibrilles et à triple hélice interrompue (FACIT, *fibril associated collagen with interrupted triple helix*). Il est composé de trois domaines en triple hélice (COL1 à COL3) séparant des domaines non collagéniques (NC1 à NC4) [5-7]. Il interagit avec la fibrille par ses domaines COL1 et COL2, C-terminaux. Le domaine COL1 constitue un module retrouvé dans tous les FACIT (voir ci-dessous et figure 3A). Les domaines COL3, hélicoïdal, et NC4, non collagénique et N-terminal, sont projetés à l'extérieur de la fibrille. Ce dernier, constitué en fait d'un module appelé PARP (homologue du *peptide arginine-proline-rich* ou NC4 [8] est ainsi en position favorable pour interagir avec d'autres éléments de la matrice, comme des protéoglycanes (figure 1A).

Nakata *et al.* ont construit un gène recombinant de la chaîne $\alpha 1$ (IX) délété d'une partie des domaines COL3 et COL2. Ce gène a été exprimé chez des souris transgéniques. La chaîne $\alpha 1$ (IX) tronquée qui est synthétisée interfère avec la formation de triples hélices stables. Les souris homozygotes pour le transgène développent une ostéoartrrose accompagnée d'une chondrodysplasie légère [9]. Ces résultats, qui suggèrent que des mutations dans les gènes de collagène de type IX pourraient être responsables de certaines formes d'ostéoartrrose et de chondrodysplasies chez l'homme, soulignent l'importance des interactions moléculaires du collagène IX, *via* ses modules, dans la cohérence de la matrice cartilagineuse.

La matrice cartilagineuse est particulièrement riche en protéoglycanes, le plus représenté dans ce tissu étant l'agrécan. Ce dernier se lie de manière spécifique et réversible avec l'acide hyaluronique. Une deuxième protéine, la protéine de liaison du cartilage participe également à cette interaction et la stabilise [10]. Cette protéine a, en effet, une double capacité de liaison : à l'agrécan, d'une part, et à l'acide hyaluronique, d'autre part (figures 1A et 2). Ces trois composants sont tous synthétisés par les chondrocytes,

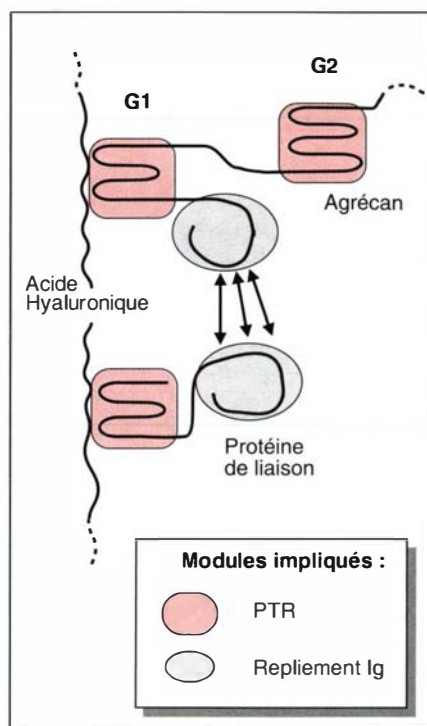


Figure 2. **Interaction de l'agrécane et de la protéine de liaison avec l'acide hyaluronique.** Les modules PTR (proteoglycan tandem repeat) et repliement Ig (IG) sont indiqués. Le domaine G1 du protéoglycane et la protéine de liaison interagissent tous deux avec l'acide hyaluronique, et interagissent entre eux via leur module IG (flèches). Ce complexe ternaire est ainsi stabilisé.

mais seuls ceux comportant une chaîne polypeptidique, l'agrécane et la protéine de liaison, suivent la voie de synthèse classique passant par le réticulum endoplasmique et l'appareil de Golgi. Ces deux macromolécules pourraient donc déjà interagir dans ces compartiments, alors qu'ils ne rencontreront l'acide hyaluronique qu'après leur exportation dans le milieu extracellulaire. Ces trois éléments interagissent de façon coopérative : des assemblages *in vitro* des trois partenaires pris deux à deux ont démontré que l'affinité de l'agrécane pour l'acide hyaluronique était fortement augmentée quand la protéine de liaison participait à l'interaction [10]. Ces éléments semblent donc chacun porter une interaction spécifique, tout en agissant en synergie (figure 2).

L'agrécane est constitué d'un grand noyau protéique à structure modulaire, sur lequel sont fixées de nombreuses chaînes de kératane sulfate et de chondroïtine. Observé en microscopie électronique après ombrage

tournant, il montre trois globules (G1 à G3) séparés par deux segments allongés. La séquence de l'ADNc de l'agrécane humain [11] a montré que son globule G1 est homologue à la protéine de liaison. Ils sont en effet tous deux constitués de deux types de modules : un module repliement Ig (*Ig fold*, IG) N-terminal, et un module répété en tandem caractéristique des protéoglycane nommé PTR (*proteoglycan tandem repeat*), formant une double boucle stabilisée par un pont disulfure. Le globule G2 de l'agrécane porte également ce module en tandem, mais pas de repliement Ig (figure 3B).

Le module à repliement Ig, présent dans la séquence N-terminale du globule G1 de l'agrécane et de la protéine de liaison présente une homologie avec les domaines hypervariables des anticorps [12]. Ce module est très probablement impliqué dans les interactions agrécane-protéine de liaison qui stabilisent la formation du complexe agrécane-acide hyaluronique-protéine de liaison (figure 2). Cette in-

teraction est donc directement responsable de l'immobilisation des charges des glycanes au sein de la matrice cartilagineuse.

