

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2005-504037

(P2005-504037A)

(43) 公表日 平成17年2月10日(2005.2.10)

(51) Int. Cl. ⁷	F I	テーマコード (参考)
C07K 14/78	C07K 14/78 ZNA	2G045
A61K 38/00	A61K 45/00	4C084
A61K 45/00	A61P 1/04	4H045
A61P 1/04	A61P 3/02 IOI	
A61P 3/02	A61P 3/10	
審査請求 有 予備審査請求 未請求 (全 227 頁) 最終頁に続く		

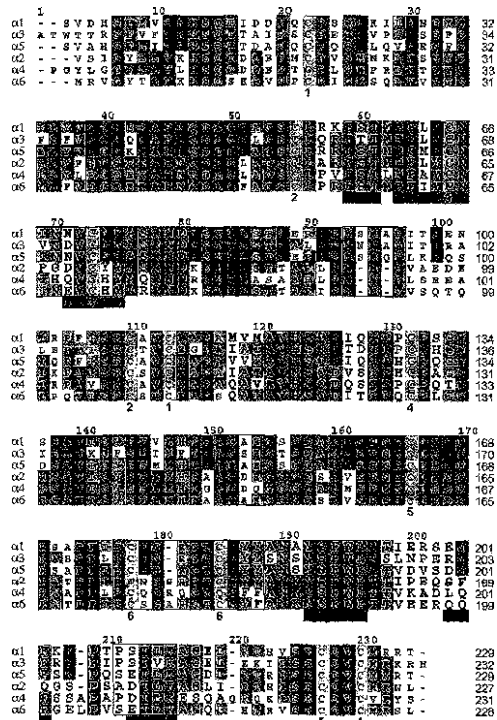
(21) 出願番号 特願2003-517296 (P2003-517296)
 (86) (22) 出願日 平成14年7月26日 (2002.7.26)
 (85) 翻訳文提出日 平成16年1月26日 (2004.1.26)
 (86) 国際出願番号 PCT/US2002/023763
 (87) 国際公開番号 W02003/012122
 (87) 国際公開日 平成15年2月13日 (2003.2.13)
 (31) 優先権主張番号 60/308, 523
 (32) 優先日 平成13年7月27日 (2001.7.27)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 60/351, 289
 (32) 優先日 平成13年10月29日 (2001.10.29)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 60/366, 854
 (32) 優先日 平成14年3月22日 (2002.3.22)
 (33) 優先権主張国 米国 (US)

(71) 出願人 500224346
 ユニヴァースティ オブ カンザス メディカル センター
 アメリカ合衆国 66160-7702
 カンザス州 カンザスシティ レインボー ブールヴァード 3901
 (74) 代理人 100083806
 弁理士 三好 秀和
 (72) 発明者 サンドラムーアシー、 ムイラシナム
 アメリカ合衆国 カンザス州 66160
 カンザス シティ レインボー ブールバード 3901 ユニバーシティ オブ カンザス メディカル センター デパートメント オブ バイオケミストリー アンド モレキュラー バイオロジー
 最終頁に続く

(54) 【発明の名称】 I V型コラーゲンNC1ドメイン六量体の結晶化構造

(57) 【要約】

本発明は、I V型コラーゲンの結晶化NC1ドメイン六量体、及びその結晶の製造方法を提供し、ここでNC1ドメイン六量体は、結晶化NC1ドメイン六量体の三次元構造を少なくとも3Å以下の解像度で決定することができるように結晶化される。本発明は、血管新生、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、及び/または基底層構築を阻害するように化合物を設計するための方法であって、本発明の方法によって生成される結晶化I V型コラーゲンNC1ドメイン六量体の三次元構造を解析し、並びにその解析によってI V型コラーゲンヘテロ三量体及び六量体構築に重要であると同定されているNC1ドメインの領域を標的とする成分を同定及び合成することを含む方法も提供する。本発明は、本発明の合理的薬物設計方法により、ここに開示されるI V型コラーゲンNC1六量体構造の解析に基づいて設計される新規ポリペプチドも提供する。



【特許請求の範囲】

【請求項 1】

少なくとも 8 個の連続するアミノ酸からなるポリペプチドで、その一般式 I は：

$P F (R 1) (R 2) C N (R 3) (R 4) (R 5) V C (R 6) (R 7) A$ (配列番号 1)

ここで、R 1 は L、M、A、V、nor L、及び I からなる群より選択され；

R 2 は F 及び Y からなる群より選択され；

R 3 は I、V、L、nor L、A、及び P からなる群より選択され；

R 4 は N、G、及び H からなる群より選択され；

R 5 は N、D、Q、及び E からなる群より選択され；

R 6 は N、Y、及び H からなる群より選択され；及び

R 7 は F 及び Y からなる群より選択される。

10

【請求項 2】

一般式 I のアミノ酸配列からなる請求項 1 に記載のポリペプチド。

【請求項 3】

R 2 が F であり；

R 4 が N であり；

R 5 が N 及び D からなる群より選択され；

R 6 が N であり；かつ

R 7 が F である、

請求項 1 に記載のポリペプチド。

20

【請求項 4】

R 2 が Y であり；

R 3 が P 及び I からなる群より選択され；

R 5 が D、Q、及び E からなる群より選択され；

R 6 が Y 及び H からなる群より選択され；かつ

R 7 が Y である、

請求項 1 に記載のポリペプチド。

【請求項 5】

ポリペプチドが、 $P F L F C N I N N V C N F A$ (配列番号 2)； $P F L F C N V N D V C N F A$ (配列番号 3)； $P F M F C N I N N V C N F A$ (配列番号 4)； $P F L Y C N P G D V C Y Y A$ (配列番号 5)； $P F A Y C N I H Q V C H Y A$ (配列番号 6)；及び $P F I Y C N I N E V C H Y A$ (配列番号 7) からなる群より選択される、請求項 1 に記載のポリペプチド。

30

【請求項 6】

少なくとも 7 個の連続するアミノ酸からなるポリペプチドで、一般式 II は：

$P F (R 1) E C (R 2) G (R 3) (R 4) G T C (R 5)$ (配列番号 8)

ここで、R 1 は L、A、V、nor L、及び I からなる群より選択され；

R 2 は H、N、Q、及び S からなる群より選択され；

R 3 は G、R、A、または不在からなる群より選択され；

R 4 は R 及び Q からなる群より選択され；かつ

R 5 は N 及び H からなる群より選択される。

40

【請求項 7】

一般式 II のアミノ酸配列からなる請求項 6 に記載のポリペプチド。

【請求項 8】

R 2 が H であり；

R 3 が R であり；

R 4 が G であり；かつ

R 5 が N である、

請求項 6 に記載のポリペプチド。

50

【請求項 9】

R 2 が N、Q、及び S からなる群より選択され；

R 3 が G、R、及び A からなる群より選択され；

R 4 が R 及び Q からなる群より選択され；かつ

R 5 が H である、

請求項 6 に記載のポリペプチド。

【請求項 10】

ポリペプチドが、P F I E C H G R G T C N (配列番号 9) ; P F L E C H G R G T C N (配列番号 10) ; P F I E C N G G R G T C H (配列番号 11) ; P F L E C Q G R Q G T C H (配列番号 12) ; 及び P F I E C S G A R G T C H (配列番号 13) からなる群より選択される、請求項 6 に記載のポリペプチド。 10

【請求項 11】

(a) E F R S A P F I E C H G R G T C N Y Y A N A (配列番号 14) 、
 (b) E F R A S P F L E C H G R G T C N Y Y S N S (配列番号 15) ;
 (c) E F R S A P F I E C H G R G T C N Y Y A N S (配列番号 16) ;
 (d) D F R A T P F I E C N G G R G T C H Y Y A) N K (配列番号 17) ;
 (e) D F R A A P F L E C Q G R Q G T C H F F A N K (配列番号 18) ; 及び
 (f) D F R A T P F I E C S G A R G T C H Y F A N K (配列番号 19)
 からなる群より選択される、少なくとも 13 アミノ酸からなるポリペプチド。

【請求項 12】

(a) 請求項 1 によるポリペプチド；
 (b) 請求項 6 によるポリペプチド；及び
 (c) 0 から 20 の間のアミノ酸からなるポリペプチドリンカー
 からなるキメラポリペプチド。 20

【請求項 13】

(a) 請求項 5 によるポリペプチド；
 (b) 請求項 10 によるポリペプチド；及び
 (c) 2 アミノ酸からなるポリペプチドリンカー、
 からなる請求項 12 に記載のキメラポリペプチド。

【請求項 14】

一般式 I I I :

F (R 1) T (R 2) (配列番号 20)

(ここで、R 1 は S 及び T からなる群より選択され；かつ

R 2 は M 及び L からなる群より選択される)

の配列からなるポリペプチド。

【請求項 15】

ポリペプチドが、F S T M (配列番号 21) 、 F T T M (配列番号 22) 及び F T S L (配列番号 23) からなる群より選択される、請求項 14 に記載のポリペプチド。

【請求項 16】

(a) X 1 - F S T M - Z 1、ここで、X 1 は配列 S C L R K (配列番号 24) の 0、1、2、3、4、または 5 アミノ酸であり、及び Z 1 は配列 P F L F C (配列番号 25) の 0、1、2、3、4、または 5 アミノ酸である； 40
 (b) X 3 - F T T M - Z 3、ここで、X 3 は配列 S C L Q R (配列番号 27) の 0、1、2、3、4、または 5 アミノ酸であり、及び Z 3 は配列 P F L F C (配列番号 25) の 0、1、2、3、4、または 5 アミノ酸である；
 (c) X 5 - F S T M - Z 5、ここで、X 5 は配列 S C L R R (配列番号 29) の 0、1、2、3、4、または 5 アミノ酸であり、及び Z 5 は配列 P F M F C (配列番号 30) の 0、1、2、3、4、または 5 アミノ酸である；
 (d) X 2 - F S T M - Z 2、ここで、X 2 は配列 S C L A R (配列番号 32) の 0、1、2、3、4、または 5 アミノ酸であり、及び Z 2 は配列 P F L Y C (配列番号 33) の 50

0、1、2、3、4、または5アミノ酸である；

(e) X4 - F S T L - Z4、ここで、X4は配列 S C L P V (配列番号35)の0、1、2、3、4、または5アミノ酸であり、及びZ4は配列 P F A Y C (配列番号36)の0、1、2、3、4、または5アミノ酸である；及び

(f) X6 - F S T M - Z6、ここで、X6は配列 S C L P R (配列番号38)の0、1、2、3、4、または5アミノ酸であり、及びZ6は配列 P F I Y C (配列番号39)の0、1、2、3、4、または5アミノ酸である、

からなる群より選択されるポリペプチド。

【請求項17】

一般式IV：

(R1) M F (R2) K (配列番号41)

(ここで、R1はE、R、及びDからなる群より選択され；かつ

R2はK、R、及びSからなる群より選択される)

の配列からなるポリペプチド。

【請求項18】

ポリペプチドが、E M F K K (配列番号42)、R M F R K (配列番号43)、及びD M F S K (配列番号44)からなる群より選択される、請求項17に記載のポリペプチド。

【請求項19】

(a) X1 - E M F K K - Z1、ここで、X1は配列 T I E R S (配列番号48)の0、1、2、3、4、または5アミノ酸であり、Z1は配列 P T P S T (配列番号49)の0、1、2、3、4、または5アミノ酸である；

(b) X3 - R M F R K - Z3、ここで、X3は配列 S L N P E (配列番号51)の0、1、2、3、4、または5アミノ酸であり、Z3は配列 P I P S T (配列番号52)の0、1、2、3、4、または5アミノ酸である；

(c) X5 - D M F S K - Z5、ここで、X5は配列 T V D V S (配列番号54)の0、1、2、3、4、または5アミノ酸であり、Z5は配列 P Q S E T (配列番号55)の0、1、2、3、4、または5アミノ酸である；

からなる群より選択されるポリペプチド。

【請求項20】

S F Q (配列番号45)；L Q F (配列番号46)、及びQ Q F (配列番号47)からなる群より選択されるポリペプチド。

【請求項21】

(a) X2 - S F Q - Z2、ここで、X2は配列 T I P E Q (配列番号57)の0、1、2、3、4、または5アミノ酸であり、Z2は配列 G S P S A (配列番号58)の0、1、2、3、4、または5アミノ酸である；

(b) X4 - L Q F - Z4、ここで、X4は配列 T V K A D (配列番号60)の0、1、2、3、4、または5アミノ酸であり、Z4は配列 S S A P A (配列番号61)の0、1、2、3、4、または5アミノ酸である；及び

(c) X6 - Q Q F - Z6、ここで、X6は配列 T V E E R (配列番号63)の0、1、2、3、4、または5アミノ酸であり、Z6は配列 G E L P V (配列番号64)の0、1、2、3、4、または5アミノ酸である、

からなる群より選択されるポリペプチド。

【請求項22】

一般式V：

(R1) A H (R2) Q D (配列番号66)

(ここで、R1はR及びKからなる群より選択され；かつ

R2はG及びNからなる群より選択される)

の配列からなるポリペプチド。

【請求項23】

ポリペプチドが、R A H G Q D (配列番号67)及びK A H N Q D (配列番号68)から

10

20

30

40

50

なる群より選択される配列からなる、請求項 22 に記載のポリペプチド。

【請求項 24】

(a) X1 - R A H G Q D - Z1、ここで、X1 は配列 V Q G N E (配列番号 69) の 0、1、2、3、4、または 5 アミノ酸であり、Z1 は配列 L G T A G (配列番号 70) の 0、1、2、3、4、または 5 アミノ酸である；

(b) X3 - R A H G Q D - Z3、ここで、X3 は配列 V Q G N Q (配列番号 72) の 0、1、2、3、4、または 5 アミノ酸であり、Z3 は配列 L G T L G (配列番号 73) の 0、1、2、3、4、または 5 アミノ酸である；

(c) X5 - R A H G Q D - Z5、ここで、X5 は配列 V Q G N K (配列番号 75) の 0、1、2、3、4、または 5 アミノ酸であり、Z5 は配列 L G T A G (配列番号 70) の 0、1、2、3、4、または 5 アミノ酸である；

(d) X2 - K A H N Q D - Z2、ここで、X2 は配列 F E G Q E (配列番号 77) の 0、1、2、3、4、または 5 アミノ酸であり、Z2 は配列 L G L A G (配列番号 78) の 0、1、2、3、4、または 5 アミノ酸である；

(e) X4 - K A H N Q D - Z4、ここで、X4 は配列 L E G Q E (配列番号 80) の 0、1、2、3、4、または 5 アミノ酸であり、Z4 は配列 L G L A G (配列番号 78) の 0、1、2、3、4、または 5 アミノ酸である；及び

(f) X6 - K A H N Q D - Z6、ここで、X6 は配列 V E G Q E (配列番号 82) の 0、1、2、3、4、または 5 アミノ酸であり、Z6 は配列 L G F A G (配列番号 83) の 0、1、2、3、4、または 5 アミノ酸である、

からなる群より選択されるポリペプチド。

【請求項 25】

一般式 V I :

(R1) G (R2) G Q (配列番号 85)

(ここで、R1 は E 及び Q からなる群より選択され；かつ

R2 は S、T、及び G からなる群より選択される)

の配列からなるポリペプチド。

【請求項 26】

ポリペプチドが、E G S G Q (配列番号 86)、E G T G Q (配列番号 87)、E G G G Q (配列番号 88) 及び Q G G G Q (配列番号 89) からなる群より選択される、請求項 25 に記載のポリペプチド。

【請求項 27】

(a) X1 - E G S G Q - Z1、ここで、X1 は配列 T S A G A (配列番号 90) の 0、1、2、3、4、または 5 アミノ酸であり、Z1 は配列 A L A S P (配列番号 91) の 0、1、2、3、4、または 5 アミノ酸である；

(b) X3 - E G T G Q - Z3、ここで、X3 は配列 T S A G S (配列番号 93) の 0、1、2、3、4、または 5 アミノ酸であり、Z3 は配列 A L A S P (配列番号 91) の 0、1、2、3、4、または 5 アミノ酸である；

(c) X2 - E G G G Q - Z2、ここで、X2 は配列 T A A G D (配列番号 95) の 0、1、2、3、4、または 5 アミノ酸であり、Z2 は配列 S L V S P (配列番号 96) の 0、1、2、3、4、または 5 アミノ酸である；

(d) X4 - Q G G G Q - Z4、ここで、X4 は配列 T G A G D (配列番号 98) の 0、1、2、3、4、または 5 アミノ酸であり、Z4 は配列 A L M S P (配列番号 99) の 0、1、2、3、4、または 5 アミノ酸である；及び

(e) X6 - E G G G Q - Z6、ここで、X6 は配列 T A A G A (配列番号 101) の 0、1、2、3、4、または 5 アミノ酸であり、Z6 は配列 S L V S P (配列番号 96) の 0、1、2、3、4、または 5 アミノ酸である、

からなる群より選択されるポリペプチド。

【請求項 28】

一般式 V I I :

10

20

30

40

50

(R 1) G (R 2) (R 3) (配列番号 1 0 3)
 (ここで、 R 1 は Q 及び E からなる群より選択され ;
 R 2 は N 及び Q からなる群より選択され ; かつ
 R 3 は E、 Q、 及び K からなる群より選択される)
 の配列からなるポリペプチド。

【請求項 2 9】

R 1 が Q であり、 かつ R 2 が N である、 請求項 2 8 に記載のポリペプチド。

【請求項 3 0】

ポリペプチドが、 Q G N E (配列番号 1 0 4)、 Q G N Q (配列番号 1 0 5)、 Q G N K
 (配列番号 1 0 6)、 及び E G Q E (配列番号 1 0 7) からなる群より選択される、 請求
 項 2 8 に記載のポリペプチド。 10

【請求項 3 1】

(a) X 1 - Q G N E - Z 1、 ここで、 X 1 は配列 S L L Y V (配列番号 1 0 8) の 0、
 1、 2、 3、 4、 または 5 アミノ酸であり、 Z 1 は配列 R A H G Q (配列番号 1 0 9) の
 0、 1、 2、 3、 4、 または 5 アミノ酸である ;
 (b) X 3 - Q G N Q - Z 3、 ここで、 X 3 は配列 S F L F V (配列番号 1 1 1) の 0、
 1、 2、 3、 4、 または 5 アミノ酸であり、 Z 3 は配列 R A H G Q (配列番号 1 0 9) の
 0、 1、 2、 3、 4、 または 5 アミノ酸である ;
 (c) X 5 - Q G N K - Z 5、 ここで、 X 5 は配列 S L L Y V (配列番号 1 0 8) の 0、
 1、 2、 3、 4、 または 5 アミノ酸であり、 Z 5 は配列 R A H G Q (配列番号 1 0 9) の
 0、 1、 2、 3、 4、 または 5 アミノ酸である ; 20
 (d) X 2 - E G Q E - Z 2、 ここで、 X 2 は配列 S L L Y F (配列番号 1 1 4) の 0、
 1、 2、 3、 4、 または 5 アミノ酸であり、 Z 2 は配列 K A H N Q (配列番号 1 1 5) の
 0、 1、 2、 3、 4、 または 5 アミノ酸である ;
 (e) X 4 - E G Q E - Z 4、 ここで、 X 4 は配列 S L L Y L (配列番号 1 1 7) の 0、
 1、 2、 3、 4、 または 5 アミノ酸であり、 Z 4 は配列 K A H N Q (配列番号 1 1 5) の
 0、 1、 2、 3、 4、 または 5 アミノ酸である ; 及び
 (f) X 6 - E G Q E - Z 6、 ここで、 X 6 は配列 S L L F V (配列番号 1 1 9) の 0、
 1、 2、 3、 4、 または 5 アミノ酸であり、 Z 6 は配列 K A H N Q (配列番号 1 1 5) の
 0、 1、 2、 3、 4、 または 5 アミノ酸である、 30
 からなる群より選択されるポリペプチド。

【請求項 3 2】

ポリペプチドが、 V Q G N E R (配列番号 1 2 1)、 V Q G N Q R (配列番号 1 2 2)、
 V Q G N K R (配列番号 1 2 3)、 F E G Q E K (配列番号 1 2 4)、 L E G Q E K (配
 列番号 1 2 5)、 及び V E G Q E K (配列番号 1 2 6) からなる群より選択される、 請求
 項 3 1 に記載のポリペプチド。

【請求項 3 3】

一般式 V I I I :

M (R 1) M (R 2) P (配列番号 1 2 7)

(ここで、 R 1 は S、 N からなる群から選択されるか、 または存在せず ; かつ
 R 2 は A、 Q からなる群より選択されるか、 または存在しない) 40
 の配列からなるポリペプチド。

【請求項 3 4】

ポリペプチドが、 M S M A P (配列番号 1 2 8)、 M N M A P (配列番号 1 2 9)、 M S
 M Q P (配列番号 1 3 0)、 及び M M P (配列番号 1 3 1) からなる群より選択される、
 請求項 3 3 に記載のポリペプチド。

【請求項 3 5】

(a) X 1 - M S M A P - Z 1、 ここで、 X 1 は配列 P E P M P (配列番号 1 3 2) の 0
 、 1、 2、 3、 4、 または 5 アミノ酸であり、 Z 1 は配列 I T G E N (配列番号 1 3 3)
 の 0、 1、 2、 3、 4、 または 5 アミノ酸である ; 50

(b) X3 - MNMAP - Z3、ここで、X3は配列PALMP(配列番号135)の0、1、2、3、4、または5アミノ酸であり、Z3は配列ITGRA(配列番号136)の0、1、2、3、4、または5アミノ酸である；

(c) X5 - MSMQP - Z5、ここで、X5は配列PEPMP(配列番号132)の0、1、2、3、4、または5アミノ酸であり、Z5は配列LKGQS(配列番号138)の0、1、2、3、4、または5アミノ酸である；

(d) X2 - MMP - Z2、ここで、X2は配列TAPLP(配列番号140)の0、1、2、3、4、または5アミノ酸であり、Z2は配列VAEDE(配列番号141)の0、1、2、3、4、または5アミノ酸である；

(e) X4 - MMP - Z4、ここで、X4は配列AAPLP(配列番号143)の0、1、2、3、4、または5アミノ酸であり、Z4は配列LSEEA(配列番号144)の0、1、2、3、4、または5アミノ酸である；及び

(f) X6 - MMP - Z6、ここで、X6は配列TAPIP(配列番号146)の0、1、2、3、4、または5アミノ酸であり、Z6は配列VSQTQ(配列番号147)の0、1、2、3、4、または5アミノ酸である、
 からなる群より選択されるポリペプチド。

【請求項36】

PMPMSMAPIITG(配列番号149)；LMPMNMAPIITG(配列番号150)；PMPMSMQPLKG(配列番号151)；PLPMMPVAE(配列番号152)；PLPMMPLSE(配列番号153)；及びPIPMMPVSQ(配列番号154)からなる群より選択されるポリペプチド。

【請求項37】

一般式IX：

AG(R1)(R2)(配列番号155)

(ここで、R1はA、S及びDからなる群より選択され；かつ

R2はE及びQからなる群より選択される)

の配列からなるポリペプチド。

【請求項38】

ポリペプチドが、AGAE(配列番号156)、AGSE(配列番号157)、AGDE(配列番号158)、及びAGDQ(配列番号159)からなる群より選択される、請求項37に記載のポリペプチド。

【請求項39】

(a) X1 - AGAE - Z1、ここで、X1は配列VMHTS(配列番号160)の0、1、2、3、4、または5アミノ酸であり、Z1は配列GSGQA(配列番号161)の0、1、2、3、4、または5アミノ酸である；

(b) X3 - AGSE - Z3、ここで、X3は配列IMFTS(配列番号163)の0、1、2、3、4、または5アミノ酸であり、Z3は配列GTGQA(配列番号164)の0、1、2、3、4、または5アミノ酸である；

(c) X5 - AGAE - Z5、ここで、X5は配列MMHTS(配列番号166)の0、1、2、3、4、または5アミノ酸であり、Z5は配列GSGQA(配列番号161)の0、1、2、3、4、または5アミノ酸である；

(d) X2 - AGDE - Z2、ここで、X2は配列LMHTA(配列番号168)の0、1、2、3、4、または5アミノ酸であり、Z2は配列GGGQS(配列番号169)の0、1、2、3、4、または5アミノ酸である；

(e) X4 - AGDQ - Z4、ここで、X4は配列LMHTG(配列番号171)の0、1、2、3、4、または5アミノ酸であり、Z4は配列GGGQA(配列番号172)の0、1、2、3、4、または5アミノ酸である；及び

(f) X6 - AGAE - Z6、ここで、X6は配列LMHTA(配列番号168)の0、1、2、3、4、または5アミノ酸であり、Z6は配列GGGQS(配列番号169)の0、1、2、3、4、または5アミノ酸である、

からなる群より選択されるポリペプチド。

【請求項 40】

一般式 X :

EC(R1)G(R2)(R3)GTC(R4)(R5)(R6) (配列番号 175)

(ここで、R1 は H、N、Q、及び S からなる群より選択され；

R2 は G、R、A からなる群から選択されるか、または存在せず；

R3 は R 及び Q からなる群より選択され

R4 は N 及び H からなる群より選択され；

R5 は F 及び Y からなる群より選択され；かつ

R6 は F 及び Y からなる群より選択される)