Il existe d'autres composants du cartilage, dont les interrelations ne sont pas encore complètement connues et qui ont aussi une structure modulaire comme la CMP (*cartilage matrix protein*) et la fibromoduline. Il apparaît donc que la matrice cartilagineuse constitue un réseau d'interactions moléculaires auquel de nombreuses protéines et leurs chaînes de glycosaminoglycane participent. Ce réseau est fondé par l'interaction de modules de différentes macromolécules et détermine les fonctions du tissu. Des protéines, présentes dans d'autres tissus et encore moins bien caractérisées, sont analogues à certaines protéines du cartilage et pourraient avoir des fonctions similaires. C'est le cas du protéoglycane appelé versican décrit pour le derme, et qui a une structure modulaire voisine de celle de l'agrécane.

La jonction derme-épiderme

La peau est composée de deux tissus, l'épiderme et le derme. L'épiderme, formé à partir d'une fine couche de kératinocytes, recouvre le derme, un tissu fibreux élastique beaucoup plus épais. Ces deux tissus sont solidarisés par une zone d'adhérence centrée autour de la membrane basale épidermique, la jonction dermo-épidermique. Cette dernière se distingue par la présence de structures d'ancrage caractéristiques. Du côté de l'épiderme, les hémidesmosomes assurent la jonction entre les kératinocytes et la membrane basale. Du côté du derme, les fibres d'ancrage, composées de collagène VII, interagissent avec la membrane basale et des fragments de celle-ci, les plaques d'ancrage, pour former un réseau qui piège les fibres de collagène de la couche supérieure du derme (figure 1B).

Cette structure est très importante pour le maintien de la cohésion derme-épiderme. En effet, si l'ensemble de ces interactions n'est pas maintenu, on peut arriver à une situation pathologique où cette zone est fragilisée donnant lieu à un décollement de l'épiderme. C'est le cas dans deux formes d'épidermolyses bulleuses (EB), l'EB jonctionnelle dans laquelle