の少なくとも 5 個の連続するアミノ酸からなるポリペプチド。

10

【請求項 41】

ポリペプチドが一般式 X のアミノ酸配列からなる請求項 40 に記載のポリペプチド。

【請求項 42】

R2 が G、R、A からなる群より選択され；かつ

R4 が H である、

請求項 40 に記載のポリペプチド。

【請求項 43】

ポリペプチドが、ECHGRGTCNYY (配列番号 176)、ECNGGRGTCHY

Y (配列番号 177)、ECQGRQGTCHEFF (配列番号 178)、及び EC SG A

RGTCHYF (配列番号 179) からなる群より選択される、請求項 40 に記載のポリ

ペプチド。

20

【請求項 44】

一般式 XI :

(R1)(R2)T(R3)K (配列番号 180)

(ここで、R1 は P、S、及び A からなる群より選択され；

R2 は S、E、及び D からなる群より選択され；かつ

R3 は L 及び V からなる群より選択される)

のアミノ酸配列からなるポリペプチド。

【請求項 45】

R3 が L である請求項 44 に記載のポリペプチド。

30

【請求項 46】

R2 が D 及び E から選択される請求項 44 に記載のポリペプチド。

【請求項 47】

ポリペプチドが、PSTLK (配列番号 181)、PSTVK (配列番号 182)、SE

TLK (配列番号 183)、ADTLK (配列番号 184)、及び PDTLK (配列番号

185) からなる群より選択される、請求項 44 に記載のポリペプチド。

【請求項 48】

(a) X1 - PSTLK - Z1、ここで、X1 は配列 FKKPT (配列番号 186) の 0

、1、2、3、4、または 5 アミノ酸であり、Z1 は配列 AGE LR (配列番号 187)

の 0、1、2、3、4、または 5 アミノ酸である；

(b) X3 - PSTVK - Z3、ここで、X3 は配列 FRKPI (配列番号 189) の 0

、1、2、3、4、または 5 アミノ酸であり、Z3 は配列 AGE LE (配列番号 190)

の 0、1、2、3、4、または 5 アミノ酸である；

(c) X5 - SETLK - Z5、ここで、X5 は配列 FSKPQ (配列番号 192) の 0

、1、2、3、4、または 5 アミノ酸であり、Z5 は配列 AGDLR (配列番号 193)

の 0、1、2、3、4、または 5 アミノ酸である；

(d) X2 - ADTLK - Z2、ここで、X2 は配列 QGSPS (配列番号 195) の 0

、1、2、3、4、または 5 アミノ酸であり、Z2 は配列 AGLIR (配列番号 196)

の 0、1、2、3、4、または 5 アミノ酸である；

50

(e) X4 - P D T L K - Z4、ここで、X4は配列 S S A P A (配列番号198)の0、1、2、3、4、または5アミノ酸であり、Z4は配列 E S Q A Q (配列番号199)の0、1、2、3、4、または5アミノ酸である；及び

(f) X6 - S E T L K - Z6、ここで、X6は配列 G E L P V (配列番号201)の0、1、2、3、4、または5アミノ酸であり、Z6は配列 A G Q L H (配列番号202)の0、1、2、3、4、または5アミノ酸である、
からなる群より選択されるポリペプチド。

【請求項49】

一般式 X I I :

A (R 1) R N D (配列番号204)

10

(ここで、R1はS、Q、及びRからなる群より選択される)
のアミノ酸配列からなるポリペプチド。

【請求項50】

ポリペプチド配列が、A S R N D (配列番号205)、A Q R N D (配列番号206)、及びA R R N D (配列番号207)からなる群より選択される、請求項49のポリペプチド。

【請求項51】

(a) X1 - A S R N D - Z1、ここで、X1は配列 N V C N F (配列番号208)の0、1、2、3、4、または5アミノ酸であり、Z1は配列 Y S Y W L (配列番号209)の0、1、2、3、4、または5アミノ酸である；

20

(b) X3 - A S R N D - Z3、ここで、X3は配列 D V C N F (配列番号211)の0、1、2、3、4、または5アミノ酸であり、Z3は配列 Y S Y W L (配列番号209)の0、1、2、3、4、または5アミノ酸である；

(c) X2 - A S R N D - Z2、ここで、X2は配列 D V C Y Y (配列番号213)の0、1、2、3、4、または5アミノ酸であり、Z2は配列 K S Y W L (配列番号214)の0、1、2、3、4、または5アミノ酸である；

(d) X4 - A Q R N D - Z4、ここで、X4は配列 Q V C H Y (配列番号216)の0、1、2、3、4、または5アミノ酸であり、Z4は配列 R S Y W L (配列番号217)の0、1、2、3、4、または5アミノ酸である；及び

(e) X6 - A R R N D - Z6、ここで、X6は配列 E V C H Y (配列番号219)の0、1、2、3、4、または5アミノ酸であり、Z6は配列 K S Y W L (配列番号214)の0、1、2、3、4、または5アミノ酸である、
からなる群より選択されるポリペプチド。

30

【請求項52】

一般式 X I I I :

(R 1) (R 2) (R 3) N (R 4) (配列番号221)

(ここで、R1はY及びFからなる群より選択され；

R2はY及びFからなる群より選択され；

R3はA及びSからなる群より選択され；かつ

R4はA、S、及びKからなる群より選択される)

40

のアミノ酸配列からなるポリペプチド。

【請求項53】

ポリペプチド配列が、Y Y A N A (配列番号222)、Y Y S N S (配列番号223)、Y Y A N S (配列番号224)、Y Y A N K (配列番号225)、F F A N K (配列番号226)及びY F A N K (配列番号227)からなる群より選択される、請求項49に記載のポリペプチド。

【請求項54】

(a) X1 - Y Y A N A - Z1、ここで、X1は配列 R G T C N (配列番号228)の0、1、2、3、4、または5アミノ酸であり、Z1は配列 Y S F W L (配列番号229)の0、1、2、3、4、または5アミノ酸である；

50

(b) X3 - YYSNS - Z3、ここで、X3は配列RGTCN(配列番号228)の0、1、2、3、4、または5アミノ酸であり、Z3は配列YSFWL(配列番号229)の0、1、2、3、4、または5アミノ酸である；

(c) X1 - YYANS - Z2、ここで、X1は配列RGTCN(配列番号228)の0、1、2、3、4、または5アミノ酸であり、Z5は配列YSFWL(配列番号229)の0、1、2、3、4、または5アミノ酸である；

(d) X2 - YYANK - Z2、ここで、X2は配列RGTCH(配列番号233)の0、1、2、3、4、または5アミノ酸であり、Z2は配列YSFWL(配列番号229)の0、1、2、3、4、または5アミノ酸である；

(e) X4 - FFANK - Z4、ここで、X4は配列QGTCH(配列番号235)の0、1、2、3、4、または5アミノ酸であり、Z4は配列YSFWL(配列番号229)の0、1、2、3、4、または5アミノ酸である；及び

(f) X6 - YFANK - Z6、ここで、X6は配列RGTCH(配列番号233)の0、1、2、3、4、または5アミノ酸であり、Z6は配列YSFWL(配列番号229)の0、1、2、3、4、または5アミノ酸である、
からなる群より選択されるポリペプチド。

【請求項55】

I E R S E M F K K P T (配列番号238)、L N P E R M F R K P I (配列番号239)、V D V S D M F S K P Q (配列番号240)、I P E Q S F Q G S P S (配列番号241)、V K A D L Q F S S A P A (配列番号242)、及びV E E R Q Q F G E L P V (配列番号243)からなる群より選択されるポリペプチド。

【請求項56】

(a) X1 - I E R S E M F K K P T - Z1、ここで、X1は配列FWLAT(配列番号244)の0、1、2、3、4、または5アミノ酸であり、Z1は配列PSTLK(配列番号181)の0、1、2、3、4、または5アミノ酸である；

(b) X3 - L N P E R M F R K P I - Z3、ここで、X3は配列FWLAS(配列番号246)の0、1、2、3、4、または5アミノ酸であり、Z3は配列PSTVK(配列番号182)の0、1、2、3、4、または5アミノ酸である；

(c) X1 - V D V S D M F S K P Q - Z2、ここで、X1は配列FWLAT(配列番号244)の0、1、2、3、4、または5アミノ酸であり、Z5は配列SETLK(配列番号183)の0、1、2、3、4、または5アミノ酸である；

(d) X2 - I P E Q S F Q G S P S - Z2、ここで、X2は配列FWLTT(配列番号249)の0、1、2、3、4、または5アミノ酸であり、Z2は配列ADTLK(配列番号184)の0、1、2、3、4、または5アミノ酸である；

(e) X4 - V K A D L Q F S S A P A - Z4、ここで、X4は配列FWLTT(配列番号249)の0、1、2、3、4、または5アミノ酸であり、Z4は配列PDTLK(配列番号185)の0、1、2、3、4、または5アミノ酸である；及び

(f) X6 - V E E R Q Q F G E L P V - Z6、ここで、X6は配列FWLTT(配列番号249)の0、1、2、3、4、または5アミノ酸であり、Z6は配列SETLK(配列番号183)の0、1、2、3、4、または5アミノ酸である、
からなる群より選択されるポリペプチド。

【請求項57】

F S T M P F L F C N I N N V C N F A (配列番号253)、F T T M P F L F C N V N D V C N F A (配列番号254)、F S T M P F M F C N I N N V C N F A (配列番号255)、F S T M P F L Y C N P G D V C Y Y A (配列番号256)、F S T L P F A Y C N I H Q V C H Y A (配列番号257)、F S T M P F I Y C N I N E V C H Y A (配列番号258)、P F L F C N I N N V C N F A S R N D (配列番号259)、P F L F C N V N D V C N F A S R N D (配列番号260)、P F M F C N I N N V C N F A S R N D (配列番号261)、P F L Y C N P G D V C Y Y A S R N D (配列番号262)、P F A Y C N I H Q V C H Y A Q R N D (配列番号263)、P F I Y C N I N E V C H

YARRND (配列番号264)、FSTMPFLFCNINNVCNFA SRND (配列番号265)、FTTMPFLFCNVNDVCNFA SRND (配列番号266)、FSTMPFMFCNINNVCNFA SRND (配列番号267)、FSTMPFLYCNPGDVCYYASRND (配列番号268)、FSTLPFAYCNIHQVCHYAQRND (配列番号269)、FSTMPFIYCNINEVCHYARRND (配列番号270)、PFIECHGRGTCNY (配列番号271)、PFLECHGRGTCNY (配列番号272)、PFIECNGGRGTCHYY (配列番号273)、PFLECQGRQGTC HFF (配列番号274)、PFIECSGARGTCHYF (配列番号275)、IERSEMFKKPTPSTL KAG (配列番号276)、LNPERMFRKPI PSTVKAG (配列番号277)、VDVSDMFSK PQSE TLKAG (配列番号278)、IPEQS FQGS PSADTLKAG (配列番号279)、VKADLQFSSAPADTLKES (配列番号280)、VEERQQFGE L P V S E T L K A G (配列番号281)、GSC L R K F S T M (配列番号282)、GSC L Q R F T T M (配列番号283)、GSC L R R F S T M (配列番号284)、GSC L A R F S T M (配列番号285)、GSC L P V F S T L (配列番号286)、GSC L P R F S T M (配列番号287)、L R K F S T M P F L F C N I N N V C N F (配列番号288)、LQRFTTMPFLFCNVNDVCNF (配列番号289)、LRRFSTMPFMFCNINNVCNF (配列番号290)、LARFSTMPFLYCNPGDVCYY (配列番号291)、LPVFSTLPFAYCNIHQVCHY (配列番号292)、LPRFSTMPFIYCNINEVCHY (配列番号293)、IPE (配列番号294)、IER (配列番号295)、QD (配列番号296)、NYY (配列番号297)、FCNVNDVCNF (配列番号298)、YCNPGDVCYY (配列番号299)、YCNHQVCHY (配列番号300)、及びYCNINEVCHY (配列番号301) からなる群より選択されるポリペプチド。

【請求項58】

- (a) 請求項1～57のいずれか1項に記載のポリペプチド；及び
(b) 薬学的に許容し得る担体、
を含む医薬組成物。

【請求項59】

組織において血管新生を阻害するための方法であって、該組織を有効阻害量の請求項1～57のいずれか1項のポリペプチドと接触させることを含む方法。

【請求項60】

組織において血管新生を阻害するための方法であって、該組織を有効阻害量の請求項58の医薬組成物と接触させることを含む方法。

【請求項61】

血管新生が腫瘍誘導性である請求項59または60の方法。

【請求項62】

哺乳動物において血管新生介在疾患または状態を治療するための方法であって、血管新生介在疾患または状態を患う哺乳動物に、血管新生の阻害に有効な量の請求項1～57のいずれか1項のポリペプチドを投与することを含む方法。

【請求項63】

哺乳動物において血管新生介在疾患または状態を治療するための方法であって、血管新生介在疾患または状態を患う哺乳動物に、血管新生の阻害に有効な量の請求項58の医薬組成物を投与することを含む方法。

【請求項64】

血管新生介在疾患または状態が、固形及び血液由来の腫瘍、糖尿病性網膜症、関節リウマチ、網膜新血管新生、脈絡膜新血管新生、黄斑変性症、角膜新血管新生、未熟児の網膜症、角膜移植片拒絶反応、新生血管緑内障、水晶体後線維増殖症、流行性角結膜炎、ビタミンA欠乏症、コンタクトレンズ疲労、アトピー性角膜炎、上方輪部角結膜炎、翼状片乾燥角膜炎、sogrens (シェーグレン症?)、赤鼻、phlyectenulosis

、梅毒、マイコバクテリア感染、脂質変性、化学的熱傷、細菌性潰瘍、真菌性潰瘍、単純ヘルペス感染、带状疱疹感染、原生動物感染、カポジ肉腫、モーレン潰瘍、テリエン周辺変性症、辺縁角質溶解、外傷、全身性狼瘡、多発性動脈炎、ウェゲナー類肉腫症、強膜炎、スティーブン・ジョンソン病、放射状角膜切除術、鎌状赤血球貧血、類肉腫、弾性線維性仮性黄色腫、パジェット病、静脈閉塞症、動脈閉塞症、頸動脈閉塞病、慢性ブドウ膜炎、慢性 *v i t r i t i s*、ライム病、イールズ病、*B e c h e t s* (ペーチェット?) 病、近視、視窩、スタルガルト病、扁平部炎、慢性網膜剥離、過粘稠度症候群、トキシプラズマ症、レーザー後合併症、線維性血管性組織の異常増殖、血管腫、*O s l e r - W e b e r - R e n d u* 病、後天性免疫不全症候群、眼血管新生病、骨関節炎、慢性炎症、クローン病、潰瘍性大腸炎、乾癬、アテローム性動脈硬化、及び類天疱瘡からなる群より選択される、請求項 6 2 または 6 3 に記載の方法。 10

【請求項 6 5】

腫瘍転移を阻害するための方法であって、腫瘍または組織を、腫瘍転移の阻害に有効な量の請求項 1 ~ 5 7 のいずれか 1 項のポリペプチドと接触させることを含む方法。

【請求項 6 6】

腫瘍転移を阻害するための方法であって、腫瘍または組織を、腫瘍転移の阻害に有効な量の請求項 5 8 に記載の医薬組成物と接触させることを含む方法。

【請求項 6 7】

腫瘍増殖を阻害するための方法であって、腫瘍または組織を、腫瘍増殖の阻害に有効な量の請求項 1 ~ 5 7 のいずれか 1 項に記載のポリペプチドと接触させることを含む方法。 20

【請求項 6 8】

腫瘍転移を阻害するための方法であって、腫瘍または組織を、腫瘍増殖の阻害に有効な量の請求項 5 8 に記載の医薬組成物と接触させることを含む方法。

【請求項 6 9】

動物組織における細胞外基質との内皮細胞相互作用を阻害するための方法であって、腫瘍または動物組織を、細胞外基質との内皮細胞相互作用の阻害に有効な量の請求項 1 ~ 5 7 のいずれか 1 項に記載のポリペプチドと接触させることを含む方法。

【請求項 7 0】

動物組織における細胞外基質との内皮細胞相互作用を阻害するための方法であって、腫瘍または動物組織を、細胞外基質との内皮細胞相互作用の阻害に有効な量の請求項 5 8 に記載の医薬組成物と接触させることを含む方法。 30

【請求項 7 1】

細胞における基底膜形成または組織発達を阻害するための方法であって、細胞または組織を、基底膜形成の阻害に有効な量の請求項 1 ~ 5 7 のいずれか 1 項に記載のポリペプチドと接触させることを含む方法。

【請求項 7 2】

細胞における基底膜形成または組織発達を阻害するための方法であって、細胞または組織を、基底膜形成の阻害に有効な量の請求項 5 8 に記載の医薬組成物と接触させることを含む方法。

【請求項 7 3】

I V 型コラーゲンの N C 1 ドメイン六量体の結晶であって、I V 型コラーゲンの $[(1) _ 2 \cdot _ 2] _ 2$ N C 1 六量体を含み、結晶化 N C 1 ドメイン六量体の三次元構造を 3 以下の解像度まで決定できるようにおおよそ $a = 1 2 7 . 1 6 \sim 1 2 9 . 4 1$ 、 $b = 1 3 9 . 5 7 \sim 1 4 3 . 8 7$; $c = 1 6 0 . 2 0 \sim 1 6 2 . 9 2$; $\beta = 9 1 . 3 ^ \circ$ を有する空間群 $P 2 _ 1$ からなる結晶。 40

【請求項 7 4】

結晶化 N C 1 ドメイン六量体の三次元構造を 2 . 2 以下の解像度まで決定することができる、請求項 7 3 に記載の結晶。

【請求項 7 5】

結晶化 N C 1 ドメイン六量体の三次元構造を 2 以下の解像度まで決定することができる 50

、請求項 7 3 に記載の結晶。

【請求項 7 6】

I V 型コラーゲン構築の阻害剤を同定するための方法であって：

(a) I V 型コラーゲンの N C 1 六量体の結晶であって、I V 型コラーゲンの [(1)
2 · 2]₂ N C 1 六量体を含み、結晶化 N C 1 ドメイン六量体の三次元構造を 3 以下の
解像度まで決定できるように、およそ a = 1 2 7 . 1 6 ~ 1 2 9 . 4 1 、 b = 1 3
9 . 5 7 ~ 1 4 3 . 8 7 ; c = 1 6 0 . 2 0 ~ 1 6 2 . 9 2 ; $\beta = 9 1 . 3 ^{\circ}$ を
有する空間群 P 2₁ からなる結晶を得、

(b) 請求項 7 5 の I V 型コラーゲンの結晶化 N C 1 ドメイン六量体の三次元構造を解析
し；及び

(b) I V 型コラーゲン N C 1 鎖の：

(i) 鎖間ドメインスワッピング領域；

(i i) 鎖内ドメインスワッピング領域；

(i i i) 特異性領域；

(i v) 特異性領域パートナー；

(v) 六量体界面；

(v i) 単量体 - 単量体界面；及び

(v i i) 超可変領域

からなる群より選択される 1 つ以上の領域を標的として、I V 型コラーゲン構築の潜在的
阻害剤を設計する、

ことを含む方法。

【請求項 7 7】

(a) 潜在的阻害剤を合成し；及び

(b) 潜在的阻害剤が I V 型コラーゲンの構築を阻害するかどうかを決定する、
ことをさらに含む、請求項 7 6 に記載の方法。

【請求項 7 8】

(a) 潜在的阻害剤を合成し；及び

(b) 潜在的阻害剤が、血管新生、腫瘍増殖、腫瘍転移、内皮細胞接着、内皮細胞増殖、
及び基底層構築のうち、1 つ以上を阻害するかどうかを決定するアッセイを実施する、
ことをさらに含む、請求項 7 6 に記載の方法。

【請求項 7 9】

請求項 7 6 ~ 7 8 のいずれか 1 項の方法によって同定される、I V 型コラーゲン構築の阻
害剤。

【請求項 8 0】

請求項 7 6 ~ 7 8 のいずれか 1 項の方法によって同定される、血管新生、腫瘍増殖、腫瘍
転移、内皮細胞接着、内皮細胞増殖、及び基底層構築からなる群より選択される 1 つ以上
のプロセスの阻害剤。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

(相互参照)

本願は 2 0 0 1 年 7 月 2 7 日出願の米国仮出願 6 0 / 3 0 8 , 5 2 3 号；2 0 0 1 年 1 0
月 2 9 日出願の 6 0 / 3 5 1 , 2 8 9 号；2 0 0 2 年 3 月 2 2 日出願の 6 0 / 3 6 6 , 8 5
4 号；及び 2 0 0 2 年 6 月 3 日出願の 6 0 / 3 8 5 , 3 6 2 号に対する優先権を主張する
。

【0 0 0 2】

(政府の利益の記載)

この研究は国立衛生研究所からの認可 D K 1 8 3 8 1 及び D K 5 3 7 6 3 によって支持さ
れており、そしてひいては、米国政府は本発明において特定の権利を有し得る。

【0 0 0 3】

10

20

30

40

50

(発明の分野)

本発明は結晶学、分子生物学、蛋白質化学、血管新生、腫瘍増殖及び転移、並びに基底膜構築に関する。

【背景技術】

【0004】

基底膜(基底層)はシート様細胞外基質(ECM)であり、これは全ての組織の基本成分である。基底層は組織の区画化に対する備えを有し、組織区画間を移動する物質のフィルターとして作用する。典型的には、基底層は、血管及び毛細管を含む、動物の全ての組織における上皮または内皮と密接に関連して見出される。基底層区画は、細胞によって選択された後、自己構築して複雑な細胞外ネットワークを形成する。生物学的に活性の基底層の形成は関連細胞の発達及び分化に重要である。

10

【0005】

IV型コラーゲンは基底膜の主要構成成分であることが示されており、1(IV)から6(IV)と呼ばれる6本の相同鎖のファミリーからなる。各々の鎖はカルボキシ末端の非コラーゲン性(NC1)ドメイン;中間領域の長いらせん状コラーゲン性ドメイン;及びアミノ末端の7Sコラーゲン性ドメインを特徴とする(Martinら、1988年、Adv. Protein Chem. 39:1~50;Gunwarら、1991年、J. Biol. Chem. 266:14088~14094)。3本の鎖が三重らせん分子、「ヘテロ三量体」を構築する。このヘテロ三量体は、内質の内腔においてひとたび形成されると、細胞外間隙に分泌され、そこで2つのそのようなヘテロ三量体がC末端相互作用によって六量体を構築し、次いでN末端の会合によって超分子ネットワークを構築する。NC1ドメインは、六量体の構築を導くC末端二量体会合を決定することにより、この構築において主要な役割を果たす。

20

【0006】

鎖の構成、そしてひいては、IV型コラーゲンネットワークの特性は2つの因子の影響を受ける。第1に、ネットワークの鎖の組成は鎖の利用可能性によって制限される:6本の鎖は組織特異的発現パターンを示し、1及び2鎖は遍在性であり、3-6鎖はより制限された組織分布を有する。第2に、NC1ドメインはネットワークの鎖特異的構築に特異性を付与する。従って、鎖の分泌を三重らせんプロモーターの形成に方向付け、かつ三重らせんプロモーターを六量体の形成に、従って、コラーゲンネットワークの形成に方向付ける、今のところ同定されていない認識配列がNC1ドメイン内に存在しなければならない。本質的かつ鎖化学量論的に異なる、多くのIV型コラーゲン六量体が理論的に可能ではあるが、3つのみのが同定されている:[1₂ 2]₂、[3 4 5]₂、及び[(1₂ 2)(5₂ 6)]。

30

【0007】

血管新生、新しい血管の形成プロセス、は生理学的プロセス、例えば、胚性及び出生後の発達に加えて、外傷修復において重要な役割を果たす。血管の形成は、炎症(例えば、糖尿病性網膜症及び関節炎)または腫瘍形成(例えば、癌)を含む病理学的プロセスによっても誘導され得る(Folkman、1985年、Perspect. Biol. Med., 29, 10)。血管新生は、腫瘍もしくは正常細胞が分泌する血管新生成長因子の他に、細胞外基質の組成及び内皮酵素の活性によって調節される(NicosiaとOttinetti、1990年、Lab. Invest., 63, 115)。

40

【0008】

全ての固形腫瘍増殖の共通の特徴は血液供給の必要性である。従って、多くの研究室が成長因子及びそれらの受容体に基づく抗血管新生性化合物の開発に焦点を当てている。このアプローチは幾らかの成功を導いてはいるが、血管新生の役割を果たすことが知られる成長因子の数は多い。従って、おそらくは腫瘍及び関連炎症性細胞が血管新生を誘導し得る様々な因子を産生するため、成長因子アンタゴニストが癌の治療において限られた用途のみを有し得る可能性が存在する。

【0009】

50

これに関して、血管新生に共通の特徴、例えば、細胞外基質（ECM）への内皮細胞接着を標的とする戦略がヒトにおける腫瘍増殖に対して強力な生理学的衝撃を有するものと予想される。この概念は、特異的ECM細胞接着受容体、例えば、 α_3 及び α_5 インテグリンのアンタゴニストが血管新生を遮断するという事実によって支持される。さらに、 α_3 インテグリンはサイトカイン活性化内皮及び平滑筋細胞上に最も顕著に発現し、血管新生に必要であることが示されている。（Varnerら、Cell Adhesion and Communication 3: 367~374（1995年）；Brooksら、Science 264: 569~571（1994年））。これらの知見に基づくと、抗血管新生治療に対する潜在的に強力な新規アプローチは、個々のECM成分内の重要な調節ドメインを特異的に標的とすることである。

10

【0010】

特定のIV型コラーゲン（IV）NC1ドメインが血管新生、腫瘍増殖、腫瘍転移、基底膜への細胞結合、及びIV型コラーゲン分子の構築の有効な阻害剤であることが示されている（例えば、米国特許第5,691,182号；第5,856,184号；第6,361,994号；及び第6,358,735号を参照）。上記にもかかわらず、これらのプロセスを阻害することが可能な更なる化合物を同定することは当該技術分野に対する重要な価値を有する。

【発明の開示】**【発明が解決しようとする課題】****【0011】**

従って、IV型コラーゲンNC1ドメインの結晶構造を推定する方法、及びIV型コラーゲンヘテロ三量体及び/またはIV型コラーゲン六量体の構築を阻害する化合物の設計にこの構造を用いる方法を提供することが非常に望ましい。

20

【課題を解決するための手段】**【0012】**

一側面において、本発明はIV型コラーゲンの結晶化NC1ドメイン六量体、及びその結晶の作製方法を提供し、ここで、NC1ドメイン六量体は、その結晶化NC1ドメイン六量体の三次元構造を少なくとも3以下の解像度で決定することができるように結晶化される。

【0013】

別の側面においては、本発明は、血管新生、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、及び/または基底層構築を阻害するように化合物を設計するための方法であって、本発明の方法によって生成された結晶化IV型コラーゲンNC1ドメイン六量体の三次元構造を解析し、並びに該解析によってIV型コラーゲンヘテロ三量体及び六量体構築に重要であるものと同定されているNC1ドメインの領域を標的とする化合物を同定及び合成することを含む方法を提供する。そのような化合物は血管新生、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、並びに基底膜構築を阻害するのに用いることができる。

30

【0014】

別の側面においては、本発明は、ここに開示されるIV型コラーゲンNC1六量体構造の解析に基づき、本発明の合理的薬物設計方法によって設計された新規ポリペプチドを提供する。結晶構造から入手可能な情報の結果として、IV型コラーゲンヘテロ三量体及び/または六量体の構築に重要である個々のNC1ドメイン配列を推定することが可能である。従って、それらの相互作用を妨害する治療用ポリペプチドを設計し、IV型コラーゲンヘテロ三量体及び/またはIV型コラーゲン六量体の構築を阻害することが可能である。そのような治療用ポリペプチドはIV型コラーゲンの構築を阻害または中断させるのに用いることができ、そしてひいては、血管新生、血管新生介在疾患、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、並びに基底層構築の阻害に有用である。

40

【発明を実施するための最良の形態】**【0015】**

50

本願では、他に述べられない限り、用いられる技術はいくつかの公知の参考文献、例えば：Molecular Cloning: A Laboratory Manual (Sambrookら、1989年、Cold Spring Harbor Laboratory Press)、Gene Expression Technology (Methods in Enzymology, Vol. 185, D. Goeddel編、1991年、Academic Press, San Diego, カリフォルニア州)、Methods in Enzymology (M. P. Deutscher, 編、(1990年) Academic Press, Inc.) 中の「蛋白質精製の手引き」；PCR Protocols: A Guide to Methods and Applications (Innisら、1990年、Academic Press, San Diego, カリフォルニア州)、Culture of Animal Cells: A Manual of Basic Technique, 第二版 (R. I. Freshney, 1987, Liss, Inc. New York, ニューヨーク州)、及び Gene Transfer and Expression Protocols, pp. 109~128, E. J. Murray 編, The Humana Press Inc., Clifton, ニュージャージー州) のいずれかにおいて見出すことができる。

【0016】

IV型コラーゲンは細胞の内側でヘテロ三量体として合成及び構築され、これは次に細胞外に分泌され、そこで六量体構築体、続いて基底膜(基底層)構築が生じる。

【0017】

本研究はIV型コラーゲン[(1)₂(2)]₂NC1六量体の構造を解明している。この構造の知識は、IV型コラーゲンヘテロ三量体及び六量体の構築を阻害することができる化合物の設計における有用性を有し、そしてひいては、血管新生、血管新生介在疾患、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、並びに基底層構築の阻害において有益である。

【0018】

本発明によって提供されるIV型コラーゲン[(1)₂(2)]₂NC1六量体構造の構造の知識も、IV型コラーゲン構築体の理解、そしてひいては、基底層/基底膜構造及び機能全般の理解をも高めるツール及び試薬を提供することによってヘテロ三量体及び六量体構築を促進する化合物の設計における有用性を有する。

【0019】

一側面において、本発明は、結晶化IV型コラーゲン[(1)₂(2)]₂NC1六量体の三次元構造を3.0以下、好ましくは2.2以下、最も好ましくは2.0以下の解像度まで決定することができるように、単離及び精製されたIV型コラーゲン[(1)₂(2)]₂NC1ドメイン六量体(「六量体」)の三次元構造に向けられ、ここで、これらの結晶は、室温でおおよそa = 129.41 ; おおよそb = 143.87 ; おおよそc = 162.92 ; 及びおおよそ = 91.3°であり、かつ非対称単位内に4つの六量体を有する空間群P2₁のものである。その代わりに、これらの結晶はおおよそa = 127.16 ; おおよそb = 139.57 ; おおよそc = 160.20 ; 及びおおよそ = 91.3°並びに非対称単位内に4つの六量体を有する。更なる代替においては、これらの結晶は室温でおおよそa = 79.79 ; おおよそb = 137.20 ; おおよそc = 126.69 ; 及びおおよそ = 90.3°並びに非対称単位内に2つの六量体を有する。

【0020】

別の側面において、本発明は、IV型コラーゲン[(1)₂(2)]₂NC1六量体を約3.0以下、好ましくは2.2以下、最も好ましくは2.0以下の解像度まで結晶化するための方法であって、IV型コラーゲン[(1)₂(2)]₂NC1六量体を0.5mg/mlから約50mg/ml、より好ましくは約1mg/mlから約15mg/ml、最も好ましくは約10mg/mlの濃度で提示し、結晶化を4から32

、より好ましくは10 から26 、さらにより好ましくは約16°から24 、さらにより好ましくは20 で行い、それにより空間群 $P2_1$ の結晶を得ることを含む方法を提供する。これらの結晶は、室温でおおよそ $a = 129.41$; おおよそ $b = 143.87$; おおよそ $c = 162.92$; 及びおおよそ $\beta = 91.3^\circ$ を有し、かつ非対称単位内に4つの六量体を有する。その代わりに、これらの結晶の低温冷却により、おおよそ $a = 127.16$; おおよそ $b = 139.57$; おおよそ $c = 160.20$; 及びおおよそ $\beta = 91.3^\circ$ を有し、かつ非対称単位内に4つの六量体を有する結晶が生じ得る。更なる代替においては、これらの結晶は室温でおおよそ $a = 79.79$; おおよそ $b = 137.20$; おおよそ $c = 126.69$; 及びおおよそ $\beta = 90.3^\circ$ を有し、かつ非対称単位内に2つの六量体を有し得る。

10

【0021】

結晶化は、一実施形態においては、10% (w/v)以上のPEG20Kでの懸滴及び蒸気拡散法を用いて生じ得る。また代替法として、他の結晶化方法を用いることができる。例えば、温度変化を結晶の生成に用いることができ、または宇宙空間での結晶化を解像度の改善に用いることができる。結晶化は、別の実施形態においては、20%以上のPEG3350で生じ得る。加えて、PEG20Kまたは3350の代わりに他の化学物質を用いることができる。例えば、有機化学物質(例えば、イソプロパノール)、無機化学物質(例えば $(NH_4)_2SO_4$ 、 NaH_2PO_4)、及び他の分子量のPEGを用いることができる。この方法の更なる詳細は以下に説明される通りである。

【0022】

更なる側面において、本発明は、結晶化IV型コラーゲン $[(1)_2(2)]_2NC1$ 六量体の三次元構造を決定するための方法であって、IV型コラーゲン $[(1)_2(2)]_2NC1$ 六量体を上述の通りに結晶化する工程、及び続いて、IV型コラーゲン $[(1)_2(2)]_2NC1$ 六量体を解析してその三次元構造を決定する工程を含む方法を提供する。好ましい実施形態において、解析はX線回折による。回折解析から生成されたデータの組はあらゆる適切なソフトウェアを用いて解析することができ、これにはHKL2000スイツ(39)のDENZO及びSCALEPACKプログラム、SOLVEプログラム(40)、RESOLVE(41)プログラム、及び/またはCCP4スイツ(42)のFFTプログラムが含まれるが、これらに限定されるものではない。得られる解析からのポリペプチドの追跡はあらゆる適切なソフトウェアを用いて達成することができ、これにはTOM FRODOグラフィックスプログラム(43)が含まれるがこれに限定されるものではない。最終構造解析はあらゆる適切なソフトウェアを用いて達成することができ、これにはSETOR(45)、GRASP(46)、及びSURFNET(47)グラフィックスソフトウェアパッケージ、CCP4スイツ内の様々なユーティリティ・プログラム、及びHBPLUS(48)及び蛋白質-蛋白質相互作用ウェブサーバー(<http://www.biochem.ucl.ac.uk/bism/PP/server/>)が含まれるがこれらに限定されるものではない。

20

30

【0023】

IV型コラーゲン $[(1)_2(2)]_2$ 六量体の三次元構造を解析することにより、当該技術分野における熟練者は、以下に説明されるように、IV型コラーゲンNC1ドメインヘテロ三量体及び六量体構築に重量な部位を決定することができる。

40

【0024】

本発明の別の側面は、IV型コラーゲン $[(1)_2(2)]_2$ 六量体の三次元構造を異なるIV型コラーゲンNC1ドメイン六量体結晶、または変異体の結晶、IV型コラーゲンNC1ドメイン六量体の類似体もしくは共複合体の三次元構造の解析に用いることである。

【0025】

本発明の更なる側面は、IV型コラーゲン $[(1)_2(2)]_2$ 六量体の三次元構造を、IV型コラーゲン $[(1)_2(2)]_2NC1$ 六量体を含むIV型コラーゲンのヘテロ三量体及び六量体の構築の阻害剤の設計に用いることである。これらの阻害剤は、

50

望ましくない血管新生、血管新生介在疾患、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、並びに基底層構築を阻害する治療薬として用いることができる。この実施形態は：

(a) 結晶化 NC1 ドメイン六量体の三次元構造を 3 Å 以下の解像度まで決定できるように I V 型コラーゲンの NC1 六量体の結晶を得、ここで、該結晶は $[(1)_{2 \cdot 2}]_2$ NC1 六量体を含み、該結晶はおおよそ $a = 127.16$ Å から 129.41 Å、 $b = 139.57$ Å から 143.87 Å； $c = 160.20$ Å から 162.92 Å； $\beta = 91.3^\circ$ を有する空間群 $P2_1$ からなり；

(b) I V 型コラーゲンの結晶化 NC1 ドメイン六量体の三次元構造を解析し；並びに

(c) I V 型コラーゲン NC1 鎖の：

(i) 鎖間ドメインスワッピング領域；

(ii) 鎖内ドメインスワッピング領域；

(iii) 特異性領域；

(iv) 特異性領域パートナー；

(v) 六量体界面；

(vi) 単量体 - 単量体界面；及び

(vii) 超可変領域；

からなる群より選択される 1 つ以上の領域を標的として、I V 型コラーゲン構築の潜在的阻害剤を設計する、

ことを含む。

【0026】

ここで用いられる場合、「標的」または「ターゲティング」は、共有結合的または非共有結合的手段により、この領域と相互作用する化合物を指す。様々な領域の定義は以下で論じられる。

【0027】

上で論じられるように、NC1 ドメインは I V 型コラーゲン鎖構築の選択プロセスを駆動し、従って、NC1 ドメイン構築の解析は I V 型コラーゲン構築と相関する。さらに、異なる NC1 ドメインの高い相同性の程度を考えると、 $[(1)_2(2)]_2$ NC1 六量体結晶構造の解析は、他の六量体型の構造の他にそのような構築体の阻害剤についての洞察をもたらす。

【0028】

ここで用いられる場合、「I V 型コラーゲンのヘテロ三量体及び六量体の構築体の阻害」は、そのようなヘテロ三量体及び/または六量体の初期構築体を阻害すること、または I V 型コラーゲン NC1 ドメインの既に構築されたヘテロ三量体及び六量体を破壊することを意味する。非常に好ましい実施形態においては、ここで同定された治療用化合物はそのような I V 型コラーゲン NC1 ドメインのヘテロ三量体及び/または六量体の初期構築を阻害する。

【0029】

阻害剤はペプチド、または I V 型コラーゲンヘテロ三量体及び/または六量体構築を妨害することが予想される重要な領域から誘導されるペプチドに対する抗体を含み得る。その代わりに、そのような構築を阻害するそれらの潜在能力に基づいて同定される小分子。巨大な、構造的に多様な化合物のライブラリ、例えば、Available Chemical Directory (ACD) の電子スクリーニングで、同定された重要領域と相互作用することが予想されるそのような分子の新規構造クラスを同定することができる。加えて、I V 型コラーゲン $[(1)_2(2)]_2$ NC1 六量体構造の知識が、化合物の「新規設計」があらゆる I V 型コラーゲン NC1 ドメインヘテロ三量体及び/または六量体の構築を阻害することを可能にする。

【0030】

潜在的阻害剤は、GRAM、DOCK、または AUTODOCK [Dunbrackら、1997年、前出] のようなドッキング・プログラムを用いて、コンピュータ・モデリン

10

20

30

40

50

グを用いることによりコンピュータ・シミュレーション上で試験することができる。これらの手順は、候補化合物をIV型コラーゲン[(1)₂(2)]₂NC1六量体にコンピュータ・フィッティングさせ、候補化合物の形状及び化学構造が以下にしてIV型コラーゲンヘテロ三量体及び/または六量体の構築を阻害するのかを推定することを含む。コンピュータ・プログラムは、IV型コラーゲン[(1)₂(2)]₂六量体上の関連結合部位に対する候補化合物の誘引、相反、及び立体的妨害の推定に用いることもできる。一般には、フィットが緊密である(例えば、立体妨害が小さい、及び/または誘引力が大きい)ほどその候補化合物は強力であり、その候補化合物が他の蛋白質との望ましくない相互作用のために重大な副作用を誘導することはほとんどありそうもない。

【0031】

潜在的な小分子阻害剤は、例えば、例えば組換えバクテリオファージ(ScottとSmith, Science, 249:386~390(1990年); Cwirllaら, Proc. Natl. Acad. Sci., 87:6378~6382(1990年); Devlinら, Science, 249:404~406(1990年))において生成された無作為ペプチドライブラリ、または組合せ化学ライブラリをスクリーニングすることによって得ることができる。このようにして選択された候補化合物は、1つ以上の見込みのある候補化合物が同定されるまで、コンピュータ・モデリング・プログラムによって系統的に修飾することができる。そのような解析は、例えばHIVプロテアーゼ阻害剤の開発(Lamら, Science 263:380~384(1994年); Wlodawerら, Ann. Rev. Biochem. 62:543~585(1993年); Appel, Perspectives in Drug Discovery and Design 1:23~48(1993年); Erickson, Perspectives in Drug Discovery and Design 1:109~128(1993年))において有効であることが示されている。

【0032】

そのようなコンピュータ・モデリングは、作製されうる本質的に無作為な無数の化学修飾とは対照的に、有限数の合理的な化学修飾の選択を可能にする。従って、ここに開示される三次元構造をコンピュータ・モデリングと共に用いることで、コンピュータ・シミュレーション上での迅速なスクリーニングが可能となり、これはスクリーニング速度及び効率を劇的に高める。

【0033】

ひとたびそのような候補化合物が同定されると、以下に論じられるように、それらを化学的に合成し、かつそれらの生物活性をアッセイする。活性を示す化合物については、それらをIV型コラーゲン[(1)₂(2)]₂NC1六量体結晶と、その化合物と結晶構造との相互作用をマッピングする更なるX線回折解析のため、複合体化することができる。この補足結晶の三次元構造はMolecular Replacement Analysisによって決定することができ、これは既知三次元構造を緊密に関連する分子または新規結晶形態にある蛋白質-リガンド複合体の構造を決定する探索モデルとして用いることを含む。新規結晶の測定されたX線回折特性をこの探索モデル構造と比較し、新規結晶における蛋白質の位置及び方向を算出する。このアプローチを用いることで、ここに開示されるIV型コラーゲン[(1)₂(2)]₂NC1六量体の構造をあらゆるそのようなIV型コラーゲン六量体または共複合体の三次元構造の解析に用いることができる。

【0034】

機能的アッセイ

IV型コラーゲンヘテロ三量体及び/または六量体のインビトロまたはインビボ構築に対する候補化合物の効果を試験するのに用いることができるあらゆるアッセイを本発明によって同定された候補化合物の効力の検証に用いることができる。さらに、血管新生、腫瘍増殖、腫瘍転移、及び内皮細胞接着及び/または運動性に対する候補化合物の効果を試験することができるあらゆるアッセイをそれらの阻害活性の検証に用いることができる。そ

10

20

30

40

50

のようなアッセイには、これらに限定されるものではないが、以下のものが含まれる。

【0035】

構築アッセイ

一例において、用いられる方法はBoutaudら、JBC 275(39):30716~30724(2000年)に記載される通りである。天然GBM六量体を標準法によって単離し、トリス塩基を用いてpH3.0で緩衝させた50mMギ酸の溶液中に希釈(<50µg/ml)することによって解離させる。これらの条件下において、HPLCまたはFPLCゲル濾過によって検証できるように、NC1単量体及び二量体への完全な解離が生じる。緩衝液からの塩の不在が完全な六量体解離に最適である。解離したNC1ドメインの構築は、反復希釈-濃縮サイクルにより緩衝液をトリス緩衝性食塩水(50mM トリス、pH7.4、150mM NaCl)に変更することによって行う。NC1ドメインを約1mg/mlの濃度、室温で24時間、所望の濃度の候補化合物の存在または不在下においてインキュベートした後、ゲル濾過クロマトグラフィーを用いて反応生成物をそれらの分子量に従って分離する。混合物中の様々な種の相対量の定量をHPLCプロフィールからのピーク面積解析によって行う。精製 1 - 6 NC1ドメインからの六量体構築を同様に行う。

10

【0036】

全ての実験において、会合混合物中のNC1ドメインの比は好ましくは1:1に維持される。単離されたNC1六量体は、次に、免疫沈殿とそれに続くウェスタンブロッティングにより組成について;電子顕微鏡検査により全体的な外見(サイズ及び形状)について;及び沈降平衡超遠心により分子量について解析することができる。

20

【0037】

血管新生に対するインビトロ効果

改変を加えて、NicosiaとOttinetti、(1990年, Lab. Invest., 63, 115)及びNicosiaら、(1994年, Exp. Biology, 164, 197~206)の手順を、インビトロ条件下において血管新生に対する薬物候補の効果を試験するように設計された実験に用いる。このモデルは血管新生応答に対する成長因子及び細胞外基質分子の効果を研究するのに用いられており、大動脈環を三次元コラーゲンゲル中で血清非含有条件下において用いる。

【0038】

1から3月齢Swiss Webster雄マウスを用いて実験を行う。麻酔の後、胸部大動脈を無菌条件下で切除し、抗生物質を含有するMCDB131無菌成長培地(Clonetics, San Diego, カリフォルニア州)に移す。動脈から脂肪を切り取り、約6から8の1mm胸部セグメントを各々の検体から得る。セグメントを48ウェル組織培養プレートに移す。動脈セグメントの移送前に、これらのプレートのウェルを100マイクロリットルのMatrigel(商標)(EHS基底膜, Collaborative Biomedical Products, Bedford, マサチューセッツ州)で層状化する。Matrigel(商標)は使用前にMCDB131成長培地で1:1に希釈する。セグメントをウェルの中央に配置した後、さらに100マイクロリットルのMatrigel(商標)を検体に被せる。従って、動脈セグメントは基底膜基質中に埋め込まれる。次に、各々のウェルに300µリットルのMCDB131成長培地を入れる。これらのプレートを37、5%CO₂に維持されたインキュベーターに入れる。検体を7日にわたって毎日観察する。新たに成長する微小血管を、培養期間中の様々な時間に、逆相顕微鏡を用いてカウントする。血管新生に対する薬物候補の効果について試験するため、薬物候補をMatrigel(商標)及びMCDB131成長培地と混合し、培養組織から基質への微小血管の成長を分析する。

30

40

【0039】

皮下フィブリン移植片血管新生

フィブリン移植片が外科的に皮下に配置されたラットに薬物候補を静脈注射する、Dvorakら、(Lab. Invest., 57(6):673~686(1987年))によ

50

って記載された方法の改変バージョン。例えば、ラットに対照または様々な濃度の薬物候補を尾静脈注射する。次に、適切な時間に移植片を取り出し、倒立顕微鏡を用いて直接分析する。この分析は、対照及び実験群においてフィブリン内部に成長した移植片当たりの血管数をカウントすることを含んでいた。

【0040】

鶏胚CAM血管新生アッセイ

記載されるように (Brooksら、Cell 92:391~400 (1998年))、bFGFで10日齢鶏胚のCAMに血管新生を誘導する。24時間後、その胚を、総容量100 μ lの無菌リン酸緩衝生理食塩水 (PBS) 中の様々な濃度の薬物候補で全体的に処理する。2日後、胚を犠牲にし、フィルターディスク及びCAM組織を取り出す。フィルターディスクの限定面積内での血管原性血管分岐点の数をカウントすることにより、血管新生を定量する。血管新生指数は、実験的処理からの分岐点の数マイナス対照処理と定義される。

10

【0041】

鶏胚腫瘍増殖アッセイ

簡潔に述べると、異なる腫瘍型の単一細胞懸濁液を10日齢鶏胚のCAMに適用する。これらの腫瘍には、例えば、CS-1メラノーマ細胞、HT1080ヒト繊維肉腫細胞、及びHeP-3ヒト類表皮カルチノーマ細胞が含まれ得る。24時間後、これらの胚に様々な濃度の薬物候補を全体的に注射する。胚を合計で7日間インキュベートし、その時間でそれらを犠牲にする。生じた腫瘍を切除し、決定された湿重量を対照と比較する。

20

【0042】

固定化NC1ドメインはヒト内皮細胞の接着を支持する

新たな血管を形成するため、内皮細胞はECMに接着し、それを通して移動する能力を有していなければならない。さらに、この内皮細胞-ECM相互作用は新たな血管の形成に必要なシグナル伝達を促進し得る。従って、内皮細胞接着を支持する薬物候補の能力を評価することができる。

【0043】

マイクロタイプレートを変量の薬物候補でコートした後、1%ウシ血清アルブミン (BSA) と共にインキュベートして非特異的相互作用を遮断する。次に内皮細胞、例えば、ヒトECV304細胞を固定化ポリペプチドに可変時間接着させる。非接着細胞を洗浄によって除去し、接着細胞から溶出されるクリスタルバイオレットの光学密度を測定することにより接着細胞を定量する。

30

【0044】

インビトロ内皮細胞移動

浸潤性細胞プロセス、例えば、血管新生及び腫瘍転移も細胞の移動を必要とする。従って、ヒト内皮細胞の移動を支持する薬物候補の能力をインビトロで試験することができる。これらの実験は、本質的には、Brooksら、J. Clin. Invest. 99:1390~1398 (1997年) における方法に従って行う。

【0045】

インビボ内皮細胞移動

ヒト内皮細胞の移動を支持する薬物候補の能力をインビボで試験することができる。例えば、肺腫瘍の転移及び血管新生の両者の良好なモデルであると考えられる転移性Lewis肺マウス腫瘍モデルにおいて、標準プロトコルを用いて、薬物候補を試験することができる (例えば、Teicherら、Anticancer Res. 18:2567~2573 (1998年); Guibaudら、Anticancer Drugs 8:276~282 (1997年); Andersonら、Cancer Res. 56:715~718 (1996年) を参照)。

40

【0046】

腫瘍接種の1日後に開始して、2日毎に1回、所望の投与回数だけ薬物候補を静脈内投与する。全ての動物を、その研究を通して毎週2回秤量する。最終処置の1日後に開始して

50

、各々の対照群から1匹以上のマウスを周期的に犠牲にして肺腫瘍負荷を測定する。対照動物の肺が意味のある評価をもたらすのに十分な腫瘍質量を有するとき、この実験を終了する。その時点で、残りの全ての動物の肺を切除して秤量し、直径が2mmを上回る腫瘍病巣の数をカウントする。

【0047】

別の側面においては、本発明は、上述の方法のいずれかによって同定されるIV型コラーゲン構築の阻害剤を提供する。

【0048】

別の側面においては、本発明は、上述の方法のいずれかによって同定される、血管新生、腫瘍増殖、腫瘍転移、内皮細胞接着、内皮細胞増殖、及び基底層構築からなる群より選択される、1つ以上のプロセスの阻害剤を提供する。

10

【0049】

別の側面においては、本発明は、IV型コラーゲン構築の阻害または崩壊に用いることができ、従って、血管新生、血管新生介在疾患、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、並びに基底層構築の、阻害に有用である新規ポリペプチドを提供する。

【0050】

「ポリペプチド」という用語は、2つ以上のサブユニット・アミノ酸、アミノ酸類似体、またはペプチド模倣体の化合物を指すようにその最も広い意味で用いられる。これらのサブユニットはペプチド結合によって連結する。ここに記載されるポリペプチドは化学的に合成されるものであっても、組換えで発現されるものであってもよい。

20

【0051】

好ましくは、本発明のポリペプチドは化学的に合成される。固相、液相、もしくはペプチド縮合技術、またはそれらのあらゆる組合せの公知技術を用いて調製される合成ポリペプチドは、天然または非天然アミノ酸を含むことができる。ペプチド合成に用いられるアミノ酸は、Merrifield(1963年、J. Am. Chem. Soc. 85: 2149~2154)の元の固相手順の標準脱保護、中和、カップリング及び洗浄プロトコルを伴う標準Boc(N-アミノ保護N-t-ブチルオキシカルボニル)アミノ酸樹脂であっても、またはCarpinoとHan(1972年、J. Org. Chem. 37: 3403~3409)により最初に記載された塩基不安定性N-アミノ保護9-フルオレニルメトキシカルボニル(Fmoc)であってもよい。Fmoc及びBoc-N-アミノ保護アミノ酸の両者とも、Sigma、Cambridge Research Biochemicalまたは当該技術分野における熟練者に知られた他の化学企業から入手することができる。加えて、これらのポリペプチドは、当該技術分野における熟練者に馴染みである他のN-保護基を用いて合成することができる。