RÉFÉRENCES

15. Aumailley M, Verrando P. Structure et pathologie des membranes basales. *médecine/sciences* 1993 ; 9 : 926-33.
16. Paulsson M. Basement membrane proteins : structure, assembly, and cellular interactions. *Crit Rev Biochem Mol Biol* 1992 ; 27 : 93-127.
17. Timpl R. Structure and biological activity of basement membrane proteins. *Eur J Biochem* 1989 ; 180 : 487-502.
18. Burgeson RE, Chiquet M, Deutzmann R, Ekblom P, Engel J, Kleinman H, Martin GR, Meneguzzi G, Paulsson M, Sanes J, Timpl R, Tryggvason K, Yamada Y, Yurchenco PD. A new nomenclature for the laminins. *Matrix Biol* 1994 ; 14 : 209-11.
19. Haaparanta T, Uitto J, Ruoslahti E, Engvall E. Molecular cloning of the cDNA encoding human laminin A chain. *Matrix* 1991 ; 11 : 151-60.
20. Gerl M, Mann K, Aumailley M, Timpl R. Localization of a major nidogen-binding site to domain III of laminin B2 chain. *FEBS Lett* 1991 ; 202 : 167-74.
21. Aeschlimann D, Paulsson M. Cross-linking of Laminin-Nidogen complexes by tissue transglutaminase. *J Biol Chem* 1991 ; 266.
22. Paulsson M, Aumailley M, Deutzmann R, Timpl R, Beck K, Engel J. Laminin-nidogen complex. Extraction with chelating agents and structural characterization. *Eur J Biochem* 1987 ; 166 : 11-9.
23. Durkin ME, Chakravarti S, Bartos BB, Liu SH, Friedman RL, Chung AE. Amino acid sequence and domain structure of entactin. Homology with epidermal growth factor precursor and low density lipoprotein receptor. *J Cell Biol* 1988 ; 107 : 2749.
24. Iozzo RV. Perlecan : a gem of a proteoglycan. *Matrix Biol* 1994 ; 14 : 203-8.
25. Murdoch AD, Dodge GR, Cohen I, Tuan RS, Iozzo RV. Primary structure of the human heparan sulfate proteoglycan from basement membrane (HSPG2/perlecan). *J Biol Chem* 1992 ; 267 : 8455-7.
26. Battaglia C, Mayer U, Aumailley M, Timpl R. Basement membrane heparan sulfate proteoglycan binds to laminin by its heparan sulfate chains and to nidogen by sites in the protein core. *Eur J Biochem* 1992 ; 208 : 359-66.
27. Christiano AM, Rosenbaum LM, Chung-Honet LC, Parente MG, Woodley DT, Pan TC, Zhang RZ, Chu ML, Burgeson RE, Uitto J. The large non collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor. *Hum Mol Genet* 1992 ; 1 : 475-81.
- le décollement se fait entre l'épiderme et la membrane basale, et l'EB dystrophique dans laquelle le décollement se fait entre le derme et la membrane basale [13-15]. La membrane basale est, elle-même, composée de laminine, de nidogène, de collagène IV et de perlecan (ou héparan sulfate protéoglycane), qui ont tous une structure modulaire (*figure 3*). Elle se forme par un processus d'auto-assemblage, ses composants contenant eux-mêmes l'information nécessaire à leur interaction [15, 16]. Le collagène IV, collagène non fibrillaire, existe sous trois formes distinctes [$\alpha 1(IV)$]₂ $\alpha 2(IV)$; [$\alpha 3(IV)$]₂ $\alpha 4(IV)$; [$\alpha 5(IV)$]₂ $\alpha 6(IV)$ avec des distributions tissulaires différentes. Chaque chaîne est divisée en trois grands domaines : une courte triple hélice N-terminale (domaine 7S), suivie d'une interruption précédant une longue triple hélice comportant de fréquentes interruptions des répétitions glycine-X-Y caractéristiques de la structure primaire des collagènes [1, 17]. Ce domaine hélicoïdal est relativement flexible. Il est suivi du côté C-terminal par un domaine globulaire NC1, composé de deux modules homologues adjacents. Ce domaine est capable d'interagir par ces modules avec le même domaine d'autres molécules de collagène IV. A l'autre extrémité de la molécule, les domaines 7S sont capables d'interagir pour réunir quatre molécules de collagène antiparallèles deux à deux. L'ensemble de ces interactions combinées permet la formation d'un filet à mailles lâches, qui constitue la charpente des membranes basales (*figure 1B*). Les laminines, molécules également trimériques, sont aussi capables d'auto-assemblage. La laminine-1, anciennement appelée laminine EHS, est un hétérotrimère composé de trois chaînes $\alpha 1$, $\beta 1$ et $\gamma 1$ (selon la nouvelle nomenclature [18], et anciennement appelées A, B1 et B2). Un trimère de laminine ressemble à une croix composée de trois bras courts porteurs de deux ou trois globules, et d'un bras long terminé par un globule, formé de l'assemblage en triple hélice α des trois chaînes $\alpha 1$, $\beta 1$ et $\gamma 1$. Ces trois chaînes ont une structure modulaire, dans laquelle des fragments de bras courts sont composés de nombreux modules EGF adjacents, et où l'extrémité C-terminale du bras long (chaîne $\alpha 1$) est composée de 5 modules G [19] (*figure 3C*). La laminine interagit par un module EGF d'un de ses bras courts avec le globule C-terminal d'une protéine plus petite, le nidogène [20]. Cette interaction peut être stabilisée par une liaison covalente établie par la transglutaminase tissulaire [21]. Par son globule N-terminal, le nidogène interagit également avec la triple hélice de collagène IV [22]. Dans le nidogène, les segments séparant les globules les uns des autres sont eux aussi constitués de modules EGF adjacents [23] (*figure 3C*). Ce complexe ternaire laminine-nidogène-collagène IV peut aussi lier des protéoglycanes, en particulier le perlecan [24] (*figure 1B*). Le noyau protéique du perlecan comprend des modules EGF et des modules homologues aux modules G C-terminaux de la laminine, et porte des chaînes glycaniques [25] (*figure 3B*). Par ces chaînes, il est capable d'interagir avec la partie C-terminale d'une molécule de laminine, et par sa partie protéique avec un globule du nidogène [26]. La laminine et le collagène IV comportent également des sites de reconnaissance pour des récepteurs cellulaires [16]. L'ensemble de ces interactions, dans lesquelles interviennent d'autres protéines modulaires moins connues comme BM40/SPARC/ostéonectine, ou la fibuline, déterminent la structure de toute membrane basale. Dans la jonction derme-épiderme, la membrane basale est jointe au derme sous-jacent par des fibres d'ancrage, composées de collagène VII, qui interagissent avec la membrane basale et des plaques d'ancrage. Le collagène VII a aussi une structure modulaire (*figure 3*). Il comprend deux domaines collagéniques en triple hélice séparés par un court domaine non hélicoïdal et deux domaines non collagéniques N et C-terminaux. Le domaine N-terminal est vaste, et composé d'un module CMP, de neuf modules Fn3 (fibronectine 3) consécutifs, suivi d'un module VWA [27] (*figure 3A*). Ces modules confèrent, par des mécanismes encore inconnus, des propriétés d'adhérence à cette extrémité du collagène VII, facilitant l'attachement des fibres qu'il forme à la membrane basale et aux plaques d'ancrage du derme. L'extrémité C-terminale est com-