30

【0052】

固相ペプチド合成は当該技術分野における者に馴染みであり、かつ、例えば、StewartとYoung、1984年、Solid Phase Synthesis, 第二版 Pierce Chemical Co., Rockford, Ill; FieldsとNoble、1990年、Int. J. Pept. Protein Res. 35: 161~214において提供される技術によって、または自動合成機を用いて達成することができる。本発明のポリペプチドは、D-アミノ酸(これは、インビボでL-アミノ酸特異的プロテアーゼに対して耐性である)、D-及びL-アミノ酸の組合せ、並びに特別な特性を搬送する様々な「デザイナー」アミノ酸(例えば、 α -メチルアミノ酸、 β -メチルアミノ酸、及びN-メチルアミノ酸等)を含むことができる。合成アミノ酸は、リジンの代わりにオルニチン、フェニルアラニンの代わりにフルオロフェニルアラニン、及びロイシンまたはイソロイシンの代わりにノロロイシンを含む。

40

【0053】

加えて、これらのポリペプチドは、新規特性を有するペプチドを調製するため、ペプチド模倣結合、例えば、エステル結合を有することができる。例えば、還元されたペプチド結合、すなわち、 $R_1-CH_2-NH-R_2$ (ここで、 R_1 及び R_2 はアミノ酸残基または

50

配列である)を組み込むペプチドを生成することができる。還元されたペプチド結合はジペプチド・サブユニットとして導入することができる。そのようなポリペプチドはプロテアーゼ活性に対して耐性であり、インビボで長期化した半減期を有する。

【0054】

上に論じられるように、細胞内でIV型コラーゲンが合成されてヘテロ三量体として構築され、次にそれが細胞外に分泌され、そこで六量体構築、次いで基底膜構築が生じる。ここに開示されるポリペプチドは細胞内で機能してヘテロ三量体構築を妨げ、それは必然的に六量体構築をも阻害し、所望の治療結果をもたらす。その代わりに(または、それに加えて)、ここに開示されるポリペプチドは細胞外で機能して六量体構築を阻害し、及び/または構築された六量体を破壊し、所望の治療結果をもたらす。

10

【0055】

そのようなポリペプチドは、一般的ヘテロ三量体構築(すなわち、鎖特異的ではない);特異的ヘテロ三量体構築(すなわち、鎖特異的);一般的六量体構築(すなわち、鎖特異的ではない);及び/または特異的六量体構築(すなわち、特異的ではない)の阻害におけるそれらの有用性に基づいて選択することができる。ここで説明されるIV型コラーゲン[(1)₂(2)]₂NC1六量体構造の知識なしに、そのような所望の特性を有する阻害剤の設計は当該技術分野における熟練者に利用可能ではない。

【0056】

アミノ酸の一文字略語がここで用いられる;「norL」はノルロイシンを指す。

【0057】

一実施形態において、ポリペプチドは一般式Iの、少なくとも8つの連続するアミノ酸からなる:

20

PF(R1)(R2)CN(R3)(R4)(R5)VC(R6)(R7)A(配列番号1)

R1はL、M、A、V、norL、及びIからなる群より選択され;

R2はF及びYからなる群より選択され;

R3はI、V、L、norL、A、及びPからなる群より選択され;

R4はN、G、及びHからなる群より選択され;

R5はN、D、Q、及びEからなる群より選択され;

R6はN、Y、及びHからなる群より選択され;並びに

R7はF及びYからなる群より選択される。

30

【0058】

この一般式Iは、以下にさらに説明されるように、結晶構造中に6-7鎖を含む鎖間ドメインスワッピング領域(「CDSR間」)のIV型コラーゲンNC1 1-6ドメインの共通配列から誘導される。この領域はヘテロ三量体内の鎖間相互作用に関与し、この配列の実質的な部分は六量体界面にも存在し、従って、六量体構築/安定化に関与する。そのようなものとして、一般式Iのペプチドは適切な鎖間相互作用の阻害に有用であり、従って、最適ヘテロ三量体及び六量体構築の破壊に有用である。

【0059】

様々な更なる実施形態において、これらのポリペプチドは一般式Iの少なくとも9、10、11、12、13、または14アミノ酸からなる。好ましい実施形態において、このポリペプチドは一般式Iの14アミノ酸からなる。

40

【0060】

好ましい実施形態において、このポリペプチドは、R2がFであり;R4がNであり;R5がN及びDからなる群より選択され;R6がNであり;並びにR7がFであるという更なる制限を伴う、一般式Iの少なくとも8個の連続するアミノ酸からなる。この実施形態のポリペプチドはCDSR間のIV型コラーゲンNC1 1、3、及び5ドメインの共通配列から誘導される。

【0061】

更なる好ましい実施形態において、これらのポリペプチドは、R2がYであり;R3がP

50

及びIからなる群より選択され；R5がD、Q、及びEからなる群より選択され；R6がY及びHからなる群より選択され；並びにR7がYであるという更なる制限を伴う、一般式Iの少なくとも8個の連続するアミノ酸からなる。この実施形態のポリペプチドは、CDSR間のIV型コラーゲンNC1 2、4、及び6ドメインの共通配列から誘導される。

【0062】

更なる好ましい実施形態において、式Iによるポリペプチドは、PFLFCNINNVCNFA(1)(配列番号2)；PFLFCNVNDVCNFA(3)(配列番号3)；PFMFCNINNVCNFA(5)(配列番号4)；PFLYCNP GDVCYYA(2)(配列番号5)；PFAYCNINH QVCHYA(4)(配列番号6)；及びPFICYCNINEVCHYA(6)(配列番号7)からなる群より選択される配列の少なくとも8個の連続するアミノ酸からなる。これらの配列は、個々のIV型コラーゲン1-6 NC1ドメインからのCDSR間配列を表す。様々な更なる実施形態において、これらのポリペプチドは列挙される配列のうちの1つの少なくとも9、10、11、12、13、または14アミノ酸からなる。好ましい実施形態において、このポリペプチドは列挙される配列のうちの1つの14アミノ酸からなる。

10

【0063】

別の実施形態においては、本発明のポリペプチドは一般式IIの、少なくとも7個の連続するアミノ酸からなる：

PF(R1)EC(R2)G(R3)(R4)GTC(R5)(配列番号8)

20

R1はL、A、V、norL、及びIからなる群より選択され；

R2はH、N、Q、及びSからなる群より選択され；

R3はG、R、A、または不在からなる群より選択され；

R4はR及びQからなる群より選択され；並びに

R5はN及びHからなる群より選択される。

【0064】

この一般式は、以下にさらに説明されるように、結晶中に6'-7'鎖を含む鎖内ドメインスワッピング領域(「CDSR内」)のIV型コラーゲンNC1 1-6ドメインの共通配列から誘導される。この領域はヘテロ三量体内の単量体-単量体相互作用に参与し、その配列の実質的な部分は六量体界面にも存在し、従って、六量体構築/安定化に参与する。そのようなものとして、この一般式のペプチドはIV型コラーゲンのヘテロ三量体及び六量体相互作用の両者の阻害に有用である。

30

【0065】

様々な更なる実施形態において、これらのペプチドは一般式IIの少なくとも8、9、10、11、12、または13アミノ酸からなる。好ましい実施形態において、このペプチドは一般式IIの13アミノ酸からなる。

【0066】

好ましい実施形態において、これらのペプチドは、R2がHであり；R3がRであり；R4がGであり；及びR5がNであるという更なる制限を伴う、一般式IIの少なくとも7個の連続するアミノ酸からなる。この実施形態のポリペプチドはIV型コラーゲン1、3、及び5 NC1ドメインのCDSR内配列の共通配列から誘導される。

40

【0067】

更なる好ましい実施形態において、これらのポリペプチドは、R2がN、Q、及びSからなる群より選択され；R3がG、R、及びAからなる群より選択され；R4がR及びQからなる群より選択され；並びにR5がHであるという更なる制限を伴う、一般式IIの少なくとも7個の連続するアミノ酸からなる。この実施形態のポリペプチドはIV型コラーゲン2、4、及び6 NC1ドメインのCDSR内配列の共通配列から誘導される。

。

【0068】

更なる実施形態において、一般式IIによるポリペプチドは、PFIECHGRGTCN

50

(1 及び 5) (配列番号 9) ; P F L E C H G R G T C N (3) (配列番号 1 0) ; P F I E C N G G R G T C H (2) (配列番号 1 1) ; P F L E C Q G R Q G T C H (4) (配列番号 1 2) ; 及び P F I E C S G A R G T C H (6) (配列番号 1 3) からなる群より選択される配列の少なくとも 7 個の連続するアミノ酸からなる。これらの配列は、個々の I V 型コラーゲン 1 - 6 N C 1 ドメインからの C D S R 内配列を表す。様々な更なる実施形態において、この実施形態のポリペプチドは列挙される配列のうちの 1 つの少なくとも 8、9、10、11、12、または 13 アミノ酸からなる。最も好ましい実施形態においては、これらのポリペプチドは列挙される配列のいずれか 1 つの 1 2 (1 , 3 , 5) または 1 3 (2 , 4 , 6) 個の連続するアミノ酸からなる。

10

【0069】

更なる実施形態においては、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、完全長 C D S R 内ポリペプチド (例えば、配列番号 9、10、11、12、または 13) は、任意に、同じ鎖から誘導される 0 ~ 5 アミノ酸をアミノ及びカルボキシ末端のいずれかまたは両者にさらに含むことができる。従って、1 様 N C 1 鎖の C D S R 内から誘導される本発明のポリペプチドは：

1 : (E) (F) (R) (S) (A) P F I E C H G R G T C N (Y) (Y) (A) (N) (A) (配列番号 1 4)、

3 : (E) (F) (R) (A) (S) P F L E C H G R G T C N (Y) (Y) (S) (N) (S) (配列番号 1 5) ; 及び

20

5 : (E) (F) (R) (S) (A) P F I E C H G R G T C N (Y) (Y) (A) (N) (S) (配列番号 1 6) ;

(ここで、括弧内の残基は C D S R 内のフランキング配列である)

からなる群より選択される配列の少なくとも 1 2、1 3、1 4、1 5、1 6、1 7、1 8、1 9、2 0、2 1、または 2 2 アミノ酸からなる群より選択することができる。

【0070】

その代わりに、2 様 N C 1 鎖の C D S R 内配列から誘導される本発明のポリペプチドは：

2 : (D) (F) (R) (A) (T) P F I E C N G G R G T C H (Y) (Y) (A) (N) (K) (配列番号 1 7) ;

30

4 : (D) (F) (R) (A) (A) P F L E C Q G R Q G T C H (F) (F) (A) (N) (K) (配列番号 1 8) ; 及び

6 : (D) (F) (R) (A) (T) P F I E C S G A R G T C H (Y) (F) (A) (N) (K) (配列番号 1 9) ;

(ここで、括弧内の残基は C D S R 内のフランキング配列である)

からなる群より選択される配列の少なくとも 1 3、1 4、1 5、1 6、1 7、1 8、1 9、2 0、2 1、2 2、または 2 3 アミノ酸からなる群より選択することができる。

【0071】

C D S R 間配列は、所定の I V 型コラーゲン N C 1 ドメインの直線配列においては同じ鎖内の C D S R 内配列から大きく分離されてはいるが (約 1 0 0 アミノ酸分離されている)、誘導された結晶構造データに基づいて同じ鎖内の C D S R 内配列と密接した空間的近位に存在する。従って、別の実施形態においては、本発明は：

40

(a) 1 つ以上の一般式 I の C D S R 間ポリペプチド ;

(b) 1 つ以上の一般式 I I の C D S R 内ポリペプチド ; 及び

(c) 0 - 2 0 アミノ酸からなる、C D S R 内ポリペプチドと C D S R 間ポリペプチドとの間のリンカーポリペプチド、を含むキメラポリペプチドを提供する。

【0072】

好ましい実施形態において、これらのキメラポリペプチドの C D S R 間及び / または C D S R 内部分は、それぞれ、一般式 I の 8、9、10、11、12、13、または 14 アミ

50

ノ酸及び一般式 I I の 7、8、9、10、11、12、13 アミノ酸からなる。様々な他の好ましい実施形態においては、リンカーポリペプチドは 0、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または 20 アミノ酸からなる。このスペーサーの最適の長さは、少なくとも部分的には、C D S R 間及び C D S R 内の長さに加えて、そのキメラを創出するのに用いられる完全長 C D S R 間及び C D S R 内内の配列の位置に依存する。例えば、完全長 C D S R 間及び完全長 C D S R 内が用いられた場合、スペーサーは好ましくは長さが 0 ~ 5 アミノ酸、より好ましくは長さが 1 - 4 アミノ酸、最も好ましくは長さが 2 - 3 アミノ酸である。ここでの教示に基づくと、更なるそのようなキメラポリペプチドを設計することは当該技術分野における熟練者には明らかであろう。

10

【0073】

これらのキメラポリペプチドの最も好ましい実施形態においては、C D S R 間ポリペプチドは、P F L F C N I N N V C N F A (配列番号 2)、P F L F C N V N D V C N F A (配列番号 3)、P F M F C N I N N V C N F A (配列番号 4)、P F L Y C N P G D V C Y Y A (配列番号 5)、P F A Y C N I H Q V C H Y A (配列番号 6)、及び P F I Y C N I N E V C H Y A (配列番号 7) からなる群より選択され；C D S R 内ポリペプチドは、P F I E C H G R G T C N (配列番号 9)、P F L E C H G R G T C N (配列番号 10)、P F I E C N G G R G T C H (配列番号 11)、P F L E C Q G R Q G T C H (配列番号 12)、及び P F I E C S G A R G T C H (配列番号 13) からなる群より選択され；並びに、リンカーポリペプチドは 1、2、3、4、または 5 アミノ酸；最も好ましくは 2 アミノ酸である。

20

【0074】

別の実施形態において、本発明のポリペプチドは一般式 I I I :

F (R 1) T (R 2) (配列番号 20)

(ここで、R 1 は S 及び T からなる群より選択され；並びに

R 2 は M 及び L からなる群より選択される)

のアミノ酸の配列からなる。

【0075】

この一般式 I I I は、以下にさらに説明されるように、結晶構造における 5 - 6 鎖の間の特異的領域 (「S R」) での I V 型コラーゲン N C 1 1 - 6 ドメインの共通配列から誘導される。この領域は、所定の 鎖の S R が相互作用する単量体中の特異的領域パートナー (「S R P」) を認識することにより、単量体間の特異的認識に参与する。そのようなものとして、一般式 I I I のペプチドは I V 型コラーゲンのヘテロ三量体及び六量体両者の相互作用の阻害に有用である。

30

【0076】

更なる実施形態において、S R ポリペプチドは、F S T M (1、2、5、及び 6) (配列番号 21)、F T T M (3) (配列番号 22) 及び F T S L (4) (配列番号 23) からなる群より選択される。

【0077】

更なる実施形態において、S R ポリペプチド (例えば、配列番号 21、22、及び 23) は、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意に、アミノ及びカルボキシル末端のいずれかまたは両者に 0 ~ 5 アミノ酸をさらに含むことができる。従って、この実施形態によると、N C 1 鎖の S R 配列から誘導される本発明のポリペプチドは：

40

1 : X 1 - F S T M - Z 1、ここで、X 1 は配列 S C L R K (配列番号 24) の 0、1、2、3、4、または 5 アミノ酸であり、及び Z 1 は配列 P F L F C (配列番号 25) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全な配列は S C L R K F S T M P F L F C である) (配列番号 26)；

3 : X 3 - F T T M - Z 3、ここで、X 3 は配列 S C L Q R (配列番号 27) の 0、1、2、3、4、または 5 アミノ酸であり、及び Z 3 は配列 P F L F C (配列番号 25) の

50

0、1、2、3、4、または5アミノ酸である（従って、完全な配列はSCLQRFTT
MPFLFCである）（配列番号28）；

5：X5-FSTM-Z5、ここで、X5は配列SCLRR（配列番号29）の0、1、2、3、4、または5アミノ酸であり、及びZ5は配列PFMFC（配列番号30）の0、1、2、3、4、または5アミノ酸である（従って、完全な配列はSCLRRFSTM
PFMFCである）（配列番号31）；

2：X2-FSTM-Z2、ここで、X2は配列SCLAR（配列番号32）の0、1、2、3、4、または5アミノ酸であり、及びZ2は配列PFLYC（配列番号33）の0、1、2、3、4、または5アミノ酸である（従って、完全な配列はSCLARFSTM
PFLYCである）（配列番号34）；

4：X4-FSTL-Z4、ここで、X4は配列SCLPV（配列番号35）の0、1、2、3、4、または5アミノ酸であり、及びZ4は配列PFAYC（配列番号36）の0、1、2、3、4、または5アミノ酸である（従って、完全な配列はSCLPVFST
LPAYCである）（配列番号37）；並びに

6：X6-FSTM-Z6、ここで、X6は配列SCLPR（配列番号38）の0、1、2、3、4、または5アミノ酸であり、及びZ6は配列PFIYC（配列番号39）の0、1、2、3、4、または5アミノ酸である（従って、完全な配列はSCLPRFSTM
PFIYCである）（配列番号40）、
からなる群より選択することができる。

【0078】

別の実施形態においては、本発明のポリペプチドは一般式IV：

(R1)MF(R2)K（配列番号41）

（ここで、R1はE、R、及びDからなる群より選択され；並びに

R2はK、R、及びSからなる群より選択される）

のアミノ酸配列からなる。

【0079】

この一般式IVは、以下により詳細に論じられるように、8'及び9'鎖の間に位置する特異的領域パートナー（「SRP」）のIV型コラーゲンNC1 1、3、及び5ドメインの共通配列から誘導される。この領域は、所定の鎖のSRPが相互作用する、単量体における特異的領域（「SR」）を認識することにより、単量体間の特異的認識に
関与する。そのようなものとして、一般式IVのペプチドはIV型コラーゲンのヘテロ三量体及び六量体相互作用の両者の阻害に有用である。

【0080】

好ましい実施形態においては、一般式IVによるSRPポリペプチドは、EMFKK（1）（配列番号42）、RMFRK（3）（配列番号43）、及びDMFSK（5）（配列番号44）からなる群より選択される。

【0081】

更なる好ましい実施形態においては、SRPポリペプチドは、SFQ（2のSRP）（配列番号45）；LQF（4のSRP）（配列番号46）、及びQQF（6のSRP）（配列番号47）からなる群より選択される。これらの配列は、示されるように、IV型コラーゲン鎖NC1ドメインのSRPを表す。2 NC1ドメインにおけるこの領域は、以下にさらに論じられるように、拡張立体配座をとり、かつ隣接1鎖における拡張構造（Phe57-Thr59）と対を形成して短い平行シートを形成し、これは構造全体における唯一の平行シートである。

【0082】

更なる実施形態において、これらのSRPポリペプチド（例えば、配列番号42-47）は、最適阻害活性にポリペプチドに適切な二次構造の特長を付与するため、任意に、同じ鎖から誘導される0-5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者にさらに含むことができる。従って、本発明のこの実施形態のSRP含有ポリペプチドは、

10

20

30

40

50

1 X1 - EMFKK - Z1、ここで、X1は配列T I E R S (配列番号48)の0、1、2、3、4、または5アミノ酸であり、Z1は配列P T P S T (配列番号49)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT I E R S E M F K K P T P S Tである)(配列番号50)；

3 : X3 - R M F R K - Z3、ここで、X3は配列S L N P E (配列番号51)の0、1、2、3、4、または5アミノ酸であり、Z3は配列P I P S T (配列番号52)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はS L N P E R M F R K P I P S Tである)(配列番号53)；

5 : X5 - D M F S K - Z5、ここで、X5は配列T V D V S (配列番号54)の0、1、2、3、4、または5アミノ酸であり、Z5は配列P Q S E T (配列番号55)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT V D V S D M F S K P Q S E Tである)(配列番号56)；

2 : X2 - S F Q - Z2、ここで、X2は配列T I P E Q (配列番号57)の0、1、2、3、4、または5アミノ酸であり、Z2は配列G S P S A (配列番号58)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT I P E Q S F Q G S P S Aである)(配列番号59)；

4 : X4 - L Q F - Z4、ここで、X4は配列T V K A D (配列番号60)の0、1、2、3、4、または5アミノ酸であり、Z4は配列S S A P A (配列番号61)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT V K A D L Q F S S A P Aである)(配列番号62)；並びに

6 : X6 - Q Q F - Z6、ここで、X6は配列T V E E R (配列番号63)の0、1、2、3、4、または5アミノ酸であり、Z6は配列G E L P V (配列番号64)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT V E E R Q Q F G E L P Vである)(配列番号65)、

からなる群より選択することができる。

【0083】

別の実施形態においては、本発明のポリペプチドは、一般式V：

(R1)AH(R2)QD (配列番号66)

(ここで、R1はR及びKからなる群より選択され；並びに

R2はG及びNからなる群より選択される)

のアミノ酸配列からなる。

【0084】

この一般式Vは、以下により詳細に論じられるように、4鎖のIV型コラーゲンNC1ドメイン - 樽様コアの共通配列から誘導される。この領域は一般的な単量体 - 単量体相互作用に参与する。そのようなものとして、一般式VのペプチドはIV型コラーゲンのヘテロ三量体及び六量体相互作用の両者の阻害に有用である。

【0085】

好ましい実施形態において、一般式Vによるポリペプチドは、RAH G Q D (1 , 3 , 5) (配列番号67)及びKAHN Q D (2 , 4 , 6) (配列番号68)からなる群より選択される。

【0086】

更なる好ましい実施形態においては、一般式Vによる - 樽ポリペプチド(例えば、配列番号67~68)は、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意に、同じ鎖から誘導される0~5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者に含むことができる。従って、本発明のこの実施形態の - 樽含有ポリペプチドは：

1 X1 - R A H G Q D - Z1、ここで、X1は配列V Q G N E (配列番号69)の0、1、2、3、4、または5アミノ酸であり、Z1は配列L G T A G (配列番号70)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はV Q G N E R A H G Q D D L G T Aである)(配列番号71)；

10

20

30

40

50

3 : X₃ - R A H G Q D - Z₃、ここで、X₃は配列V Q G N Q (配列番号72)の0、1、2、3、4、または5アミノ酸であり、Z₃は配列L G T L G (配列番号73)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はV Q G N Q R A H G Q D L G T L Gである)(配列番号74) ;

5 : X₅ - R A H G Q D - Z₅、ここで、X₅は配列V Q G N K (配列番号75)の0、1、2、3、4、または5アミノ酸であり、Z₅は配列L G T A G (配列番号70)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はV Q G N K R A H G Q D L G T A Gである)(配列番号76) ;

2 : X₂ - K A H N Q D - Z₂、ここで、X₂は配列F E G Q E (配列番号77)の0、1、2、3、4、または5アミノ酸であり、Z₂は配列L G L A G (配列番号78)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はF E G Q E K A H N Q D L G L A Gである)(配列番号79) ;

4 : X₄ - K A H N Q D - Z₄、ここで、X₄は配列L E G Q E (配列番号80)の0、1、2、3、4、または5アミノ酸であり、Z₄は配列L G L A G (配列番号78)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はL E G Q E K A H N Q D L G L A Gである)(配列番号81) ;並びに

6 : X₆ - K A H N Q D - Z₆、ここで、X₆は配列V E G Q E (配列番号82)の0、1、2、3、4、または5アミノ酸であり、Z₆は配列L G F A G (配列番号83)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はV E G Q E K A H N Q D L G F A Gである)(配列番号84)、
からなる群より選択することができる。

【0087】

別の実施形態においては、本発明のポリペプチドは一般式V I :

(R₁)G(R₂)GQ(配列番号85)

(ここで、R₁はE及びQからなる群より選択され ; 並びに

R₂はS、T、及びGからなる群より選択される)

のアミノ酸配列からなる。

【0088】

この一般式V Iは、以下により詳細に論じられるように、4'鎖のI V型コラーゲンN C1ドメイン - 樽様コアの共通配列から誘導される。この領域は一般的な単量体 - 単量体相互作用に参与する。そのようなものとして、一般式V IのペプチドはI V型コラーゲンのヘテロ三量体及び六量体の両者の阻害に有用である。

【0089】

好ましい実施形態においては、一般式V Iによるポリペプチドは、E G S G Q (1 , 5) (配列番号86)、E G T G Q (3) (配列番号87)、E G G G Q (2 , 6) (配列番号88)及びQ G G G Q (4) (配列番号89)からなる群より選択される。

【0090】

更なる実施形態においては、一般式V Iによる - 樽ポリペプチド(例えば、配列番号86 ~ 89)は、最適阻害活性に適切な二次構造の特長をペプチドに付与するため、任意に、同じ鎖から誘導される0 ~ 5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者にさらに含むことができる。従って、本発明のこの実施形態の - 樽含有ポリペプチドは :

1及び5 X₁ - E G S G Q - Z₁、ここで、X₁は配列T S A G A (配列番号90)の0、1、2、3、4、または5アミノ酸であり、Z₁は配列A L A S P (配列番号91)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT S A G A E G S G Q A L A S Pである)(配列番号92) ;