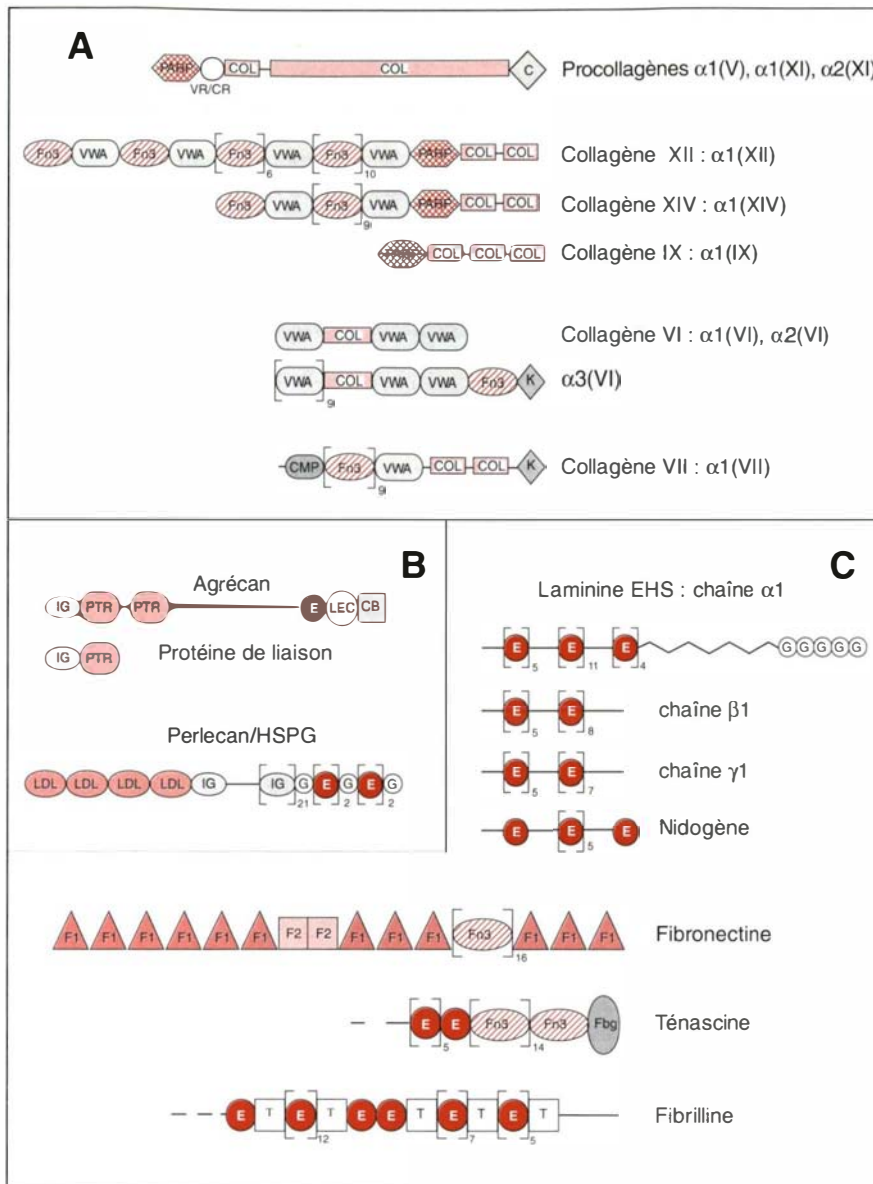


Figure 3. Structure modulaire des protéines de la matrice extracellulaire. Représentation schématique de quelques protéines de la matrice extracellulaire. **A.** Collagènes ; **B.** Protéoglycanes et protéine de liaison du cartilage ; **C.** autres molécules. Chaque module est représenté par un symbole particulier et un sigle. IG, module à repliement Ig ; PTR, protéoglycan tandem repeat ; E, module EGF ; LEC, lectine ; CB, complément B ; PARP, domaine homologue au peptide riche en proline-arginine ; VR/CR, homologie à une région constante (CR) et une région variable (VR) du globule N-terminal de certaines chaînes de procollagène [8] ; COL, domaine en triple hélice ; C, module C-propeptide ; F1, F2, Fn3, modules fibronectine de type I, II et III ; VWA, module von Willebrand de type A ; K, module Kunitz ; G, module G ; LDL, module des low density lipoproteins ; CMP, module de la cartilage matrix protein ; Fbg, module du fibrinogène ; T, modules du TGF β . Pour simplifier la figure, les longues répétitions de modules sont données entre crochets. Les zones des molécules ne présentant pas de structure modulaire sont schématisées par un trait simple. Le zig-zag des chaînes de laminine représente la portion en hélice α super-enroulée du trimère. La partie N-terminale de la chaîne $\alpha 2(XI)$ n'est connue qu'à partir du premier tiers du module PARP environ mais est représentée totalement ici.

posée d'un module rencontré dans les inhibiteurs de protéases de Kunitz, également rencontré dans une chaîne du collagène VI [28].

Assemblages supramoléculaires retrouvés dans le derme et d'autres tissus

Le derme lui-même est constitué d'éléments appartenant à plusieurs familles de macromolécules. Ces assemblages se retrouvent, dans des proportions diverses, dans bien d'autres tissus conjonctifs (figure 1C). De même que la charpente de la matrice cartilagineuse est composée de fibres mixtes comprenant à la fois les collagènes II et XI, celui de la matrice dermique est constitué de fibrilles de collagènes I et III, et d'un cœur de collagène V. En surface de ces fibres, les collagènes de la famille FACIT présents dans la peau sont les collagènes de type XII et XIV (figure 1C). Ces collagènes ont deux domaines en triple hélice (COL1 et COL2), le domaine COL1 étant constitué d'un module homologue à COL1 du collagène IX, et un large domaine non collagénique représentant environ 90 % de la protéine totale. Ce domaine non collagénique est constitué de modules Fn3 et VWA répétés un grand nombre de fois [29, 30] (figure 3A) et joue probablement un rôle dans des interactions protéine-protéine.

Les collagènes XII et XIV interagissent avec la fibre de collagène I de manière probablement indirecte. Récemment, l'interaction du collagène XIV avec la décorine, un protéoglycane, a été démontrée *in vitro*. Cette interaction se fait par la chaîne glycosaminoglycane de la décorine et semble mettre en jeu le module NC1 C-terminal [31] du collagène XIV (figure 1C). Il est connu par ailleurs que la décorine interagit, par son noyau protéique, avec les fibres de collagène I [32, 33]. On suppose donc l'existence *in vivo* de ce complexe quaternaire : fibre mixte de collagènes I et V, décorine et collagène XIV (figure 1C). Des résultats récents de notre laboratoire indiquent que le collagène XII interagit avec la décorine de manière comparable (Font *et al.*, résultats non publiés).