3 : X₃ - E G T G Q - Z₃、ここで、X₃は配列T S A G S (配列番号93)の0、1、2、3、4、または5アミノ酸であり、Z₃は配列A L A S P (配列番号91)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT S A G S E G T G

Q A L A S Pである) (配列番号94) ;

2 : X 2 - E G G G Q - Z 2、ここで、X 2は配列T A A G D (配列番号95)の0、1、2、3、4、または5アミノ酸であり、Z 2は配列S L V S P (配列番号96)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT A A G D E G G G Q S L V S Pである) (配列番号97) ;

4 : X 4 - Q G G G Q - Z 4、ここで、X 4は配列T G A G D (配列番号98)の0、1、2、3、4、または5アミノ酸であり、Z 4は配列A L M S P (配列番号99)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT G A G D Q G G G Q A L M S Pである) (配列番号100) ;並びに

6 : X 6 - E G G G Q - Z 6、ここで、X 6は配列T A A G A (配列番号101)の0、1、2、3、4、または5アミノ酸であり、Z 6は配列S L V S P (配列番号96)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT A A G A E G G G Q S L V S Pである) (配列番号102)、

からなる群より選択することができる。

【0091】

別の実施形態において、これらのポリペプチドは、推定結晶構造から決定されるように、六量体界面に存在する配列を含む。I V型コラーゲンは細胞内部で合成されて三量体として構築され、次にそれは細胞外に分泌され、そこで六量体構築、及び続いて基底膜構築が生じる。治療薬、例えば、ここに開示されるものは、細胞内で三量体構築、従って、六量体の構築を妨げるように機能し、従って、所望の治療結果をもたらすことができる。その代わりに(または、それに加えて)、治療薬は細胞外で機能することができ、それは三量体構築を非阻害のままにするが六量体構築を標的とする。

【0092】

そのようなものとして、六量体界面の領域からのポリペプチドは六量体形成の阻害または六量体形成の破壊に用いることができる。この実施形態においては、本発明のポリペプチドは一般式V I I :

(R 1) G (R 2) (R 3) (配列番号103)

(ここで、R 1はQ及びEからなる群より選択され ;

R 2はN及びQからなる群より選択され ;並びに

R 3はE、Q、及びKからなる群より選択される)

のアミノ酸配列からなる。

【0093】

この一般式V I Iは、以下により詳細に論じられるように、3鎖の終端から4鎖の最初までの六量体界面のI V型コラーゲンNC1 1 - 6ドメインの共通配列から誘導される。この領域は六量体界面に存在し、六量体の構築及び安定化に関与する。そのようなものとして、一般式V I IのペプチドはI V型コラーゲンの六量体相互作用の阻害に有用である。

【0094】

好ましい実施形態において、これらのポリペプチドは、R 1がQであり、かつR 2がNであるという更なる限定を伴う一般式V I Iからなる。更なる好ましい実施形態において、一般式V I Iによるポリペプチドは、Q G N E (1) (配列番号104)、Q G N Q (3) (配列番号105)、及びQ G N K (5) (配列番号106)からなる群より選択される。

【0095】

更なる好ましい実施形態において、一般式V I IによるポリペプチドはE G Q E (配列番号107)からなり、これは一般式V I Iにおいて2 / 4 / 6 NC1ドメインに存在する配列の配列である。

【0096】

更なる実施形態において、配列番号104 ~ 107からなる群より選択される六量体ポリペプチドは、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意

10

20

30

40

50

に、同じ鎖から誘導される0～5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者にさらに含むことができる。従って、そのようなポリペプチドは：

1：X1 - QGNE - Z1、ここで、X1は配列SLLYV（配列番号108）の0、1、2、3、4、または5アミノ酸であり、Z1は配列RAHGQ（配列番号109）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はSLLYVQGNERAHGQである）（配列番号110）；

3：X3 - QGNQ - Z3、ここで、X3は配列SFLFV（配列番号111）の0、1、2、3、4、または5アミノ酸であり、Z3は配列RAHGQ（配列番号109）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はSFLFVQGNQR AHGQである）（配列番号112）；

5：X5 - QGNK - Z5、ここで、X5は配列SLLYV（配列番号108）の0、1、2、3、4、または5アミノ酸であり、Z5は配列RAHGQ（配列番号109）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はSLLYVQGNKR AHGQである）（配列番号113）；

2：X2 - EGQE - Z2、ここで、X2は配列SLLYF（配列番号114）の0、1、2、3、4、または5アミノ酸であり、Z2は配列KAHNQ（配列番号115）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はSLLYFEGQEK AHNQである）（配列番号116）；

4：X4 - EGQE - Z4、ここで、X4は配列SLLYL（配列番号117）の0、1、2、3、4、または5アミノ酸であり、Z4は配列KAHNQ（配列番号115）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はSLLYLEGQEK AHNQである）（配列番号118）；並びに

6：X6 - EGQE - Z6、ここで、X6は配列SLLFV（配列番号119）の0、1、2、3、4、または5アミノ酸であり、Z6は配列KAHNQ（配列番号115）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はSLLFVEGQEK AHNQである）（配列番号120）、

からなる群より選択することができる。

【0097】

一般式VIIによるこれらの六量体界面ポリペプチドの特に好ましい実施形態は、以下のように、1 - 6六量体ペプチドのアミノ及びカルボキシル末端の更なる1アミノ酸からなる：

1 VQGNER（配列番号121）

3：VQGNQR（配列番号122）

5：VQGNKR（配列番号123）

2：FEGQEK（配列番号124）

4：LEGQEK（配列番号125）

6：VEGQEK（配列番号126）。

【0098】

推定結晶構造からの決定で、ポリペプチドが六量体界面に存在する配列を含む更なる実施形態においては、本発明のポリペプチドは一般式VIII：

M(R1)M(R2)P（配列番号127）

（ここで、R1はS、Nからなる群から選択されるか、または存在せず；並びに

R2はA、Qからなる群より選択されるか、または存在しない）

のアミノ酸配列からなる。

【0099】

この一般式VIIIは、以下により詳細に論じられるように、8及び9鎖の間の六量体界面のIV型コラーゲンNC1 1 - 6ドメインの共通配列から誘導される。この領域は六量体界面に存在し、六量体の構築及び安定化に参与する。そのようなものとして、一般式VIIIのペプチドはIV型コラーゲンの六量体相互作用の阻害に有用である。

【0100】

10

20

30

40

50

好ましい実施形態において、一般式V I I Iのポリペプチドは、M S M A P (1) (配列番号128)、M N M A P (3) (配列番号129)、M S M Q P (5) (配列番号130)、及びM M P (2、4、及び6) (配列番号131) からなる群より選択される。

【0101】

更なる好ましい実施形態において、配列番号128～131からなる群より選択される六量体ポリペプチドは、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意に、同じ鎖から誘導される0～5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両方にさらに含むことができる。従って、そのようなポリペプチドは：

1：X1 - M S M A P - Z1、ここで、X1は配列P E P M P (配列番号132) の0、1、2、3、4、または5アミノ酸であり、Z1は配列I T G E N (配列番号133) の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はP E P M P M S M A P I T G E Nである)(配列番号134)；

3：X3 - M N M A P - Z3、ここで、X3は配列P A L M P (配列番号135) の0、1、2、3、4、または5アミノ酸であり、Z3は配列I T G R A (配列番号136) の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はP A L M P M N M A P I T G R Aである)(配列番号137)；

5：X5 - M S M Q P - Z5、ここで、X5は配列P E P M P (配列番号132) の0、1、2、3、4、または5アミノ酸であり、Z5は配列L K G Q S (配列番号138) の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はP E P M P M S M Q P L K G Q Sである)(配列番号139)；

2；X2 - M M P - Z2、ここで、X2は配列T A P L P (配列番号140) の0、1、2、3、4、または5アミノ酸であり、Z2は配列V A E D E (配列番号141) の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT A P L P M M P V A E D Eである)(配列番号142)；

4：X4 - M M P - Z4、ここで、X4は配列A A P L P (配列番号143) の0、1、2、3、4、または5アミノ酸であり、Z4は配列L S E E A (配列番号144) の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はA A P L P M M P L S E E Aである)(配列番号145)；並びに

6：X6 - M M P - Z6、ここで、X6は配列T A P I P (配列番号146) の0、1、2、3、4、または5アミノ酸であり、Z6は配列V S Q T Q (配列番号147) の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はT A P I P M M P V S Q T Qである)(配列番号148)、からなる群より選択することができる。

【0102】

一般式V I I Iによるこれらの六量体界面ペプチドの特に好ましい実施形態は、以下の様に、1～6六量体ペプチドのアミノ及びカルボキシル末端の両者の3つの更なるアミノ酸からなる：

1：P M P M S M A P I T G (配列番号149) ；

3：L M P M N M A P I T G (配列番号150) ；

5：P M P M S M Q P L K G (配列番号151) ；

2：P L P M M P V A E (配列番号152) ；

4：P L P M M P L S E (配列番号153) ；及び

6：P I P M M P V S Q (配列番号154) 。

【0103】

推定結晶構造からの決定でポリペプチドが六量体界面に存在する配列を含む更なる実施形態においては、本発明のポリペプチドは一般式I X：

A G (R 1) (R 2) (配列番号155)

(ここで、R1はA、S及びDからなる群より選択され；並びに

R2はE及びQからなる群より選択される)

10

20

30

40

50

のアミノ酸配列からなる。

【0104】

この一般式IXは、以下により詳細に論じられるように、3'及び4'鎖の間のIV型コラーゲンNC1 1-6ドメインの共通配列から誘導される。この領域は六量体界面に存在し、六量体の構築及び安定化に關与する。そのようなものとして、一般式IXのペプチドはIV型コラーゲンの六量体相互作用の阻害に有用である。

【0105】

好ましい実施形態において、一般式IXのポリペプチドは、AGAE(1、5、及び6)(配列番号156)、AGSE(3)(配列番号157)、AGDE(2)(配列番号158)、及びAGDQ(4)(配列番号159)からなる群より選択される。 10

【0106】

更なる実施形態においては、配列番号156~159からなる群より選択される六量体ポリペプチドは、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意に、同じ鎖から誘導される0~5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者にさらに含むことができる。従って、そのようなポリペプチドは：

1: X1-AGAE-Z1、ここで、X1は配列VMHTS(配列番号160)の0、1、2、3、4、または5アミノ酸であり、Z1は配列GSGQA(配列番号161)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はVMHTSAGAE GSGQAである)(配列番号162)； 20

3: X3-AGSE-Z3、ここで、X3はIMFTS(配列番号163)の0、1、2、3、4、または5アミノ酸であり、Z3は配列GTGQA(配列番号164)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はIMFTSAGSE GTGQAである)(配列番号165)；

5: X5-AGAE-Z5、ここで、X5は配列MMHTS(配列番号166)の0、1、2、3、4、または5アミノ酸であり、Z5は配列GSGQA(配列番号161)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はMMHTSAGAE GSGQAである)(配列番号167)；

2: X2-AGDE-Z2、ここで、X2は配列LMHTA(配列番号168)の0、1、2、3、4、または5アミノ酸であり、Z2は配列GGGQS(配列番号169)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はLMHTAAGDE GGGQSである)(配列番号170)； 30

4: X4-AGDQ-Z4、ここで、X4は配列LMHTG(配列番号171)の0、1、2、3、4、または5アミノ酸であり、Z4は配列GGGQA(配列番号172)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はLMHTGAGD QGGGQAである)(配列番号173)；並びに

6: X6-AGAE-Z6、ここで、X6は配列LMHTA(配列番号168)の0、1、2、3、4、または5アミノ酸であり、Z6は配列GGGQS(配列番号169)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はLMHTAAGAE GGGQSである)(配列番号174)、 40

からなる群より選択することができる。

【0107】

推定結晶構造からの決定でポリペプチドが六量体界面に存在する配列を含む更なる実施形態においては、本発明のポリペプチドは一般式X：

EC(R1)G(R2)(R3)GTC(R4)(R5)(R6)(配列番号175)

(ここで、R1はH、N、Q、及びSからなる群より選択され；

R2はG、R、Aからなる群から選択されるか、または存在せず；

R3はR及びQからなる群より選択され

R4はN及びHからなる群より選択され；

R5はF及びYからなる群より選択され；並びに 50

R 6 は F 及び Y からなる群より選択される)
の配列の少なくとも 5 アミノ酸からなる。

【0108】

様々な好ましい実施形態において、このポリペプチドは一般式 X の少なくとも 6、7、8、9、10、11、または 12 アミノ酸からなる。好ましい実施形態においては、このポリペプチドは一般式 X の 12 アミノ酸からなる。この一般式 X は、以下により詳細に論じられるように、上で論じられる C D S R 内と広範囲にわたって重複し、6' - 7' 鎖内に存在する。この領域は六量体界面に存在し、六量体の構築及び安定化に参与する。そのようなものとして、一般式 X のペプチドは I V 型コラーゲンの六量体相互作用の障害に有用である。

10

【0109】

更なる実施形態においては、これらのポリペプチドは、R 2 が G、R、A からなる群より選択され；かつ R 4 が H であることを除いて、一般式 X について上述される通りである。この実施形態のポリペプチドは一般式 X の 2 / 4 / 6 の共通配列から誘導される。

【0110】

更なる好ましい実施形態においては、一般式 X のポリペプチドは、E C H G R G T C N Y Y (1 / 3 / 5) (配列番号 176)、E C N G G R G T C H Y Y (2) (配列番号 177)、E C Q G R Q G T C H F F (4) (配列番号 178)、及び E C S G A R G T C H Y F (6) (配列番号 179) からなる群より選択される。

20

【0111】

更なる好ましい実施形態においては、本発明のポリペプチドは一般式 X I :

(R 1) (R 2) T (R 3) K (配列番号 180)

(ここで、R 1 は P、S、及び A からなる群より選択され；

R 2 は S、E、及び D からなる群より選択され；並びに

R 3 は L 及び V からなる群より選択される)

のアミノ酸配列からなる。

【0112】

この一般式 X I は、以下により詳細に論じられるように、結晶構造中で 9' 鎖と重複して存在する。この領域は六量体界面に存在し、六量体の構築及び安定化に参与する。そのようなものとして、一般式 X I のペプチドは I V 型コラーゲンの六量体相互作用の障害に有用である。

30

【0113】

一般式 X I の好ましい実施形態においては、R 3 が L である (2 / 4 / 6 / 1 / 5 におけるように)。一般式 X の更なる好ましい実施形態においては、R 2 が D 及び E から選択される (2 / 4 / 5 / 6)。更なる好ましい実施形態においては、一般式 X I によるポリペプチドは P S T L K (1) (配列番号 181)、P S T V K (3) (配列番号 182)、S E T L K (5 及び 6) (配列番号 183)、A D T L K (2) (配列番号 184)、及び P D T L K (4) (配列番号 185) からなる群より選択される。

【0114】

更なる実施形態において、配列番号 181 ~ 185 からなる群より選択される六量体ポリペプチドは、最適障害活性に適する二次構造の特長をポリペプチドに付与するため、任意に、同じ鎖から誘導される 0 ~ 5 アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者にさらに含むことができる。従って、そのようなポリペプチドは：

40

1 : X 1 - P S T L K - Z 1、ここで、X 1 は配列 F K K P T (配列番号 186) の 0、1、2、3、4、または 5 アミノ酸であり、Z 1 は配列 A G E L R (配列番号 187) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は F K K P T P S T L K A G E L R である) (配列番号 188)；

3 : X 3 - P S T V K - Z 3、ここで、X 3 は配列 F R K P I (配列番号 189) の 0、1、2、3、4、または 5 アミノ酸であり、Z 3 は配列 A G E L E (配列番号 190) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は F R K P I P S

50

T V K A G E L Eである) (配列番号191) ;

5 : X 5 - S E T L K - Z 5、ここで、X 5は配列F S K P Q (配列番号192)の0、1、2、3、4、または5アミノ酸であり、Z 5は配列A G D L R (配列番号193)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はF S K P Q S E T L K A G D L Rである) (配列番号194) ;

2 : X 2 - A D T L K - Z 2、ここで、X 2は配列Q G S P S (配列番号195)の0、1、2、3、4、または5アミノ酸であり、Z 2は配列A G L I R (配列番号196)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はQ G S P S A D T L K A G L I Rである) (配列番号197) ;

4 : X 4 - P D T L K - Z 4、ここで、X 4は配列S S A P A (配列番号198)の0、1、2、3、4、または5アミノ酸であり、Z 4は配列E S Q A Q (配列番号199)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はS S A P A P D T L K E S Q A Qである) (配列番号200) ;並びに

6 : X 6 - S E T L K - Z 6、ここで、X 6は配列G E L P V (配列番号201)の0、1、2、3、4、または5アミノ酸であり、Z 6は配列A G Q L H (配列番号202)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はG E L P V S E T L K A G Q L Hである) (配列番号203)、
からなる群より選択することができる。

【0115】

更なる好ましい実施形態において、本発明のポリペプチドは一般式X I I :

A (R 1) R N D (配列番号204)

(ここで、R 1はS、Q、及びRからなる群より選択される)

のアミノ酸配列からなる。

【0116】

この一般式X I Iは、結晶構造において 7及び 8鎖を接続する高度に保存されたループに存在する。この領域は六量体界面に存在し、六量体の構築及び安定化に関与する。そのようなものとして、一般式X I IのペプチドはI V型コラーゲンの六量体相互作用の阻害に有用である。

【0117】

更なる好ましい実施形態においては、一般式X I Iによるポリペプチドは、A S R N D (1, 3, 5, 2) (配列番号205)、A Q R N D (4) (配列番号206)、及びA R R N D (6) (配列番号207)からなる群より選択される。

【0118】

更なる実施形態においては、配列番号205、206、及び207からなる群より選択される六量体ポリペプチドは、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意に、同じ 鎖から誘導される0~5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者にさらに含むことができる。従って、そのようなポリペプチドは :

1及び 5 : X 1 - A S R N D - Z 1、ここで、X 1は配列N V C N F (配列番号208)の0、1、2、3、4、または5アミノ酸であり、Z 1は配列Y S Y W L (配列番号209)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はN V C N F A S R N D Y S Y W Lである) (配列番号210) ;

3 : X 3 - A S R N D - Z 3、ここで、X 3は配列D V C N F (配列番号211)の0、1、2、3、4、または5アミノ酸であり、Z 3は配列Y S Y W L (配列番号209)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はD V C N F A S R N D Y S Y W Lである) (配列番号212) ;

2 : X 2 - A S R N D - Z 2、ここで、X 2は配列D V C Y Y (配列番号213)の0、1、2、3、4、または5アミノ酸であり、Z 2は配列K S Y W L (配列番号214)の0、1、2、3、4、または5アミノ酸である(従って、完全長配列はD V C Y Y A S R N D K S Y W Lである) (配列番号215) ;

4 : X 4 - A Q R N D - Z 4、ここで、X 4 は配列 Q V C H Y (配列番号 2 1 6) の 0、1、2、3、4、または 5 アミノ酸であり、Z 4 は配列 R S Y W L (配列番号 2 1 7) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は Q V C H Y A Q R N D R S Y W L である) (配列番号 2 1 8) ; 並びに

6 : X 6 - A R R N D - Z 6、ここで、X 6 は配列 E V C H Y (配列番号 2 1 9) の 0、1、2、3、4、または 5 アミノ酸であり、Z 6 は配列 K S Y W L (配列番号 2 1 4) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は E V C H Y A R R N D K S Y W L である) (配列番号 2 2 0)、
 からなる群より選択することができる。

【 0 1 1 9 】

更なる好ましい実施形態においては、本発明のポリペプチドは一般式 X I I I :

(R 1) (R 2) (R 3) N (R 4) (配列番号 2 2 1)

(ここで、R 1 は Y 及び F からなる群より選択され ;

R 2 は Y 及び F からなる群より選択され ;

R 3 は A 及び S からなる群より選択され ; 並びに

R 4 は A、S、及び K からなる群より選択される)

のアミノ酸配列からなる。

【 0 1 2 0 】

この一般式 X I I I は、結晶構造において 7 ' 及び 8 ' 鎖を接続する高度に保存されたループに存在する。この領域は六量体界面に存在し、六量体の構築及び安定化に
 20 関与する。そのようなものとして、一般式 X I I I のペプチドは I V 型コラーゲンの六量体相互作用の阻害に有用である。

【 0 1 2 1 】

更なる好ましい実施形態においては、一般式 X I I I によるポリペプチドは、Y Y A N A (1) (配列番号 2 2 2)、Y Y S N S (3) (配列番号 2 2 3)、Y Y A N S (5) (配列番号 2 2 4)、Y Y A N K (2) (配列番号 2 2 5)、F F A N K (4) (配列番号 2 2 6) 及び Y F A N K (6) (配列番号 2 2 7) からなる群より選択される。

【 0 1 2 2 】

更なる実施形態においては、配列番号 2 2 2 ~ 2 2 7 からなる群より選択される六量体ポリペプチドは、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意に、同じ鎖から誘導される 0 ~ 5 アミノ酸をアミノ及びカルボキシル末端のいずれか
 30 または両者にさらに含むことができる。従って、そのようなポリペプチドは :

1 : X 1 - Y Y A N A - Z 1、ここで、X 1 は配列 R G T C N (配列番号 2 2 8) の 0、1、2、3、4、または 5 アミノ酸であり、Z 1 は配列 Y S F W L (配列番号 2 2 9) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は R G T C N Y Y A N A Y S F W L である) (配列番号 2 3 0) ;

3 : X 3 - Y Y S N S - Z 3、ここで、X 3 は配列 R G T C N (配列番号 2 2 8) の 0、1、2、3、4、または 5 アミノ酸であり、Z 3 は配列 Y S F W L (配列番号 2 2 9) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は R G T C N Y Y
 40 S N S Y S F W L である) (配列番号 2 3 1) ;

5 : X 1 - Y Y A N S - Z 2、ここで、X 1 は配列 R G T C N (配列番号 2 2 8) の 0、1、2、3、4、または 5 アミノ酸であり、Z 5 は配列 Y S F W L (配列番号 2 2 9) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は R G T C N Y Y A N S Y S F W L である) (配列番号 2 3 2) ;

2 : X 2 - Y Y A N K - Z 2、ここで、X 2 は配列 R G T C H (配列番号 2 3 3) の 0、1、2、3、4、または 5 アミノ酸であり、Z 2 は配列 Y S F W L (配列番号 2 2 9) の 0、1、2、3、4、または 5 アミノ酸である (従って、完全長配列は R G T C H Y Y A N K Y S F W L である) (配列番号 2 3 4) ;

4 : X 4 - F F A N K - Z 4、ここで、X 4 は配列 Q G T C H (配列番号 2 3 5) の 0 50

10

20

30

40

50

、1、2、3、4、または5アミノ酸であり、Z4は配列Y S F W L（配列番号229）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はQ G T C H F F A N K Y S F W Lである）（配列番号236）；並びに

6：X6 - Y F A N K - Z6、ここで、X6は配列R G T C H（配列番号233）の0、1、2、3、4、または5アミノ酸であり、Z6は配列Y S F W L（配列番号229）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はR G T C H Y F A N K Y S F W Lである）（配列番号237）、
からなる群より選択することができる。

【0123】

更なる実施形態において、本発明は、8'及び9'鎖の間に位置するIV型コラーゲン鎖NC1ドメイン配列の超可変領域から誘導される新規ポリペプチドを提供し、これらは結晶構造から単量体-単量体界面に存在するものと同定され、並びにSRPを含み、かつSRとSRPとの間での最適相互作用に適切な二次構造の付与に關与する。この実施形態において、これらのポリペプチドは、I E R S E M F K K P T（1）（配列番号238）、L N P E R M F R K P I（3）（配列番号239）、V D V S D M F S K P Q（5）（配列番号240）、I P E Q S F Q G S P S（2）（配列番号241）、V K A D L Q F S S A P A（4）（配列番号242）、及びV E E R Q Q F G E L P V（6）（配列番号243）からなる群より選択される配列の少なくとも7アミノ酸からなる。様々な実施形態において、これらのポリペプチドは、配列番号235 - 240からなる群より選択される配列の少なくとも8、9、10、11、または12アミノ酸からなる。

【0124】

更なる実施形態において、配列番号238～243からなる群より選択されるポリペプチドは、最適阻害活性に適切な二次構造の特長をポリペプチドに付与するため、任意に、同じ鎖から誘導される0～5アミノ酸をアミノ及びカルボキシル末端のいずれかまたは両者にさらに含むことができる。従って、そのようなポリペプチドは：

1：X1 - I E R S E M F K K P T - Z1、ここで、X1は配列F W L A T（配列番号244）の0、1、2、3、4、または5アミノ酸であり、Z1は配列P S T L K（配列番号181）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はF W L A T I E R S E M F K K P T P S T L Kである）（配列番号245）；

3：X3 - L N P E R M F R K P I - Z3、ここで、X3は配列F W L A S（配列番号246）の0、1、2、3、4、または5アミノ酸であり、Z3は配列P S T V K（配列番号182）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はF W L A S L N P E R M F R K P I P S T V Kである）（配列番号247）；

5：X1 - V D V S D M F S K P Q - Z2、ここで、X1は配列F W L A T（配列番号244）の0、1、2、3、4、または5アミノ酸であり、Z5は配列S E T L K（配列番号183）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はF W L A T V D V S D M F S K P Q S E T L Kである）（配列番号248）；

2：X2 - I P E Q S F Q G S P S - Z2、ここで、X2は配列F W L T T（配列番号249）の0、1、2、3、4、または5アミノ酸であり、Z2は配列A D T L K（配列番号184）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はF W L T T I P E Q S F Q G S P S A D T L Kである）（配列番号250）；

4：X4 - V K A D L Q F S S A P A - Z4、ここで、X4は配列F W L T T（配列番号249）の0、1、2、3、4、または5アミノ酸であり、Z4は配列P D T L K（配列番号185）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はF W L T T V K A D L Q F S S A P A P D T L Kである）（配列番号251）；並びに

6：X6 - V E E R Q Q F G E L P V - Z6、ここで、X6は配列F W L T T（配列番号249）の0、1、2、3、4、または5アミノ酸であり、Z6は配列S E T L K（配列番号183）の0、1、2、3、4、または5アミノ酸である（従って、完全長配列はF W L T T V E E R Q Q F G E L P V S E T L Kである）（配列番号252）、

からなる群より選択することができる。

【0125】

更なる実施形態において、本発明は、単量体 - 単量体相互作用（従って、ヘテロ三量体の構築）、及び/または三量体 - 三量体相互作用（従って、六量体の構築）の阻害に重要であるものと同定される複数の領域を含む他のポリペプチドを提供する。本発明のこの側面によるポリペプチドは以下を含む：

S Rに加えてC D S R間：

- 1 : F S T M P F L F C N I N N V C N F A (配列番号 2 5 3)
- 3 : F T T M P F L F C N V N D V C N F A (配列番号 2 5 4)
- 5 : F S T M P F M F C N I N N V C N F A (配列番号 2 5 5)
- 2 : F S T M P F L Y C N P G D V C Y Y A (配列番号 2 5 6)
- 4 : F S T L P F A Y C N I H Q V C H Y A (配列番号 2 5 7)
- 6 : F S T M P F I Y C N I N E V C H Y A (配列番号 2 5 8)

10

C D S R間に加えて連続六量体界面領域：

- 1 : P F L F C N I N N V C N F A S R N D (配列番号 2 5 9)
- 3 : P F L F C N V N D V C N F A S R N D (配列番号 2 6 0)
- 5 : P F M F C N I N N V C N F A S R N D (配列番号 2 6 1)
- 2 : P F L Y C N P G D V C Y Y A S R N D (配列番号 2 6 2)
- 4 : P F A Y C N I H Q V C H Y A Q R N D (配列番号 2 6 3)
- 6 : P F I Y C N I N E V C H Y A R R N D (配列番号 2 6 4)

20

S Rに加えてC D S R間、それに加えて連続六量体界面領域：

- 1 : F S T M P F L F C N I N N V C N F A S R N D (配列番号 2 6 5)
- 3 : F T T M P F L F C N V N D V C N F A S R N D (配列番号 2 6 6)
- 5 : F S T M P F M F C N I N N V C N F A S R N D (配列番号 2 6 7)
- 2 : F S T M P F L Y C N P G D V C Y Y A S R N D (配列番号 2 6 8)
- 4 : F S T L P F A Y C N I H Q V C H Y A Q R N D (配列番号 2 6 9)
- 6 : F S T M P F I Y C N I N E V C H Y A R R N D (配列番号 2 7 0)

C D S R内に加えて連続六量体界面領域：

- 1 及び 5 : P F I E C H G R G T C N Y Y (配列番号 2 7 1)
- 3 : P F L E C H G R G T C N Y Y (配列番号 2 7 2)
- 2 : P F I E C N G G R G T C H Y Y (配列番号 2 7 3)
- 4 : P F L E C Q G R Q G T C H F F (配列番号 2 7 4)
- 6 : P F I E C S G A R G T C H Y F (配列番号 2 7 5)

30

S R P / 可変領域に加えて連続六量体界面：

- 1 : I E R S E M F K K P T P S T L K A G (配列番号 2 7 6)
- 3 : L N P E R M F R K P I P S T V K A G (配列番号 2 7 7)
- 5 : V D V S D M F S K P Q S E T L K A G (配列番号 2 7 8)
- 2 : I P E Q S F Q G S P S A D T L K A G (配列番号 2 7 9)
- 4 ; V K A D L Q F S S A P A P D T L K E S (配列番号 2 8 0)
- 6 : V E E R Q Q F G E L P V S E T L K A G (配列番号 2 8 1)

40

特異的単量体 - 単量体阻害剤に加えてS R：

- 1 : G S C L R K F S T M (配列番号 2 8 2)
- 3 : G S C L Q R F T T M (配列番号 2 8 3)
- 5 : G S C L R R F S T M (配列番号 2 8 4)
- 2 : G S C L A R F S T M (配列番号 2 8 5)
- 4 : G S C L P V F S T L (配列番号 2 8 6)
- 6 : G S C L P R F S T M (配列番号 2 8 7)

単量体 - 単量体阻害剤に加えてS R、加えてC D S R間、加えて六量体界面

- 1 : L R K F S T M P F L F C N I N N V C N F (配列番号 2 8 8)
- 3 : L Q R F T T M P F L F C N V N D V C N F (配列番号 2 8 9)

50

5 : L R R F S T M P F M F C N I N N V C N F (配列番号 2 9 0)
 2 : L A R F S T M P F L Y C N P G D V C Y Y (配列番号 2 9 1)
 4 : L P V F S T L P F A Y C N I H Q V C H Y (配列番号 2 9 2)
 6 : L P R F S T M P F I Y C N I N E V C H Y (配列番号 2 9 3) 。

【 0 1 2 6 】

別の側面において、本発明は、血管新生、血管新生介在疾患、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、並びに基底層構築を阻害するための方法であって、それらを必要とする被検体に血管新生、血管新生介在疾患、腫瘍増殖、腫瘍転移、内皮細胞接着及び/または増殖、並びに基底層構築の阻害に有効な量の、本発明のポリペプチドの1種類以上、そのようなポリペプチドに対する抗体、またはそれらの医薬組成物を投与することを含む方法を提供する。

10

【 0 1 2 7 】

「血管新生介在疾患」は望ましくない血管新生の付随を伴う疾患及び状態を指し、これには、これらに限定されるものではないが、固形及び血液由来の腫瘍、糖尿病性網膜症、関節リウマチ、網膜新血管新生、脈絡膜新血管新生、黄斑変性症、角膜新血管新生、未熟児の網膜症、角膜移植片拒絶反応、新生血管緑内障、水晶体後線維増殖症、流行性角結膜炎、ビタミンA欠乏症、コンタクトレンズ疲労、アトピー性角膜炎、上方輪部角結膜炎、翼状片乾燥角膜炎、s o g r e n s (シェーグレン症?)、赤鼻、p h y l e c t e n u l o s i s、梅毒、マイコバクテリア感染、脂質変性、化学的熱傷、細菌性潰瘍、真菌性潰瘍、単純ヘルペス感染、帯状疱疹感染、原生動物感染、カポジ肉腫、モーレン潰瘍、テリエン周辺変性症、辺縁角質溶解、外傷、全身性狼瘡、多発性動脈炎、ウェゲナー類肉腫症、強膜炎、スティーブン・ジョンソン病、放射状角膜切除術、鎌状赤血球貧血、類肉腫、弾性線維性仮性黄色腫、パジェット病、静脈閉塞症、動脈閉塞症、頸動脈閉塞病、慢性ブドウ膜炎、慢性v i t r i t i s、ライム病、イールズ病、B e c h e t s (ベーチェット?) 病病、近視、視窩、スタルガルト病、扁平部炎、慢性網膜剥離、過粘稠度症候群、トキシプラズマ症、レーザー後合併症、線維性血管性組織の異常増殖、血管腫、O s l e r - W e b e r - R e n d u 病 (多発性肺動静脈瘻)、後天性免疫不全症候群、眼血管新生病、骨関節炎、慢性炎症、クローン病、潰瘍性大腸炎、乾癬、アテローム性動脈硬化、及び類天疱瘡が含まれる (米国特許第 5 , 7 1 2 , 2 9 1 号を参照) 。

20

【 0 1 2 8 】

これらのポリペプチド、またはそのようなポリペプチドに対する抗体は滅菌のような通常の医薬上の操作を施すことができ、及び/または従来の補助剤、例えば保存剤、安定化剤、湿潤剤、乳化剤、緩衝剤等を含むことができる。

30

【 0 1 2 9 】

投与のため、ポリペプチドまたはそのようなポリペプチドに対する抗体は、通常、指示される投与経路に適する1種類以上の補助剤と組み合わせる。ポリペプチドまたはそのようなポリペプチドに対する抗体は、通常、投与のため、ラクトース、スクロース、デンプン粉、アルカン酸のセルロースエステル、ステアリン酸、タルク、ステアリン酸マグネシウム、酸化マグネシウム、リン酸及び硫酸のナトリウム及びカルシウム塩、アラビアゴム、ゼラチン、アルギン酸ナトリウム、ポリビニルピロリドン、及び/またはポリビニルアルコールと混合することができ、並びに錠剤化またはカプセル化することができる。また代替物として、本発明のポリペプチドまたはそのようなポリペプチドに対する抗体は、生理食塩水、水、ポリエチレングリコール、プロピレングリコール、カルボキシメチルセルロースコロイド溶液、エタノール、コーン油、ラッカセイ油、綿実油、ゴマ油、トラガカントゴム、及び/または様々な緩衝液に溶解することができる。他の補助剤及び投与方法は、医薬技術分野において公知である。担体または希釈剤には、例えば、モノステアリン酸グリセロールまたはジステアリン酸グリセロール等を単独で、またはワックスもしくは当該技術分野において公知の他の材料を伴った、時間遅延材料が含まれ得る。

40

【 0 1 3 0 】

本発明のこの側面の実施において、用いられるポリペプチド、そのようなポリペプチドに

50

対する抗体、または医薬組成物の量または投与量範囲は、血管新生、血管新生介在疾患、腫瘍増殖、腫瘍転移、及び/または内皮細胞 - 細胞外基質相互作用を有効に阻害するものである。用いることができるポリペプチドの阻害性量は、一般には、約 $0.01 \mu\text{g}/\text{kg}$ / 体重 kg ~ 約 $10 \text{mg}/\text{kg}$ の範囲であり、好ましくは約 $0.05 \mu\text{g}/\text{kg}$ ~ 約 $5 \text{mg}/\text{kg}$ の範囲である。

【0131】

ポリペプチド、そのようなポリペプチドに対する抗体、またはそれらの医薬組成物は、経口、非経口、吸入スプレー、直腸、または局所を含むあらゆる適切な経路により、通常の薬学的に許容し得る担体、補助剤、及び賦形剤を含有する投与単位処方で投与することができる。ここで用いられる非経口という用語には、皮下、静脈内、動脈内、筋肉内、胸骨内、腱内、脊髄内、頭蓋内、胸腔内、輸液技術または腹腔内が含まれる。好ましい実施形態においては、ポリペプチドは静脈内または皮下投与される。

10

【0132】

ポリペプチド、そのようなポリペプチドに対する抗体、またはそれらの医薬組成物は固体形態（顆粒、粉末または座剤を含む）または液体形態（例えば、溶液、懸濁液、または乳剤）で作製することができる。本発明のポリペプチド及びそのようなポリペプチドに対する抗体は、様々な溶液に適用することができる。本発明による使用に適する溶液は無菌であり、十分な量のポリペプチドを溶解し、かつ目的とするように対して有害ではない。

【0133】

好ましい実施形態においては、IV型コラーゲンの2つ以上の領域を構築の阻害のために標的とするため、開示されるポリペプチド、そのようなポリペプチドに対する抗体、またはそれらの医薬組成物のうちから、1種類以上が用いられる。例えば、異なる六量体領域を標的とするペプチドを組み合わせて用いて、それらの阻害効果を高めることができる。代替法として、またはそれに加えて、単量体 - 単量体相互作用を標的とするペプチドを六量体構築を標的とするペプチドと組み合わせることで更なる阻害効果を提供することができる。ここでの教示に基づき、他の組合せは当該技術分野における熟練者の知識の範囲内にある。

20

【実施例】

【0134】

蛋白質精製及び結晶化。[(1)₂ . 2]₂ NC1六量体を、Pel - Freeze Biologicals (Rogers, アーカンソー州) (37) から購入したウシ眼レンズから単離した。簡潔に述べると、LBMは、1M NaCl及びプロテアーゼ阻害剤の存在下において、レンズの超音波処理により調製した(38)。NC1ドメインを完全長IV型コラーゲンから開裂させるため、LBM調製品を、細菌コラゲナーゼを用いて37で消化した。DE-52及びS-300カラムクロマトグラフィーを用いることによりNC1六量体を精製した。

30

【0135】

市販の疎行列キット (Hampton Research, Laguna Niguel, カリフォルニア州) を用いる初期結晶化スクリーニングを、濃縮蛋白質 (10 mg/ml) 及び懸滴蒸気拡散法を用いて行った。LBM NC1結晶は、一晩、10% (w/v) PEG 20K、0.1M Bicine緩衝液 (pH 9.0) 中、室温で、小クラスタとして成長する。微量接種手順を用いて、より低い蛋白質濃度を用いる類似条件下において回折品質の結晶を成長させた。これらの結晶は単斜P2₁空間群に属し、単位セルの寸法は室温で $a = 129.41$ 、 $b = 143.87$ 、 $c = 162.92$ 、及び $\beta = 91.3^\circ$ であり、非対称単位内に4つの六量体を有する。25% 2,4-メチルペンタンジオール (MPD) またはグリセロール中でのこれらの結晶の低温冷却は単位セルの収縮を生じる ($a = 127.16$ 、 $b = 139.57$; $c = 160.20$; $\beta = 91.3^\circ$)。

40

【0136】

構造決定及び微調整。最初の重原子浸漬は結晶化pHで行い、後にリン酸緩衝液で中性p

50

Hに切り替えた。2 mM LuCl_3 または K_2PtCl_6 を含有する合成母液中に浸漬された NC1 結晶は、その格子を、 $a = 79.79$ 、 $b = 137.20$ 、 $c = 126.69$ 、 $\beta = 90.3^\circ$ 及び非対称単位内に2つの六量体の寸法のより小さい単位セルに変形する。これらの結晶を2 mM LuCl_3 に一晩浸漬することによって定型的に新たな形態に変換し、更なる重原子浸漬に用いた。0.5 M KBr に1分間浸漬し、かつ冷 N_2 流内で瞬間凍結した1個の結晶を用いて、多波長異常回折 (MAD) データ・セットをピーク、屈曲及び2つの離れた波長で集めた (表1)。重原子浸漬スクリーニングは Stanford Synchrotron Radiation Laboratory (SSRL) のビームライン 1-5 及び 9-2 並びに Brookhaven National Laboratory の National Synchrotron Light Source (NSLS) のビームライン X8C で行った。この研究で用いた Br-MAD データ・セットは SSRL で収集し、HKL2000 スーツ (39) の DENZO 及び SCALEPACK プログラムを用いて処理した。Br 部位は SOLVE プログラム (40) を用いて位置決定し、37 の最高ピーク (> 6) を解像度 2.2 の反射の位相決定に用いた。得られる位相を RESOLVE (41) を用いる溶媒平坦化によって改良し、CCP4 スーツ (42) の FFT プログラムを用いて電子密度マップを算出した。2本の 1鎖及び1本の 2鎖のポリペプチド (鎖 A~C) を、TOM FRODO グラフィックスプログラム (43) を用いて追跡した。Br⁻ 部位から得られた非結晶学的対称 (「NC5」) 関係を用いて完全な非対称単位を生成した。最初に第2三量体 (鎖 D-F) を生成して1つの六量体を完成させ、次いで最初の六量体から第2の六量体 (鎖 G-L) を生成した。0.8856 (λ_4) で収集した 2.0 データ・セットを CNS プログラム (44) を用いるモデルの微調整に用い、データの 5% は R_{free} の監視のために取って置いた。初期モデルに、30.0 ~ 3.0 解像度範囲 ($R = 0.361$ 及び $R_{free} = 0.364$) における反射、次いで 10.0 ~ 2.5 解像度範囲 ($R = 0.326$ 及び $R_{free} = 0.287$) における模擬アニリング微調整を用いる剛体微調整を施した。モデル構築並びに位置及び温度パラメータの微調整の数回の反復サイクルにおいて、解像度を 2.0 まで徐々に拡張した。微調整の最終ラウンドの間、 $2F_o - F_c$ 及び $F_o - F_c$ マップ並びに水素結合基準を用いて、溶媒分子 (水及びグリセロール) 及び Br⁻ イオンを段階的に添加した。いくつかの側鎖の複数の配座異性体を最終ラウンドにおいてモデル化した。SETOR (45)、GRASP (46)、及び SURF NET (47) グラフィックスソフトウェアパッケージ、並びに CCP4 スーツの様々なユーティリティを用いて、構造を解析した。六量体界面は HBPLUS (48) 及び蛋白質-蛋白質相互作用ウェブ・サーバー (<http://www.biochem.ucl.ac.uk/bism/PP/server/>) を用いて解析した。

10

20

30

【表1】

表1. 結晶学的解析の要約

データ収集				
データセット	ピーク	屈曲	遠隔1	遠隔2
波長(Å)	0.9195	0.9197	0.9537	0.8856
解像度(Å)	2.1	2.1	2.15	2.0
実測反射	602,172	603,309	568,640	686,286
独自反射	159,617	159,667	149,817	184,445
完全性(%) [*]	98.3(90.9)	98.2(90.5)	98.7(95.1)	97.9(87.8)
R _{sym} (%) [†]	4.0(7.7)	3.0(6.7)	2.4(4.9)	3.4(8.6)
I/σ(I)	29.2(15.0)	33.0(18.2)	37.6(26)	30.5(13.1)
フェージング統計				
解像度範囲(Å)	50.0~2.2			
B _r 部位の数	33			
全体的なZスコア	127			
性能指数	0.67/0.76			
解析/分解				
解像度範囲(Å)	8.0~2.0			
反射の数(σ > 2)	166,448/8,789			
実働/試験				
R _{cryst} /R _{free} (%) [‡]	17.0/19.6			
RMS 偏差				
結合長さ(Å)	0.0051			
結合角(°)	1.29			

10

20

30

【0137】

^{*} 全体的な完全性が示され、最高解像度シェルにおける完全性が括弧内に示される。R_{sym} 及び I/σ(I) も同様の約束事に従う。[†] $R_{sym} = \sum_{h_i} |I(h_i) - \langle I(h_i) \rangle| / \sum_{h_i} I(h_i)$ 。[‡] データの5%は洗練から除外され、R_{free}の決定に用いられた。R_{cryst}はこれらの反射を含まない。両者の場合において、 $R = (\sum |F_o| - k \sum |F_c|) / \sum |F_o|$ (加重用の反射は適切に選択する)。

【0138】

(結果及び考察)

構造決定及び概要。 1及び2鎖で構成されるウシLBM NC1六量体は、非対称単位当たり六量体4つの単車空間群P2₁(A形態)で結晶化する。これは、それぞれ非対称単位内に2つの六量体及び1つの六量体を伴って結晶化する、マウスEHS腫瘍NC1(49)及びヒト胎盤NC1六量体(50)について報告される結晶形態とは異なる。予備回折データの強度統計は、LBM NC1結晶におけるc軸に沿った偽翻訳対称(pseudo-translation symmetry)の存在を示唆した。浸漬実験を用いる重原子誘導体についての徹底的な探索は成功しなかった。しかしながら、pH7.0でLuCl₃に浸漬した結晶は、非対称単位内に2つの六量体のみを有する同じ空間群(B形態)への結晶学的翻訳を生じる偽翻訳対称の結果として、格子をより小さい単位セルに変換した。LuCl₃に浸漬した結晶のMADデータは、おそらくは格子変換の原因である単一の弱い結合部位のため、有用な相情報をもたらさなかった。しかしながら、我

50

々は、このより小さい単位セルを、新たに示唆されたハロゲン化物を用いる短浸漬戦略（51、52）を含む更なる重原子スクリーニングに利用した。LuCl₃浸漬B形態結晶構造を、解像度2.0で、異常散乱体としてのBr⁻を溶媒平坦化と組み合わせて用いるMAD法によって決定した。データ収集、フェージング及び微調整統計が表1に示される。

【0139】

このマップをヒトNC1₁及び2配列に当てはめた（図2）が、これはいずれのウシ配列も入手できなかったためである。4種類の他の哺乳動物種の1及び2配列の各々が公知であり、それらはそれらの間で95%を上回る配列同一性を共有する。ヒト配列の95%を上回る残基が実験電子密度マップと一致する。ヒト配列とこのマップとの相違は、1鎖における残基Ile15Thr、Ser22Pro、Pro129Gln並びに2鎖におけるAsp96Glu、Glu97Asp、及びGly176Alaに見出された。配列は、コラーゲン領域の最後のGly-Xaa-Yaa反復の後の残基が両鎖において最初の残基として数えられるようにナンバー付けされている。2つの六量体における12本の鎖は、各々の三量体における1、1及び2の順に鎖ID A-Lが割り当てられている。このマップは全ての鎖のN末端の5-6残基及びC末端の2残基の乱れを示す。最終モデルは2つの六量体、36のBr⁻イオン、48個のグリセロール分子及び1139個の水分子を含む。微調整の最終R因子及びR_{free}は、それぞれ、0.168及び0.197である。90%を上回る残基がラマチャンドランマップの最も好ましい領域内に存在し、各々の三量体における第1鎖のArg76及びSer148、第2鎖のSer148並びに2鎖のArg75、Glu95及びAla145が、許容し得ない領域に存在する。一握りの残基のみが複数の立体配座をとる。非対称単位内の2つの六量体は類似し、結晶の接触による明らかな相違がない。鎖A~Fを含む六量体をこのモデルの説明に用いる。

10

20

【0140】

六量体の全体的な構造が図3に示される。六量体内の2つの三量体は界面（「赤道面」）の二重NC₅軸によって関連付けられ、三量体内の単量体は三重らせん軸（「極軸」）と一致する偽三重対称性によって関連付けられる。

【0141】

単量体位相。複数配列比較を用いる我々の以前の研究（22）によって予想されるように、NC1単量体は1鎖を優先して新規三次構造内に折り畳まれる（図4）。三量体における2本の1鎖は同一であり、2鎖は類似の全体構造を有する。1鎖の一方及び2鎖における214個の一致残基のC原子は0.9のRMS偏差と重なる。各々の鎖は2つの相同サブドメイン、N-及びC-サブドメインに分割することができる。これら2つのサブドメインは類似の位相で折り畳まれ、1鎖の2つのサブドメインの96個の一致残基のC原子は1.0のRMS偏差と重なる。12個の不変システイン残基は、各々のサブドメインに3つの、6つのジスルフィドを保存位置に形成する（図2及び5）。2つのサブドメイン間の主な相違は、6つのヒト配列において最も保存されない、N-サブドメインにPro86-Pro95を、及びC-サブドメインにIle196-Thr209を包含する領域に生じる（図1）。各々のサブドメインは2つのシート-三重らせん接合点に近接する3本鎖逆平行シート（I&I'）及び、六量体を構成する2つの三量体の間の相互作用の領域からなる、六量体界面に近接する6本鎖逆平行シート（II&II'）を有する（図4及び5）。シートIは、ポリペプチドの最初の半分に属する配列の3本の非近接鎖（1、10及び2）によって形成される。しかしながら、シートIIにおいては、4本の鎖（4、3、8、及び9）のみが配列の最初の半分に属し、残りの2本の鎖（6'及び7'）は配列の第2の半分の一部を形成する。従って、配列の第2半分からのヘアピン構造（「鎖内ドメインスワッピング領域」または「CDSR内」）はN-サブドメインにスワップして6本鎖シートを形成する。ポリペプチドの2つの半分は位相的に類似し、N-サブドメイン内の6本鎖シートに相当するC-サブドメイン内の領域は、単離された単量体構造において類似のシートを形成する

30

40

50

には2本の鎖を欠く。同様に、ドメイン・スワッピング相互作用に關与するC末端における6'-7'に対応するN末端半分における6-7ヘアピンは、単量体構造において外部に伸長する。これらの2つの特徴は、次項において説明される三量体組織の基礎を形成する。

【0142】

三量体組織：1 NC1ドメインの2本の鎖及び2 NC1ドメインの1本の鎖は偽三重分子対称性を有する三量体構造を形成する。各々の鎖は位相的に類似のサブドメインで構成されるため、偽六重対称性さえ存在する。三量体の位相図が図5に示される。三量体の構造は、約65の底部直径及び内径約12-14.0の中空コアを有するほぼ円錐形状である。これはコラーゲン三重らせんの直径とほぼ同じであり、3本の鎖の全てのN末端が円錐の頂点で一緒になり、そこで三重らせんコラーゲンドメインがNC1ドメインと連結する。三量体はいくつかの鎖間疎水性及び水素結合相互作用によって強固に充填される(表2)。一方の鎖のN-サブドメイン内の5つのセグメントの残基が第2鎖のC-サブドメイン内の7つのセグメントのものと接触し、三量体内の単量体-単量体相互作用の領域からなる「単量体-単量体界面」を構成する。最も重要な相互作用は1つのN-サブドメインセグメント及び2つのC-サブドメインセグメントに限定される(図1)。単量体-単量体相互作用には2つのレベルがあり、一方は「一般的三量体」構築に必須であり、他方は三量体内の単量体-単量体相互作用のNC1鎖特異性を決定する。

10

【表2】

表2. 三量体における単量体-単量体界面の比較。

20

界面パラメータ	$\alpha 1 A - \alpha 1 B$		$\alpha 1 b - \alpha 2$		$\alpha 2 - \alpha 1 A$	
	$\alpha 1 A$	$\alpha 1 B$	$\alpha 1 B$	$\alpha 2$	$\alpha 2$	$\alpha 1 A$
セグメント数	5	7	5	7	5	8
残基数	49	60	51	65	49	59
$\Delta ASA (\text{\AA}^2)$	2137	2182	2087	2066	1985	2044
極性/非極性原子(%)	40.1/59.9	24.5/75.5	44.3/55.7	32.5/67.5	39.9/60.1	24.8/75.3
水素結合						
M-M/M-S/S-S	9/8/5		11/8/12		9/9/3	