RÉFÉRENCES

28. Greenspan DS. The carboxy-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene. *Hum Mol Genet* 1993 ; 2 : 273-8.

29. Yamagata M, Yamada KM, Yamada SS, Shinomura T, Tanaka H, Nishida Y, Obara M, Kimata K. The complete primary structure of type XII collagen shows a chimeric molecule with reiterated fibronectin type III motifs, von Willebrand factor A motifs, a domain homologous to a noncollagenous region of type IX collagen, and short collagenous domains with an Arg-Gly-Asp site. *J Cell Biol* 1991 ; 115 : 209-21.

30. Gerecke DR, Foley JW, Castagnola P, Gennari M, Dublet B, Cancedda R, Linsenmayer TF, van der Rest M, Olsen BR, Gordon MK. Type XIV collagen is encoded by alternative transcripts with distinct 5' regions and is a multidomain protein with homologies to von Willebrand's factor, fibronectin, and other matrix proteins. *J Biol Chem* 1993 ; 268 : 12177-84.

31. Font B, Aubert-Foucher E, Goldschmidt D, Eichenberger D, van der Rest M. Binding of collagen XIV with the dermatan sulfate side chain of decorin. *J Biol Chem* 1993 ; 268 : 25015-8.

32. Brown DC, Vogel KG. Characteristics of the *in vitro* interaction of a small proteoglycan (PGII) of bovine tendon with type I collagen. *Matrix* 1989 ; 9 : 468-78.

33. Hedbom E, Heinegard D. Interaction of a 59-kDa connective tissue matrix protein with collagen I and collagen II. *J Biol Chem* 1989 ; 264 : 6898-905.

34. Maslen CL, Corson GM, Maddox BK, Glanville RW, Sakai LY. Partial sequence of a candidate gene for the Marfan syndrome. *Nature* 1991 ; 352 : 334-7.

35. Cooke RM, Wilkinson AJ, Baron M, Pastore A, Tappin MJ, Campbell ID, Gregory H, Sheard B. The solution structure of human epidermal growth factor. *Nature* 1987 ; 327 : 339-41.

36. Koller E, Winterhalter KH, Trueb B. The globular domains of type VI collagen are related to the collagen binding domains of cartilage matrix protein and von Willebrand factor. *EMBO J* 1989 ; 8 : 1073-7.

37. Rebay I, Fleming RJ, Fehon RG, Cherbass L, Cherbass P, Artavanis-Tsakonas S. Specific EGF repeats of Notch mediate interactions with Delta and Serrate : implications for Notch as a multifunctional receptor. *Cell* 1991 ; 67 : 687-99.

38. Handford PA, Mayhew M, Baron M, Winship PR, Campbell ID, Brownlee GG. Key residues involved in calcium-binding motifs in EGF-like domains. *Nature* 1991 ; 351 : 164-7.

Une autre structure est présente dans la peau comme dans le cartilage et d'autres matrices extracellulaires : il s'agit d'une structure microfibrillaire composée de collagène VI. Ce collagène a un court domaine en triple hélice, et de grands domaines globulaires N-terminaux représentant plus des deux tiers de la protéine, composés principalement de la répétition de modules VWA. L'extrémité C-terminale de la chaîne $\alpha 3(\text{VI})$ comporte en plus un module de type inhibiteur de

protéase de Kunitz et un module de protéine salivaire (figure 3A). Par ses modules VWA, les molécules de collagène VI interagissent entre elles et avec l'acide hyaluronique pour former une structure microfibrillaire régulière (figure 1C).

Les fibres élastiques de la peau sont composées d'élastine et de microfibrilles, dont le composant majeur est la fibrilline. Ces mutations sont responsables du syndrome de Marfan (m/s n° 8, vol. 7, p. 872). On ne connaît