30

ΔASA 、界面溶媒接近可能面積；M、主鎖；S、側鎖

【0143】

三量体内には、以下の単量体-単量体界面が存在する： $1 A - 2 C$ ； $1 B - 2 C$ ；及び $1 A - 1 B$ 。六量体は2つのそのような三量体を含む；第2三量体における単量体-単量体界面は $1 D - 2 F$ ； $1 E - 2 F$ ；及び $1 D - 1 E$ である。

【0144】

一般的三量体：第1レベルで、単量体は互いに絡み合い、3Dドメインスワッピング相互作用によって三量体を形成する(図5及び6a)(53)。6本鎖シート(II')が、同じ鎖の2つの半分の内の鎖から形成されるN-サブドメイン内のシートIIと同様に、C-サブドメインにおいて2本の異なる鎖の鎖から形成される。これらのシートは1及び2鎖において区別不能である。従って、6つのシート(II/II')が存在し、6つのサブドメインの各々における1つがNC1三量体構造において終端閉鎖3Dドメインスワッピング相互作用を形成する。これらの6本鎖シートの各々は配列の半分の4本の鎖(4/4'、3/3'、8/8'、9/9')によって形成され、残りの2本の鎖(6/6'、7/7')は同じ鎖(6/7；CDSR間)または隣接する鎖(6'/7'；「鎖内ドメインスワッピング領域」もしくは「CDSR内」)の他方の半分による寄与を受ける。9を除く全ての鎖のアミノ酸配列は、種内及び種間に渡って、鎖において高度に保存される。周期的な様式で形成される6つの位相的に類似するシートは偽六重対称性の外観を三量体に付与する(図6a)。シートの各々において、最も外側の鎖(9/9')は六量体界面の赤道面に平行な表面上にあっ

40

50

て外環の一部を形成し、最も内側の鎖(4/4')はコア内を極軸、すなわち、偽三重軸に対してほぼ並行に走る。各々のシート内でのこれら2本の鎖の角度は約75°であり、それが右回りのねじれをもたらす。6つのシート全てからの4/4'鎖は、それらの間に主鎖水素結合が存在しないものの、直径約14Åの平行樽様コアを形成する(図6a)。しかしながら、これらのコア鎖は主鎖-側鎖水素結合により直接、または溶媒分子を介して安定化されている。4/4'鎖は疎水性及び親水性残基の混合物を有し、前者はコアを指向し、後者は隣接鎖を指向する。興味深いことに、4鎖は長鎖親水性アミノ酸を含み、それにより隣接鎖の4'鎖の主鎖原子とより直接的な水素結合が形成され、これはより強力な鎖間相互作用を示す。鎖内の4'と4との相互作用には主として溶媒分子が介在する。従って、6本鎖シートは、3Dドメインスワッピング相互作用及びコンパクトな樽様コア構造により、一般的三量体構造の組織において必須の構成成分である。しかしながら、それらは三量体の鎖特異的構築においては限られた役割しか果たし得ない。従って、CDSR内、CDSR間、及び4/4'ベースの樽様コアを標的とする化合物、例えば、これらの領域から誘導されるペプチドは、一般的な単量体-単量体相互作用の阻害に用いることができ、そしてひいては、三量体構築の阻害に用いることができる。

10

【0145】

三量体構造における鎖特異性：8'及び9'鎖を接続するループの配列は6本のヒト鎖全てにおいて最も可変の領域である(「超可変領域」と呼ぶ)。この超可変領域は、その一次配列において、結晶構造中の1及び2鎖において異なる二次構造としてそれぞれ自体を表す。それは全ての1鎖において短い3₁₀らせん(g2')('特異性領域パートナー'もしくは「SRP」; Glu200-Lys204(EMFKK))を形成するが、2鎖における対応する領域(Ser198-Gln200; SPQ)は拡張立体配座(p')をとり、隣接1B鎖における拡張構造('特異性領域'もしくは「SR」; p、Phe57-Met60; FSTM)と対を形成して短い平行シートを形成する(図8b)。1及び2に由来するSRの配列は同一である(FSTM)ことに注意しなければならない。これは全構造において唯一の平行シートであり、主として鎖で構成される。pの配列は6本の鎖全てにおいて高度に保存され、2鎖においては、1A鎖内に平行シートを形成するパートナーがないとしても、同じ拡張構造を形成する。従って、これらの2本の鎖の間の更なる主鎖水素結合相互作用は、2鎖に存在する拡張構造よりも1鎖における3₁₀らせん構造の存在のため、1B-2界面(すなわち、1のSRと2のSRPとの相互作用を含む)のみで見出され、2-1A(すなわち、1のSRPと2のSRとの相互作用を含む)または1A-1B(すなわち、1のSRと1のSRPとの相互作用を含む)界面においては見出されない。3つの界面での二次構造要素におけるこの相違に加えて、単量体-単量体界面での主鎖-側鎖及び側鎖-側鎖相互作用にも相違が存在する(図6b)。これも3つの界面での極性対非極性原子の異なる比に反映される(表2)。従って、SR、SRP、または超可変領域を標的とする化合物、例えば、これらの領域から誘導されるペプチドは、特異的単量体-単量体相互作用の阻害、そしてひいては、三量体構築の阻害に用いることができる。

20

30

【0146】

さらに、単量体-単量体界面内の個々の界面の組成を考えると、特異的三量体構築の好ましい阻害剤は1及び2において同一であるSRを標的とし、そしてひいては、そのような阻害剤は単量体-単量体界面内の各界面での相互作用を妨害し、そしてひいては三量体構築を阻害するものと予想される。1B-2界面に見られる更なるH結合相互作用に必要である2SRPを標的とする阻害剤も好ましい。

40

【0147】

Lys56(1B)の側鎖は、2鎖における平行シートに先行するループの主鎖と主鎖及びGln120(2)の側鎖の近接結合とに挟まれている。この緊密に固定された位置において、Lys56(1B)はIle194(2)のカルボニル及びAsp121(2)のカルボニルと2つの強力な水素結合を、並びにGln120(2)及

50

びGlu196(2)のカルボニルと2つのより弱い相互作用を形成する直線配座が想定される。2鎖の平行シートに対応する1様(すなわち、1/3/5ファミリー)領域は3₁₀らせんであり、これはより長い配列に及ぶ。従って、1A-1B界面において、Lys56(1A)は主鎖結合と完全に平行ではなく、これはIle196(1B)のカルボニル酸素と弱い水素結合のみを形成する、異なる回転異性体配座をとるより多くの空間をこのリジンにもたす。これは、2の親水性Asp121に代わる1鎖の疎水性Thr124の存在によっても影響を受けることがある。2-1A界面では、Arg55(2)が他の2つの界面における1鎖のLys56と同様の位置に入り、Ile196(1A)のカルボニルと強力な水素結合相互作用を1つ有する。Arg55/Ala54及びGly98/Glu95を含むアミノ酸配列における他の相違は界面での水素結合パターンに相違をもたらす。従って、Arg55(2)/Lys56(1)は最適1-2単量体-単量体相互作用に重要な残基であり、この領域を標的とする化合物、例えば、LRKF(配列番号294)(1)またはLARF(配列番号295)(2)を含むペプチドは特異的単量体-単量体相互作用の構築の障害に用いることができる。この領域はSRに先行するため、この領域をSRと組み合わせて、特異的単量体-単量体相互作用の複数の側面を妨害し、より長いペプチドを形成することができ、そしてひいては三量体構築の障害により効果を示す。

10

【0148】

さらに、領域Ile194-Glu196(2)、Ile196(1)及びGln120-Asp121(2)も最適1-2単量体-単量体相互作用に関与し、これらの領域を標的とする化合物、例えば、IPE(配列番号294)(2 184-196)、IER(配列番号295)(1 196-198)またはQD(配列番号296)(2 120-121)を含むペプチドは特異的単量体-単量体相互作用の構築の障害、そしてひいては三量体構築の障害に用いることができる。

20

【0149】

1B-2界面(すなわち、1のSRと2のSRPとの相互作用を含む)は最大数の接触残基、親水性原子の最高割合を有し、かつ他の単量体-単量体界面よりも多くの水素結合を含む(表2)。他方、埋もれている表面積は1A-1B界面(すなわち、1のSRと1のSRPとの相互作用を含む)で最大である。これらの観察から、1B-2界面は水素結合相互作用によって優先的形形成され、1A-1B界面はより強い疎水性力によって安定化される。

30

【0150】

界面での特異的相互作用に加えて、充填の考察も三量体での鎖化学量論の決定において重要な役割を果たし得る。1及び2NC1鎖は類似の三次構造に低RMS偏差で折り畳まれはするが、各々のNC1鎖における2つのサブドメインの相対方向は三重らせん接合点近傍で異なる。2鎖におけるN-サブドメインのThr13-Tyr30を含む領域は、1構造における類似領域の相対方向と比較して、2鎖におけるC-サブドメインのその等価領域Asp121-Tyr138からさらに離れている。三重らせん接合点近傍での2構造のより大きい幅は、仮定上の2ホモ三量体に充填されるとき、深刻な立体的衝突を生じる。しかしながら、たとえ弱い相互作用を有するとしても、仮定上のホモ三量体において3本の1鎖を収容させることが可能である。

40

【0151】

単量体-単量体相互作用を妨害するように設計されたペプチドを、好ましくは、そのような単量体-単量体構築が生じる細胞に送達することが好ましい。また代替法としては、それらのペプチドを、細胞によって分泌されている構築された三量体の破壊に用いることができる。

【0152】

六量体構築：IV型コラーゲン三量体は、ひとたび内質の内腔において形成されると、細胞外間隙に分泌され、そこでN及びC末端会合によって六量体、次いで超分子ネットワークへと構築される。NC1ドメインは、六量体構築につながるC末端二量体会合を決定す

50

ることにより、この組立において主要な役割を果たす。この項においては、結晶構造において観察されるそのような六量体構築に影響を及ぼす力を説明し、I V型コラーゲンネットワーク構築における特異性の理論的根拠を提供する。

【0153】

フットボール型六量体は2つの同一三量体で構成され、各々は前項において説明されるように2本の1鎖及び1本の2鎖を含む。緊密に絡み合った三量体によって形成される各々のプロトマー（すなわち、NC1ドメインを含む、完全なI V型コラーゲン三量体）は単一の存在と考えられ、それ故に六量体を他のホモ二量体蛋白質複合体と比較して解析することができる（43）。我々は、2つの三量体間の相互作用の強度を評価し、かつI V型コラーゲンネットワークにおける六量体構築を解析するため、六量体界面を定義するいくつかのパラメータを決定している（表3）。

【表3】

表3. NC1六量体における三量体-三量体相互作用を定義する界面パラメータの比較及び32のホモ二量体複合体について観察された平均。

界面パラメータ	NC1六量体	観察された平均(43) (32ホモ二量体)
$\Delta ASA (\text{\AA}^2)$	4173.1	1685.03
平坦性	1.91	3.46
環状性	0.87	0.71
分断化	18	5.22
100 \AA^2 当たりの水素結合	1.2	0.7
ギャップ指数	1.24	2.2

極性及び非極性原子のパーセンテージは、それぞれ45.5及び54.5である。

【0154】

ほとんどのホモ二量体と同様に、2つのNC1三量体は、赤道面に存在し、かつ個々の三量体内の偽三重対称軸に対して垂直である、二重NC5軸によって関連付けられる（図4）。この対称性の制約は、より効率的な充填に加えて、1様及び2様配列の界面残基におけるいくつかの相違によって部分的に影響を受けることがある。六量体界面は2つの三量体のほぼ平坦な表面によって形成され、平均面からの全ての六量体界面原子のRMS偏差は1.9である（図9a）。これは、最近の総説（43）において考察された32のホモ二量体についての平均平面値3.5よりも有意に低い。3つの単量体の各々の6つのセグメントによって形成される、三量体当たり合計で109残基の六量体界面はほぼ円形であり、平均面の長軸及び短軸の長さは、それぞれ、約69及び61と測定される。この平坦な円形六量体界面は三量体当たり約4400²の溶媒接近可能面積を覆い、これはより大きい界面を有するより大きい分子の観察と相関する（54）。そのような大きい界面は、疎水性及び親水性残基の両者を含む三量体間の強力な相互作用を促進する。六量体界面における極性（45.5%）及び非極性原子（54.5%）は、ほぼ等しい割合であり、これは六量体の安定化における両タイプの相互作用の重要性を強調する。

【0155】

考察は、ここまで、六量体界面の全体的な性質に焦点を当てていた。次に、六量体界面での個々の鎖間の相互作用をより詳細に解析する。一方の三量体の各々の単量体は他方の三量体の2つの単量体と接触し、これらは接触面積の程度及び水素結合の数に基づいて「主」及び「副」接触と呼ばれる。主接触を作製する2つの単量体は、六量体の変性実験（55）において用いられる用語と同様の意味で、「二量体」と呼ばれる。これらの2つの三量体の間の二重NC5は2本の1鎖によって形成される「ホモ二量体」を唯1つ生じ（図7A）、残りは1及び2鎖によって形成される2つの「ヘテロ二量体」である（図7A~B）。

10

20

30

40

50

【0156】

偽三重軸の周りでの一方の三量体の他方に対する 120° の回転は「全ホモ二量体」構造を生じる。そのような配置が不可能である理由は主として対称性の考察で説明することができる：対称性の破壊は、おそらくは相互作用がより少なく、かつ幾らかの好ましからざる接触を伴う、効率に劣る充填を生じる。界面での複雑な水素結合相互作用を理解するため、各々の単量体とその「主」及び「副」相互作用パートナーとの相互作用を研究することが必須である。このレベルで提示される複雑性は、相互作用を構造中の3つの領域：「主」接触の「コア」及び「外側」領域並びに「主-副接合部」に分解することによって、さらに単純化することができる。

【0157】

主接触のコア領域：3Dドメインスワッピング相互作用によって形成される2つの6本鎖シート、 II 及び II' は、三量体組織の場合のように、六量体構築の形成において重要な役割を果たす。六量体界面は、コアにおいて $3-4$ 及び $3'-4'$ を接続するターンによって占められている。これらのターンは、シート II/II' の残りの鎖と共に、多数の保存残基を、六量体界面での広範な水素結合相互作用のため位置付ける。2つの三量体内のコアターン(2つの等価サブドメインによる寄与で単量体当たり2つ)は、一方の三量体内の各ターンが他方の三量体内の2つのターンと接触するように、ねじれた立体配置で充填される。 N -サブドメイン内のターンは、第2($Asn39/Gln38$)及び第3位置($Glu40/39$)に親水性アミノ酸を含む I'/III' 型のものである。 C -サブドメインターンは1鎖においては II 型、2鎖においては II' 型のものであり、小疎水性アミノ酸、 $Ala149/146-Gly150/147-Ala151/Asp148$ 、膨隆を誘導する $Ala149$ 1または $Asp148$ 2を伴う。従って、 N -サブドメイン内のターンの親水性側鎖は水素結合に関与し、 C -サブドメイン内のターンの疎水性残基は疎水性相互作用に加えてペプチド面のスタッキング効果によって充填される(図7A)。 N -サブドメイン内の $Asn39(Gln38)$ 側鎖は C -サブドメインターン内の主鎖アミドと水素結合を形成するが、保存 $Glu40(39)$ は他方の三量体内の単量体鎖の N -及び C -サブドメインの間に侵入し、保存 $Gln37(36)$ の側鎖と水素結合を形成する。 $1-1$ 二量体内の $Glu40$ 残基は互いに強力な水素結合を形成し、これは $1-2$ 二量体においては失われている。ターン及び側鎖の充填はCPKモデルにおけるコア界面では緊密であるように見え、これは明確な水素結合相互作用に加えて、強力なファン・デル・ワールス相互作用を示す。従って、六量体界面での主接触のコア領域を標的とする化合物、例えばこれらの領域から誘導されるペプチドを、六量体構築の障害に用いることができる。例えば、 $3-4$ 接続領域または $3'-4'$ 接続領域を含むペプチドを、主接触のコア領域での六量体構築の障害に用いることができる。

【0158】

主接触の外側領域： $Arg179(177)$ に先行する配列可変性は $1-2$ (六量体)界面での潜在的H結合の数に影響を及ぼす。外側領域における相互作用は、 7 及び 8 、並びに $7'$ 及び $8'$ シートを接続する高度に保存されたループを含む。六量体の $1-1$ 主界面においては、一方の鎖の高度に保存された $Ala74-Asp78$ の5個の近接カルボニル酸素が対称セットの他方の鎖の側鎖 $Asn77$ 、 $Arg179$ 、及び $Tyr185$ と水素結合を形成する(図9c)。これらの側鎖は1及び2鎖の両者においても保存される。しかしながら、2配列における $Gly176$ の挿入及び $Asn174$ の置換は保存 $Asn78$ 及び $Arg177$ 残基の方向を変化させ、これが $1-2$ 界面においていくつかの水素結合を生じる。従って、六量体界面での主接触の外側領域を標的とする化合物、例えば、これらの領域から誘導されるペプチドを六量体構築の障害に用いることができる。例えば、配列ASRND(配列番号201)(1)もしくはYYANA(配列番号218)(1)、または他の鎖における対応配列を含むペプチドを、主接触の外側領域での六量体構築の障害に用いることができる。

【0159】

10

20

30

40

50

主 - 副接合部：主 - 副接合部は、一方の三量体からの2本鎖が、他方の三量体の2本鎖と接触する六量体界面の領域である。2つの型の接合部が存在し、一方は3本の1及び1本の2鎖を含み、他方は1及び2鎖の各々を2本含む。これらの2つの接合部における水素結合パターンは高度に保存される(図7C)。1-1及び2-2の両者は界面においてAsn187(185)-Tyr189(188)(NY Y)(配列番号297)水素結合対を形成する。これに加えて、Asn187(185)は対向する三量体からの別の鎖(上で論じられる主接触の外側領域内)のArg76(75)と水素結合の対を形成する。2本の異なる鎖からの残基を含む、Asn187(185)によって形成される複数の水素結合は、好ましくは、三量体 - 三量体界面を安定化する主要因子の1つである。従って、六量体界面での主 - 副接合部を標的とする化合物、例えば、これらの領域から誘導されるペプチドを六量体構築の障害に用いることができる。例えば、配列NY Y(配列番号288)(1)(例えば、ECHGRGTCNY Y(配列番号172))、もしくは他の鎖における対応配列を含み、これらの全てが六量体界面に存在する(及びCDS R内の大部分を含む)か、またはAS RND(配列番号201)(1(ARG76(75)残基を含む)、もしくは他の鎖における対応配列を含むペプチドを、主 - 副接合部での六量体構築の障害に用いることができる。従って、配列AS RND(配列番号201)を含むペプチドは、主接触の外側領域及び主 - 副接合部の両者での相互作用を妨害することにより、六量体構築を妨害することができる。同様に、CDS R内を標的とし、かつ配列「NY Y」(配列番号288)からの2個の追加Y残基を含むように伸長するペプチドは、六量体構築に加えて、三量体構築の障害に用いることができる。

10

20

【0160】

六量体界面に位置し、かつ六量体構築に重要であると信じられる他の残基には、(1)MSMAP(配列番号129)(残基91-951)/MMP(配列番号132)(2)、及び他の鎖における対応配列；(2)PSTLK(配列番号177)(1における残基208-212；9'鎖；2におけるADTLK(配列番号180))、及び他の鎖における対応配列；(3)FCNINNVCNFA(配列番号289)(1及び5 - -CDS R間と共通の広汎性を有する)、及び他の鎖における対応配列；

3：FCNVNDVCNF(配列番号298)

2：YCNPGDVCYY(配列番号299)

4：YCNIHQVCHY(配列番号300)

6：YCNINEVCHY(配列番号301)。

30

【0161】

従って、これらの配列を含むペプチド、またはそれらの一部は、六量体構築の障害に用いることができる。

【0162】

ジスルフィド結合：鎖間または鎖内？

ジスルフィド架橋はコラーゲン構築において繰り返されるテーマであり、三量体構造の安定化において重要な役割を果たすものと信じられる(11)。原繊維プロコラーゲンは、C-テロペプチドまたはC-プロペプチドのいずれかにおいてプロテインジスルフィドイソメラーゼによって触媒される鎖間ジスルフィド結合を形成するものと信じられる(56, Kiovu, 1987#343)。鎖間ジスルフィドはIV型コラーゲンのコラーゲン及びNC1ドメインの両者において形成されるものと提唱されている。コラーゲンドメインにおける鎖間ジスルフィドがプロトマー内で形成されてコラーゲン三重らせんを安定化するのに対して、NC1ドメインにおけるものはプロトマー間に生じてC末端でネットワークを安定化するものと信じられる。2つの異なるプロトマーからの類似鎖のNC1ドメイン間でのジスルフィド交換が、六量体構築における主要安定化力の1つであるとして提唱された(57)。変性条件下において、ヒト胎盤誘導NC1六量体は二量体及び単量体として解離した。これらの二量体は主としてジスルフィド架橋によって架橋することが示された。しかしながら、いくつかのBMから単離されたNC1六量体を比較するLangefeldらの後の研究(55)はどちらかといえば複雑な結果を明らかにした。胎盤

40

50

B M及び腎系球体B M N C 1六量体の結果は、前の観察、変性時に二量体として解離、と一致するが、L B M N C 1六量体は主として単量体として解離し、ジスルフィド架橋の不在が暗示された。L B M N C 1六量体の結晶構造はまさにそれ、全てのシステインが鎖内ジスルフィドに關与することを明らかにする。

【0163】

S i e b o l dら(57)は、1鎖におけるN-サブドメイン内のC y s 2 0 (2 0 ') - C y s 1 1 1 ' (1 1 1) 及びC y s 5 3 (5 3 ') - C y s 1 0 8 ' (1 0 8) 対(並びにC-サブドメイン内の同様の位置のもの)を含むジスルフィド交換が、各サブドメイン内で臭化シアンに基づいて合計4つのジスルフィド架橋を生じていると提唱した。結晶構造において観察されるジスルフィドの幾何位相学的配置は、そのような再配置の可能性が極端に離れていることを示唆する(図7A)。N C 1単量体内のジスルフィドは3段に配置され、C y s 2 0 - C y s 1 1 1 及びC y s 1 3 0 - C y s 2 2 5 は三重らせん接合部に近く、C y s 6 5 - C y s 7 1 及びC y s 1 7 6 - C y s 1 8 2 は界面に近く、並びにC y s 5 3 - C y s 1 0 8 及びC y s 1 6 4 - C y s 2 2 2 はその間に存在する。

1 A - 1 D二量体の単量体内の、ジスルフィド対C y s 2 0 - C y s 1 1 1 及びC y s 5 3 - C y s 1 0 8 は、それぞれ約70°及び50°離れている。従って、ジスルフィド交換の可能性は、あるとすれば、C y s 6 5 - C y s 7 1 及びC y s 1 7 6 - C y s 1 8 2 対のみに存在する。しかしながら2つの三量体のねじれ型配列は、1 A鎖のC y s 6 5 - C y s 7 1 対を、N-サブドメイン内のその対応物C y s 6 5 ' - C y s 7 1 ' よりも、1 D鎖のそのC-サブドメイン等価C y s 1 7 6 ' - C y s 1 8 2 ' 対へと近づける。1 A - 1 D二量体内のこれら2つの最も近いジスルフィド対は、互いから約16°にある。さらにより重要なことには、これらの鎖内ジスルフィドは3 Dドメインスワップヘアピン領域内に位置する。ジスルフィド交換がこれらの対の間で実際に可能であったならば、それは大きな立体配剤の変化を含む。「交換可能な」システイン残基を含むヘアピンのそのような変動は鎖間及び鎖内両方の3 Dドメインスワッピング相互作用を破壊し、このため三量体構造を不安定化する。これらの理由から、本構造内の2つのプロトマーに属する単量体間のジスルフィド架橋を予見することは困難である。我々はまた、プロトマー内ジスルフィドの可能性も試験したが、これも大きな立体配座の変化を必要とし、潜在的に3本の鎖のN末端を移動させてコラーゲン-N C 1連結に選択的に影響を及ぼす。プロトマー間ジスルフィド架橋を説明するには、全ての他のB MからのN C 1ドメインに対し、代替の立体配座が存在しなければならない。

【0164】

生物学的意義。非コラーゲンドメインに関する利用可能な結晶学的データは、ほとんど存在しない。唯一の利用可能な非コラーゲンドメインの構造は、エンドスタチンのものであり(58, 59)、これはX V I I I 及びX V型コラーゲンからの一本鎖の相同断片である。

【0165】

本研究は、I V型コラーゲン1, 2ネットワークの鎖化学量論の最初の明白な構造的基礎に加えて、I V型コラーゲンの鎖特異的構築の構造的基礎を提供する。N C 1単量体は新規三次構造内に折り畳まれ、(1)2, 2の末端閉鎖三量体が、独自3 Dドメインスワッピング相互作用を通じて組織化される。これらの特徴は、ドメインスワッピングに關与する領域の、全体的な配列類似性及び非常に高い配列同一性により、全ての種からの、全てのI V型コラーゲンネットワークにおいて保存されなければならない。鎖特異性は、構成鎖のN C 1ドメインの超可変領域の一次配列における相違によって決定され、これは単量体-単量体界面での異なる二次構造として表される。六量体構造は、ジスルフィド架橋の必要なしに、三量体-三量体界面での広範な疎水性及び親水性相互作用によって安定化される。L B M N C 1六量体の結晶構造及びいくつかのB MからのN C 1六量体の変性研究は、鎖間ジスルフィドによって架橋する六量体においては、代わりにの立体配座が存在しなければならないことを示唆する。いくつかのこれまで不明の酵素的プロセスが、同じアミノ酸配列の、異なる組織における異なる立体配座への折り畳みの原因であり得