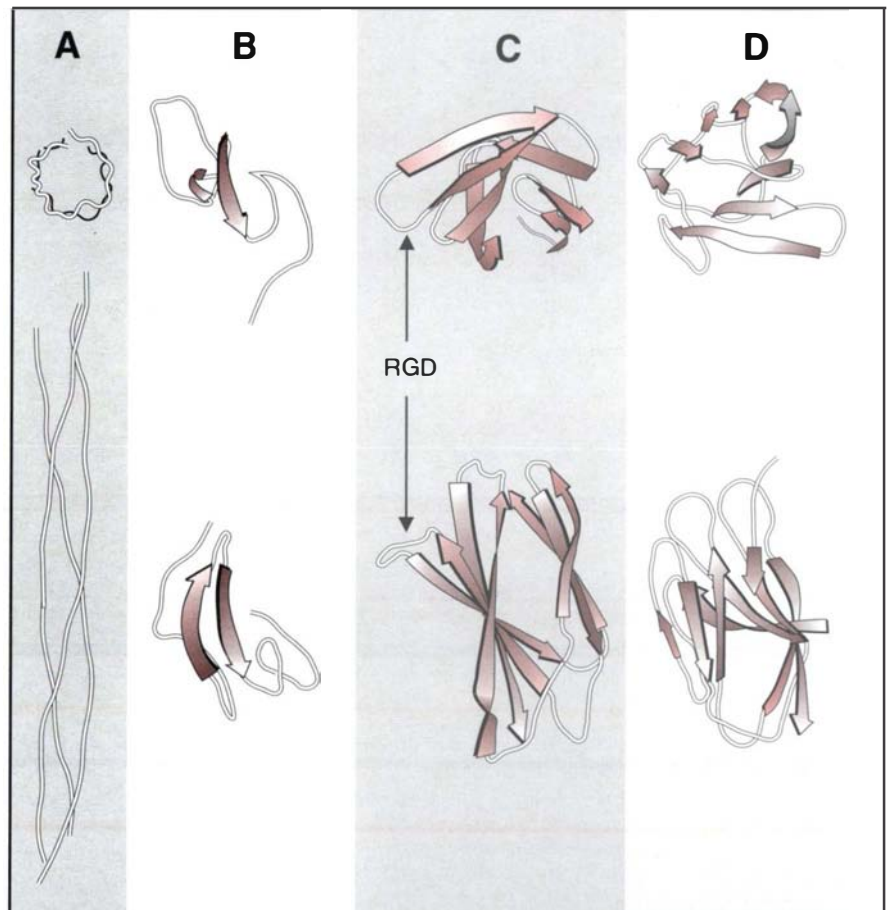


Figure 4. **Structure tridimensionnelle de quelques modules.** A. domaine collagénique ; B. module EGF ; C. module Fn3 ; D. module à repliement Ig. Les diagrammes en lacets (représentant les brins β) et en boucles ont été dessinés avec MOLSCRIPT [48] d'après les structures déduites de données cristallographiques présentes dans la base de données PDB (Protein Data Bank) [49] sous les références suivantes : A) 2CLG, B) P1EGF, C) 1TEN, D) 1REI. Les modules sont représentés en vue axiale (haut de la figure) et perpendiculaire (bas). Par commodité, le domaine collagénique représenté comprend 18 résidus par chaîne. La structure du module EGF est celle du facteur de croissance EGF humain. La structure du module Fn3 représentée est issue de la ténascine. La position de la séquence RGD dans cette structure est indiquée par une flèche.

pas encore le mode d'assemblage précis de cette molécule, mais sa séquence montre un grand nombre de modules EGF répétés [34] (*figure 3C*). Les microfibrilles se retrouvent aussi indépendamment de l'élastine dans de très nombreux tissus, en particulier dans le tissu de l'œil qui maintient le cristallin, la zonule (*figure 1C*).

In vivo, l'ensemble de ces structures est probablement très imbriqué, par le jeu d'interactions croisées. D'autres molécules, telles que des glycoprotéines à expression transitoire (vitronectine, ténascine, thrombospondine...), ou d'autres protéoglycanes (versican, biglycan, fibromoduline...), sont présents dans ces tissus. Leur rôle et les interactions dont elles sont capables sont encore mal connus.

Ainsi, la structure des matrices extracellulaires met en jeu des réseaux d'interactions entre des partenaires multiples, celles-ci se réalisant d'une molécule à l'autre principalement par des modules particuliers.

Les modules des protéines matricielles

Nous avons rencontré, dans les exemples discutés ci-dessus, des composants distincts construits cependant à partir de modules homologues. Les homologies structurales des modules n'impliquent pas forcément une stricte homologie de fonction pour les différentes protéines qui les portent. Pour faciliter l'analyse, nous avons regroupé ici les modules, suivant des caractéristiques générales. Cependant, en raison de la diversité des fonctions que peuvent porter des modules homologues, ces regroupements ne peuvent pas constituer en eux-mêmes une classification.

Les modules arrangés par groupes et jouant un rôle structural

Dans de nombreuses macromolécules de la matrice extracellulaire, certains modules sont fréquemment arrangés en groupes, c'est-à-dire qu'ils sont répétés, parfois un grand nombre de fois, et sont souvent adjacents. C'est notamment le cas des modules EGF, Fn3 et VWA (*figure 3*). L'observation des molécules consti-

tuées de ces groupes en microscopie électronique, après ombrage tournant, montre que fréquemment, ces répétitions constituent dans la protéine des segments en bâtonnet, des bras. Ainsi, les bras courts de la laminine forment des bâtonnets ponctués de deux ou trois globules. Dans la séquence de la molécule, les courts segments séparant les globules sont constitués de plusieurs modules EGF adjacents. La structure d'un module EGF isolé a été déduite de la structure connue à partir de l'EGF cristallisé. [35]. La protéine est repliée en feuillets β stabilisés par trois ponts disulfure (*figure 4B*). Elle forme donc une structure très compacte, dont la répétition en chaîne donne souvent lieu à des segments résistants aux protéases.

De même, la ténascine est constituée de six chaînes polypeptidiques disposées en tentacules, chacune étant constituée de 14 modules EGF formant un bras fin, suivis de 11 modules Fn3 continuant ce bras en l'épaississant. La fibronectine, dimère en forme de V, comporte également un long segment constitué de la répétition de modules Fn3.

La triple hélice de type «collagène» (*figures 3C et 4A*) est une conformation que l'on retrouve dans de nombreuses protéines de la matrice extracellulaire (les collagènes), mais également dans des protéines transmembranaires ou circulantes (Clq, acétylcholinestérase, récepteur du macrophage, ficoline, etc.). La notion de module peut s'appliquer à cette structure, même s'il est parfois difficile d'en définir les limites : en effet, les triples hélices participant à la structure de ces protéines ont des tailles variables, et peuvent être plus ou moins interrompues par de courtes séquences non répétitives.

Modules impliqués dans des interactions protéine-protéine

L'implication de certains modules dans des interactions spécifiques protéine-protéine a pu être démontrée. Ainsi, le module VWA du facteur de von Willebrand, facteur circulant jouant un rôle dans l'agrégation des plaquettes au cours de la coagulation, est capable d'interagir avec le collagène. Parmi les nombreuses protéines

de la matrice portant ce module, certaines interagissent aussi avec le collagène, mais l'implication de leurs modules VWA dans cette interaction n'a pas encore été démontrée. Le collagène VI, par exemple, possède plusieurs modules VWA à ses extrémités. On suppose ainsi que l'assemblage de plusieurs molécules en une structure microfibrillaire (*figure 1C*), se fait grâce à l'interaction de ces modules d'une molécule avec la triple hélice collagénique d'une autre molécule [36].

Les modules EGF, outre leur rôle structural quand ils sont arrangés en groupes, peuvent aussi, dans certains cas, jouer un rôle dans des interactions protéine-protéine. Cela a été démontré pour deux molécules jouant un rôle dans le développement de la drosophile, Notch et Delta. Dans ces molécules, il semble que chaque module EGF d'un groupe soit impliqué dans une interaction différente. Tous ne seraient donc pas équivalents et interchangeable [37]. Dans la fibrilline, qui comprend plus de quarante modules EGF [33], on connaît plusieurs mutations dans ces modules. Ces mutations entraînent l'apparition d'un syndrome de Marfan, caractérisé par des défauts des systèmes squelettique, oculaire et cardiovasculaire. L'existence de cette maladie nous prouve, là encore, que le défaut porté par un seul module peut affecter la fonction de la protéine entière. Ici, la mutation de la protéine entraîne un défaut de l'assemblage en une structure microfibrillaire, mettant en jeu des interactions protéine-protéine.

Certains modules EGF de la fibrilline ou d'autres molécules portent une séquence consensus de β -hydroxylation sur une asparagine ou un acide aspartique particulier. Quand cet acide aminé est ainsi modifié, le module acquiert la capacité de lier le calcium, ce qui peut jouer un rôle dans les interactions protéine-protéine [38]. Plusieurs mutations de la fibrilline chez des patients atteints par la maladie de Marfan interviennent dans cette séquence consensus [39, 40]. D'autres mutations ont aussi été mises en évidence, en particulier au niveau de cystéines de modules EGF.

La fibronectine joue un rôle important dans l'interaction des cellules

RÉFÉRENCES

39. Lee B, Godfrey M, Vitale E, Hori H, Mattei M-G, Sarfarazi M, Tsipouras P, Ramirez F, Hollister DW. Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes. *Nature* 1991 ; 352 : 330-4.
40. Dietz HC, Cutting GR, Pyeritz RE, Maslen CL, Sakai LY, Corson GM, Puffenberger EG, Hamosh A, Nanthakumar EJ, Currisin SM, Stetten G, Meyers DA, Francómano CA. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 1991 ; 352 : 337-9.
41. Humphries MJ. The molecular basis and specificity of integrin-ligand interactions. *J Cell Sci* 1990 ; 97 : 585-92.
42. Bork P. Mobile modules and motifs. *Curr Opin Struct Biol* 1992 ; 2 : 413-21.
43. Leahy DJ, Hendrickson WA, Aukhil I, Erickson HP. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* 1992 ; 258 : 987-91.
44. Moradi-Améli M, Deléage G, Geourgon C, van der Rest M. Common topology within a non-collagenous domain of several different collagen types. *Matrix Biol* 1994 ; 14 : 233-9.
45. Engel J, Prockop DJ. The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annu Rev Biophys Chem* 1991 ; 20 : 137-52.
46. Labourdette L, van der Rest M. Analysis of the role of the COL1 domain and its adjacent cysteine-containing sequence in the chain assembly of type IX collagen. *FEBS Lett* 1993 ; 320 : 211-4.
47. Mazzorana M, Gruffat H, Sergeant A, van der Rest M. Mechanisms of collagen trimer formation. *J Biol Chem* 1993 ; 268 : 3029-32.
48. Kraulis PJ. MOLSCRIPT : a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 1991 ; 24 : 946-50.
49. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice M, Rodgers JR, Kennard O, Shinamouchi T, Tasumi M. The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 1977 ; 112 : 535-42.

avec la matrice extracellulaire en particulier grâce à des récepteurs spécifiques de type intégrines. L'interaction avec certaines intégrines met en jeu la reconnaissance d'une séquence arginine-glycine-acide aspartique (RGD). Un seul des modules Fn3 de la fibronectine porte ce tripeptide RGD, et constitue le site d'interaction avec les cellules [41]. L'étude par RMN de modules Fn3 et de la diffraction des rayons X par ces modules issus de différentes macromolécules ont démontré une structure formée de deux feuillets comprenant respectivement quatre et trois brins β [42] (figure 4C). Le tripeptide RGD est situé dans une boucle particulièrement accessible de ce module Fn3 (figure 4). Un autre module Fn3, appartenant à une autre molécule, la ténascine, porte une séquence RGD à la même position [43]. Aucun autre des 300 modules Fn3 connus à ce jour n'a de séquence RGD, et ne semble jouer un rôle dans l'interaction avec les cellules. Inversement, si le module Fn3 de la fibronectine portant la séquence RGD est déterminant pour la spécificité des interactions de la molécule avec la cellule, cette interaction est plus forte si l'ensemble de la molécule y participe, ce qui suggère l'existence de sites secondaires d'adhérence, agissant en synergie avec le site RGD [41].

La structure de certains modules trouvés isolément et non en groupe suggère leur implication dans des interactions protéine-protéine. Par exemple, les modules à repliement Ig et PARP/NC ont des structures voisines des boucles hypervariables des immunoglobulines, impliquées dans les phénomènes de reconnaissance antigène-anticorps. Le module à repliement Ig est constitué de sept brins constituant deux feuillets β et intercalant trois boucles hypervariables (figure 4D). La structure exacte du module PARP/NC n'a pas encore été déterminée, mais un modèle prédisant un repliement en neuf brins β antiparallèles est proposé [44]. D'autres structures en feuillets β sont rencontrées dans des protéines impliquées dans des mécanismes de reconnaissance moléculaire. Il est donc légitime de postuler un rôle d'interaction aux domaines protéiques portant des modules à repliement Ig ou PARP/NC.

Modules jouant un rôle dans l'assemblage moléculaire

Les collagènes fibrillaires sont tous construits sur le même schéma avec un N-propeptide et un C-propeptide, de part et d'autre d'une longue triple hélice centrale, qui seront clivés après sécrétion du trimère. Si les N-propeptides ont des structures diverses, les C-propeptides sont tous homologues (figure 3). La traduction terminée, les chaînes s'assemblent en trimères par leurs C-propeptides. C'est ce module qui porte les fonctions de reconnaissance et d'assemblage des chaînes en trimères. L'assemblage est stabilisé par des ponts disulfure. À partir de l'amorce trimérique formée par l'association des trois C-propeptides de trois chaînes, et après certaines modifications post-traductionnelles (hydroxylation et glycosylation), la formation de la triple hélice commence et progresse de proche en proche à la manière d'une fermeture éclair, jusqu'à l'extrémité N-terminale [45]. Les molécules assemblées sont exportées, et leurs N et C-propeptides sont clivés dans l'espace extracellulaire, ce qui permet la formation de fibres.

Pour les autres collagènes, les mécanismes d'assemblage sont moins connus. Pour les FACIT, qui comprennent les collagènes IX, XII, XIV et XVI, nous avons montré au laboratoire que c'est le module C-terminal COL1 qui est impliqué dans la formation du trimère. Ces collagènes ne comportent pas de C-propeptide clivé. Cependant, ils comportent tous le module très conservé COL1, dont la jonction avec NC1 comporte deux cystéines espacées de quatre acides aminés conservés. Un fragment protéolytique du collagène IX, comprenant COL1 avec ces deux cystéines, est capable *in vitro* de reformer des assemblages trimériques en triple hélice à partir des chaînes séparées et purifiées [46]. Il a par ailleurs été démontré par transfection de mini-gènes de collagène XII recombinants, codant pour ce module, qu'il contient toute l'information pour l'assemblage des chaînes de collagène XII, et leur repliement en triple hélice [47].

Conclusion

Nous avons ici discuté les rôles joués par quelques modules parmi les

mieux connus. Il en existe cependant un nombre beaucoup plus important : on connaît en effet actuellement une cinquantaine de modules différents participant aux structures des protéines extracellulaires. La comparaison systématique des séquences des protéines telles qu'elles apparaissent dans les bases de données, avec les motifs des modules répertoriés, montre que certains modules sont représentés jusqu'à 1 200 fois au sein des protéines connues à ce jour. La grande variété des protéines extracellulaires, et en particulier des protéines de la matrice, s'est ainsi établie à partir d'un nombre relativement restreint de modules constitutifs, associés dans la structure d'une protéine en de multiples combinaisons.

Les différentes matrices extracellulaires sont constituées de l'assemblage ordonné de macromolécules, spécifiant les fonctions d'un tissu. L'analyse au niveau moléculaire des interactions mises en jeu dans les assemblages des matrices extracellulaires met en évidence le rôle clé que jouent les modules dans ces interactions. La connaissance des mécanismes d'assemblage impliqués est essentielle à une meilleure compréhension des matrices formées dans des conditions normales et pathologiques ■

Remerciements

Nous remercions G. Deléage et C. van Herwege pour leur aide à l'élaboration des figures.

Summary

Molecular interactions and modularity of extracellular matrix proteins

During the past ten years, numerous reports have demonstrated the presence of homologous sequences repeated in different proteins. These motifs, that share common structures, are called modules. The great diversity of secreted proteins and particularly of those found in the extracellular matrix can in fact be reduced to a relatively small number of distinct modules. In this review, we underline the importance of modularity of extracellular matrix macromolecules in their supra-molecular assemblies and in their physiological roles. The modules from which a particular protein is built up can be involved both in the structure of the protein, acting for example as spacers, and in functional aspects, leading to specific interactions with other extracellular proteins or with specific membrane receptors. The resulting interactions account for the integrity of extracellular matrices. A mutation affecting a single module can be responsible for a loss of function of the entire protein. This can lead to severe heritable disorders.

TIRÉS À PART

M. van der Rest.