10

20

30

40

50

る。

【0166】

(参照文献)

1. Timpl, R., 及び Brown, J. C. (1996) *Bioessays* 18 (2), 123 - 131
2. Weber, M. (1992) *Kidney International* 41, 620 - 628
3. Pihlajaniemi, T. (1996) in *Molecular Pathology 及び Genetics of Alport Syndrome* (Tryggvasson, K., 編) Vol. 117, pp. 46 - 79, Karger, Basel 10
4. Miner, J. (1999) *Kidney International* 56, 2016 - 2024
5. Prockop, D. J., 及び Kivirikko, K. I. (1995) *Ann. Rev. Biochem.* 64, 403 - 34
6. Myllyharju, J., 及び Kivirikko, K. I. (2001) *Ann Med* 33, 7 - 21
7. Kadler, K. (1994)
8. Bachinger, H. - P., Bruckner, P., Timpl, R., Prockop, D. J., 及び Engel, J. (1980) *Eur. J. Biochem.* 106, 619 - 632 20
9. Bachinger, H. - P., Fessler, L. I., Timpl, R., 及び Fessler, J. H. (1981) *J Biol. Chem.* 256, 13193 - 13199
10. Dolz, R., Engel, J., 及び Kulm, K. (1988) *Eur J Biochem* 178 (2), 357 - 66
11. McLaughlin, S. H., 及び Bulleid, N. J. (1998) *Matrix Biology* 16, 369 - 377
12. Lees, J. F., Tasab, M., 及び Bulleid, N. J. (1997) *EMBO J.* 16 (5), 908 - 916 30
13. Dion, A. S., 及び Myers, J. C. (1987) *J Mol Biol* 193 (1), 127 - 43
14. Rosenbloom, J., Endo, R., 及び Harsch, M. (1976) *J. Biol. Chem.* 251, 2070 - 2076
15. Schofield, D. J., Uitto, J., 及び Prockop, D. J. (1974) *Biochemistry* 13, 1801 - 1806
16. Uitto, V., Uitto, J., 及び Prockop, D. J. (1981) *Arch. Biochem. Biophys.* 210, 445 - 454
17. Boutaud, A., Borza, D. - B., Bondar, O., Gunwar, S., Netzer, K. - O., Singh, N., Ninomiya, Y., Sado, Y., Noelken, M. E., 及び Hudson, B. G. (2000) *J. Biol. Chem.* 275, 30716 - 30724 40
18. Borza, D. B., Bondar, O., Ninomiya, Y., Sado, Y., Naito, I., Todd, P., 及び Hudson, B. G. (2001) *J Biol Chem* 276 (30), 28532 - 40.
19. Hudson, B. G., Reeder, S. T., 及び Tryggvasson, K. (1993) *J Biol Chem* 268 (35), 26033 - 6
20. Timpl, R., Wiedemann, H., van Delden, V., Urthmayr, H., 及び Kuhn, K. (1981) *Eur J Biochem* 120 (2), 203 - 11 50

21. Zhou, J., Ding, M., Zhao, Z., 及び Reeder, S. T. (1994) *J. Biol. Chem.* 269, 13193 - 13199
22. Netzer, K. O., Suzuki, K., Itoh, Y., Hudson, B. G., 及び Khalifah, R. G. (1998) *Protein Sci* 7 (6), 1340 - 51
23. Fowler, S. J., Jose, S., Zhang, X., Deutzmann, R., Sarras, M. P., Jr., 及び Boot Handford, R. P. (2000) *J Biol Chem* 275 (50), 39589 - 99.
24. Boute, N., Exposito, J. Y., Boury-Esnault, N., Vacelet, J., Noro, N., Miyazaki, K., Yoshizato, K., 及び Garrone, R. (1996) *Biol Cell* 88 (1-2), 37 - 44
25. Guo, X. D., 及び Kramer, J. M. (1989) *J Biol Chem* 264 (29), 17574 - 82.
26. Sibley, M. H., Johnson, J. J., Mello, C. C., 及び Kramer, J. M. (1993) *J Cell Biol* 123 (1), 255 - 64.
27. Blumberg, B., MacKrell, A. J., 及び Fessler, J. H. (1988) *J Biol Chem* 263 (34), 18328 - 37.
28. Exposito, J. Y., D' Alessio, M., Di Liberto, M., 及び Ramirez, F. (1993) *J Biol Chem* 268 (7), 5249 - 54.
29. Gunwar, S., Ballester, F., Noelken, M. E., Sado, Y., Ninomiya, Y., 及び Hudson, B. G. (1998) *J Biol Chem* 273 (15), 8767 - 75
30. Zhang, X., Hudson, B. G., 及び Sarras, M. P., Jr. (1994) *Dev Biol* 164 (1), 10 - 23
31. Guo, X., Johnson, J. J., 及び Kramer, J. M. (1991) *Nature* 349, 707 - 709
32. Sibley, M. H., Graham, P. L., von Mende, N., 及び Kramer, J. M. (1994) *EMBO J.* 13, 3278 - 3285
33. Kashtan, C. E., 及び Michael, A. F. (1993) *Am. J. Kid. Dis.* 22, 627 - 640
34. Kashtan, C. E., 及び Michael, A. F. (1996) *Kidney Int* 50, 1445 - 1463
35. Cosgrove, D., Meehan, D. T., Grunkemeyer, J. A., Kornak, J. M., Sayers, R., Hunter, W. J., 及び Samuelson, G. C. (1996) *Genes Dev* 10 (23), 2981 - 92.
36. Miner, J. H., 及び Sanes, J. R. (1996) *J Cell Biol* 135 (5), 1403 - 13.
37. Gunwar, S., Noelken, M. E., 及び Hudson, B. G. (1991) *J Biol Chem* 266 (21), 14088 - 94
38. Peczon, B. D., McCarthy, C. A., 及び Merritt, R. B. (1982) *Exp. Eye. Res.* 35, 643 - 651
39. Otwinowski, Z., 及び Minor, W. (1997) *Methods in Enzymology* 276, 307 - 326
40. Terwilliger, T. C., 及び Berendzen, J. (1997) *Acta Crystallogr.* D55, 849 - 861
41. Terwilliger, T. C. (2000) *Acta Crystallog* 50

- r D 56 (Pt 8), 965 - 72 .
- 42 . Dodson , E . J . , Winn , M . , 及び Ralph , A . (1997) Methods in Enzymology 277 , 620 - 633
- 43 . Jones , S . , 及び Thornton , J . M . (1996) Proceedings of The National Academy of Science (U . S . A) 93 , 13 - 20
- 44 . Brunger , A . T . , Adams , P . D . , Clore , G . M . , DeLano , W . L . , Gros , P . , Grosse - Kunstleve , R . W . , Jiang , J . S . , Kuszewski , J . , Pannu , N . S . , 及び al . , e . (1998) Acta Crystallogr . D 54 , 905 - 921 10
- 45 . Evans , S . V . (1993) J Mol . Graphics 11 , 134 - 138
- 46 . Nicholls , A . , Sharp , K . A . , 及び Honig , B . (1991) Proteins 11 , 281 - 296
- 47 . Laskowski , R . A . (1995) J . Mol . Graph . 13 . , , 323 - 330
- 48 . McDonald , I . K . , 及び Thornton , J . M . (1994) J . Mol . Biol . 238 , 777 - 793 .
- 49 . Timpl , R . , Oberbaumer , I . , von der Mark , H . , Bode , W . , Wick , G . , Weber , S . , 及び Engel , J . (1985) Ann NY Acad Sci 460 , 58 - 72 20
- 50 . Stubbs , M . , Summers , L . , Mayr , I . , Schneider , M . , Bode , W . , Huber , R . , Ries , A . , 及び Kuhm , K . (1990) J Mol Biol 211 , 683 - 684
- 51 . Dauter , Z . , 及び Dauter , M . (1999) J . Mol . Biol . 289 , 93 - 101
- 52 . Dauter , Z . , Dauter , M . , 及び Rajashankar , K . R . (2000) Acta Crystallogr . D 56 , 232 - 237
- 53 . Schlunegger , M . P . , Bennett , M . J . , 及び Eisenberg , D . (1997) Advances in Protein Science 30 50 , 61 - 132
- 54 . Jones , T . A . (1978) J . Appl . Crystallogr . 11 , 268 - 272
- 55 . Langeveld , J . P . , Wieslander , J . , Timoneda , J . , McKinney , P . , Butkowski , R . J . , Wisdom , B . J . , Jr . , 及び Hudson , B . G . (1988) J Biol Chem 263 (21) , 10481 - 8
- 56 . Uitto , J . , 及び Prockop , D . J . (1973) Biochem . Biophys . Res . Commun . 55 , 904 - 911
- 57 . Siebold , B . , Deutzmann , R . , 及び Kuhn , K . (1988) Eur J Biochem 176 (3) , 617 - 24 40
- 58 . Hoheneister , E . , Sasaki , T . , Olsen , B . R . , 及び Timpl , R . (1998) Embo J . 17 , 1656 - 1664
- 59 . Sasaki , T . , Larsson , H . , Tisi , D . , Claesson - Welsh , L . , Hoheneister , E . , 及び Timpl , R . (2000) J . Mol . Biol . 301 , 1179 - 1190
- 60 . Petitclerc , E . , Boutaud , A . , Prestayko , A . , Xu , J . , Sado , Y . , Ninomiya , Y . , Sarras , M . P . , Jr . , Hudson , B . G . , 及び Brooks , P . C . (2000) J Biol Chem 275 (11) , 8051 - 61 50

61. Laskowski, R. A., MacArthur, M. W., Moss, D. S., 及び Thornton, J. M. (1993) J. Appl. Cryst. . 26, 283 - 291

62. Barton, G. J. (1993) Prot. Eng. 6, 37 - 40

【図面の簡単な説明】

【0167】

【図1】 1様(1、3、及び5)及び 2様(2、4、及び6)ファミリーにグループ分けされる、6本のヒト NC1鎖の配列比較。システム対鎖内ジスルフィドは、下部に同一番号が付されている。三量体-三量体界面を形成する6つのセグメントには囲みが付され、単量体-単量体での3つの主要セグメントは、より大きなフォントサイズで強調されている。一般的及び特異的相互作用を形成する最も重要なセグメントは、下部に暗影付きのバーでそれぞれ特定されている。

10

【図2】(a) 1鎖及び(b) 2鎖。結晶構造に基づいて二次構造要素が割り当てられている。 1及び 2構造の両者は、鎖 1 - 10及び 1' - 10'並びに 3_{10} らせん g_1 及び g_1' を含む。二次構造における相違は、2つの配列における同じ領域での、 1における 3_{10} らせん及び 2における鎖 p'である。 2鎖の p'鎖のパートナーは、2本の 1鎖のうち一方である。 2における対応領域及び他方の 1鎖は、拡張された構造である。これらの領域は囲みによって印付けられている。二次構造はPROCHECK(61)からのものであった。

【図3】推定NC1六量体構造の立体図。三量体-三量体界面(「赤道面」)、コラーゲン三重鎖接合部、及び偽三重軸もしくは三重らせん軸(「極軸」)が特定されている。これらの2つの三量体は、極軸と紙面に対し垂直な、二重NC5軸によって関連付けられている。この図及び図5、8、9及び10bは、SETOR(45)を用いて作成した。

20

【図4】(a)六量体中の 1単量体構造の図示。4つの -シート領域がI、II、II'及びIIとして特定され、3つの短い 3_{10} らせんも示される。

【図5】鎖間及び鎖内3Dドメイン・スワッピング相互作用(一般的構築体)及び異なる二次構造要素との鎖界面(特異的構築体)を示す、NC1三量体の位相図。二次構造要素は 1A鎖についてのみ表示される。N-サブドメインにおける -シート、I及びII、並びにC-サブドメインにおけるI'及びII'が同定されている。各々のサブドメインは10本の鎖(1 - 10及び 1' - 10')及び2本の短い 3_{10} (g_1 及び g_2')らせんを有する。加えて、3つの界面に異なる二次構造が存在する - 1B - 2界面の平行 -シート(p - p')並びに、 1A - 1B及び 2 - 1A界面での 3_{10} らせん(g_1')及び拡張構造。

30

【図6】a)三量体における一般的相互作用。鎖間及び鎖内3Dドメイン・スワッピング相互作用によって形成される6本鎖 -シートは三量体組織における主要な力を形成する。サブドメインに属するこれらのシートはそのような相互作用を強調するために囲み内に示される。円の内側に示される中央樽様コアも、この骨格のパッキング及び安定化において役割を果たす。(b)3つの界面での独自二次構造及び顕著な側鎖相互作用が示される。 1b - 2界面は他の界面よりも多数の水素結合を有する。

【図7】三量体-三量体界面。 1 - 1及び 1 - 2二量体の「コア」(図7A)、「外側」(図7B)及び主-副接合部(図7C)の界面における、必須水素結合相互作用の比較(詳細は本文を参照)。

40

【 図 2 】

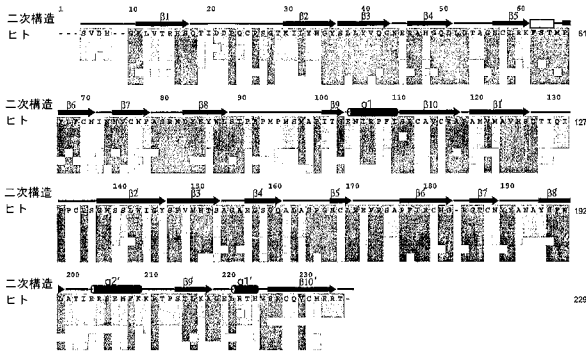


FIGURE 2a

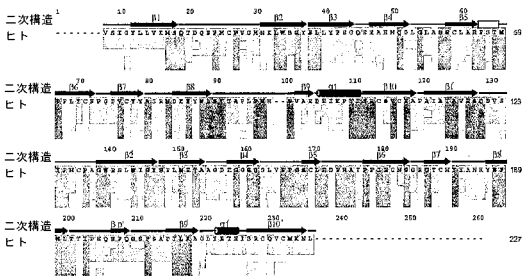


FIGURE 2b

【 図 3 】

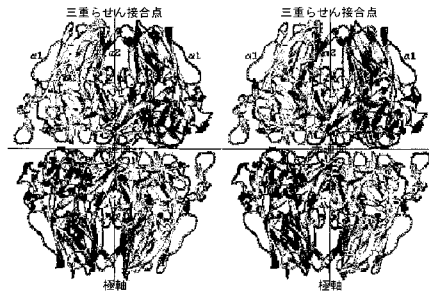


FIGURE 3

【 図 6 】



FIGURE 6a

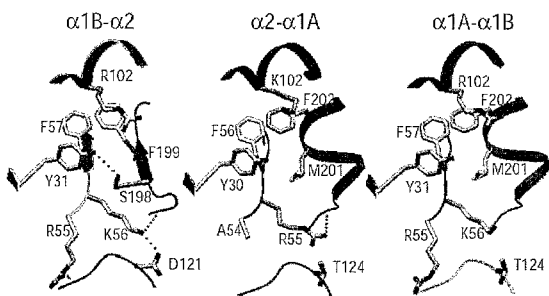


FIGURE 6b

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 February 2003 (13.02.2003)

PCT

(10) International Publication Number
WO 03/012122 A2

(51) International Patent Classification: C12Q 3901 Rainbow Blvd., Kansas City, KS 66160 NC 27703 (US).

(21) International Application Number: PCT/US02/23763

(22) International Filing Date: 26 July 2002 (26.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/398,523 27 July 2001 (27.07.2001) US
60/351,289 29 October 2001 (29.10.2001) US
60/366,854 22 March 2002 (22.03.2002) US
60/385,362 3 June 2002 (03.06.2002) US

(71) Applicants and
(72) Inventors: SUNDARAMOORTHY, Muirathinam [US/US]; Dept. of Biochemistry and Molecular Biology, University of Kansas Medical Center, 3901 Rainbow Blvd., Kansas City, KS 66160 (US). HUDSON, Billy [US/US]; Dept. of Biochemistry and Director, S-3223, Medical Center North, 1161 Twenty-first Avenue South, Nashville, TN 37232-2372, Kansas 66219 (US).

(74) Agent: HARPER, David, S.; McDonnell Boehnen Hulbert & Berghoff, 300 South Wacker Drive, Suite # 3200, Chicago, IL 60606 (US).

(81) Designated States (national): A1, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI,

[Continued on next page]

(71) Applicant (for all designated States except US): UNIVERSITY OF KANSAS MEDICAL CENTER [US/US];



(54) Title: CRYSTALLIZED STRUCTURE OF TYPE IV COLLAGEN NC1 DOMAIN HEXAMER

(57) Abstract: The present invention provides a crystallized NC1 domain hexamer of Type IV collagen, and methods for making the crystal, wherein the NC1 domain hexamer is crystallized such that the three dimensional structure of the crystallized NC1 domain hexamer can be determined to a resolution of at least 3 Å or better. The present invention also provides a method for designing compounds to inhibit angiogenesis, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and/or basal lamina assembly, comprising analyzing the three dimensional structure of a crystallized Type IV collagen NC1 domain hexamer produced by the methods of the invention, and identifying and synthesizing compounds that target regions of the NC1 domain that have been identified by the analysis as being important for type IV collagen heterotrimer and hexamer assembly. The present invention also provides novel polypeptides designed by the rational drug design methods of the present invention, based on an analysis of the type IV collagen NC1 hexamer structure disclosed herein.

WO 03/012122 A2

888888 1 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995

WO 03/012122 A2 

GM, HR, HU, ID, IL, IN, IS, JP, KH, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MY, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NI, SN, TD, TG).

Published:

*without international search report and to be republished
upon receipt of that report*

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, FI,
ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK,

*For two-letter codes and other abbreviations, refer to the "Guide
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

WO 03/012122

PCT/US02/23763

Crystallized structure of Type IV Collagen NCI Domain Hexamer

CROSS REFERENCE

5 This application claims priority to U.S. Provisional Patent Application Serial Nos. 60/308,523 filed July 27, 2001; 60/351,289 filed October 29, 2001; 60/366,854 filed March 22, 2002; and 60/385,362 filed June 3, 2002.

STATEMENT OF GOVERNMENT INTEREST

10 This work was supported by Grants DK18381 and DK53763 from the National Institutes of Health, and thus the U.S. government may have certain rights in the invention.

FIELD OF THE INVENTION

15 The present invention relates to the fields of crystallography, molecular biology, protein chemistry, angiogenesis, tumor growth and metastasis, and basement membrane assembly

BACKGROUND OF THE INVENTION

20 The basement membrane (basal lamina) is a sheet-like extracellular matrix (ECM), which is a basic component of all tissues. The basal lamina provides for the compartmentalization of tissues, and acts as a filter for substances traveling between tissue compartments. Typically the basal lamina is found closely associated with an epithelium or endothelium in all tissues of an animal, including blood vessels and
25 capillaries. The basal lamina components are secreted by cells and then self assemble to form an intricate extra-cellular network. The formation of biologically active basal lamina is important to the development and differentiation of the associated cells.

Type IV collagen has been shown to be a major structural component of basement membranes, and consists of a family of six homologous α chains,
30 designated $\alpha 1(IV)$ through $\alpha 6(IV)$. Each α chain is characterized by a non-

WO 03/012122

PCT/US02/23763

collagenous (NC1) domain at the carboxyl terminus; a long, helical collagenous domain in the middle region; and a 7S collagenous domain at the amino terminus. (Martin, et. al., 1988, Adv. Protein Chem. 39:1-50; Gunwar, et. al. 1991, J. Biol. Chem. 266:14088-14094). Three α chains assemble into triple helical molecules, the
5 "heterotrimer." The heterotrimer, once formed in the endoplasmic lumen, is secreted into the extracellular space, where two such heterotrimers assemble into a hexamer via C-terminal interactions, and then into a supramolecular network through N-terminal associations. The NC1 domains play the dominant role in this assembly, by determining the C-terminal dimeric association, leading to hexamer assembly.

10 The chain composition, and thus the properties of type IV collagen networks, are influenced by two factors. First, the chain composition of networks is limited by chain availability: the six α chains show a tissue-specific expression pattern, with the $\alpha 1$ and $\alpha 2$ chains being ubiquitous, and the $\alpha 3$ - $\alpha 6$ chains having a more restricted tissue distribution. Second, the NC1 domain confers specificity to the chain-specific
15 assembly of networks. Thus, as yet unidentified recognition sequences must exist within the NC1 domain that direct the selection of chains to form triple helical protomers, and that direct triple helical protomers to form hexamers and, thus, collagen networks. While numerous type IV collagen hexamers are theoretically possible that differ in kind and α chain stoichiometry, only three have been identified:
20 $[\alpha 1_2\alpha 2]_2$, $[\alpha 3\alpha 4\alpha 5]_2$, and $[(\alpha 1_2\alpha 2)(\alpha 5_2\alpha 6)]$.

Angiogenesis, the process of formation of new blood vessels, plays an important role in physiological processes such as embryonic and postnatal development, as well as in wound repair. Formation of blood vessels can also be induced by pathological processes involving inflammation (e.g., diabetic retinopathy
25 and arthritis) or neoplasia (e.g., cancer) (Folkman, 1985, *Perspect. Biol. Med.*, 29, 10). Neovascularization is regulated by angiogenic growth factors secreted by tumor or normal cells as well as by the composition of the extracellular matrix and the activity of endothelial enzymes (Nicosia and Ottinetti, 1990, *Lab. Invest.*, 63, 115).

A common feature of all solid tumor growth is the requirement for a blood
30 supply. Therefore, numerous laboratories have focused on developing anti-angiogenic compounds based on growth factors and their receptors. While this approach has led to some success, the number of growth factors known to play a role an angiogenesis is large. Therefore, the possibility exists that growth factor antagonists may have only limited use in treating cancer, since tumors and associated

WO 03/012122

PCT/US02/23763

inflammatory cells likely produce a wide variety of factors that can induce angiogenesis.

In this regard, a strategy that targets a common feature of angiogenesis, such as endothelial cell adhesion to the extracellular matrix (ECM), might be expected to have a profound physiological impact on tumor growth in humans. This notion is supported by the fact that antagonists of specific ECM cell adhesion receptors such as $\alpha v\beta 3$ and $\alpha v\beta 5$ integrins can block angiogenesis. Furthermore, the $\alpha v\beta 3$ integrin is expressed most prominently on cytokine-activated endothelial and smooth muscle cells, and has been shown to be required for angiogenesis. (Varner et al., Cell Adhesion and Communication 3:367-374 (1995); Brooks et al., Science 264:569-571 (1994)). Based on these findings, a potentially powerful new approach to anti-angiogenic therapy is to specifically target critical regulatory domains within distinct ECM components.

Specific type IV collagen $\alpha(IV)$ NC1 domains have been demonstrated to be effective inhibitors of angiogenesis, tumor growth, tumor metastasis, cell binding to basement membranes, and assembly of Type IV collagen molecules (see, for example, U.S. Patent Nos. 5,691,182; 5,856,184; 6,361,994; and 6,358,735). Despite the above, it would be of significant value to the art to identify further compounds capable of inhibiting these processes.

It is therefore highly desirable to provide a method of deducing the crystal structure of type IV collagen NC1 domains, and of providing a method of using this structure to design compounds that inhibit assembly of the type IV collagen heterotrimer and/or the type IV collagen hexamer.

25 SUMMARY OF THE INVENTION

In one aspect, the present invention provides a crystallized NC1 domain hexamer of Type IV collagen, and methods for making the crystal, wherein the NC1 domain hexamer is crystallized such that the three dimensional structure of the crystallized NC1 domain hexamer can be determined to a resolution of at least 3 Å or better.

In another aspect, the present invention provides a method for designing compounds to inhibit angiogenesis, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and/or basal lamina assembly, comprising analyzing the

WO 03/012122

PCT/US02/23763

three dimensional structure of a crystallized Type IV collagen NC1 domain hexamer produced by the methods of the invention, and identifying and synthesizing compounds that target regions of the NC1 domain that have been identified by the analysis as being important for type IV collagen heterotrimer and hexamer assembly.

5 Such compounds can be used to inhibit angiogenesis, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and basal lamina assembly.

In another aspect, the present invention provides novel polypeptides designed by the rational drug design methods of the present invention, based on an analysis of the type IV collagen NC1 hexamer structure disclosed herein. As a result of the information available from the crystal structure, it is possible to predict individual NC1 domain sequences that are critical for assembly of the type IV collagen heterotrimer and/or hexamer. Thus, it is also possible to design therapeutic polypeptides that will interfere with those interactions, and to inhibit assembly of the type IV collagen heterotrimer and/or the type IV collagen hexamer. Such therapeutic polypeptides can be used to inhibit or disrupt type IV collagen assembly, and thus are useful to inhibit angiogenesis, angiogenesis-mediated disorders, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and basal lamina assembly.

Brief Description of the Figures

20 **Figure 1.** Alignment of six human α NC1 chains grouped as α 1-like (1, 3, & 5) and α 2-like (2, 4, & 6) families. The cysteine pairs intrachain disulfides are labeled with identical numbers at the bottom. Six segments that form the trimer-trimer interface are boxed and three major segments at the monomer-monomer are highlighted with larger font size. The most important segments forming generic and specific interactions are identified at the bottom with darkly shaded bars, respectively.

30 **Figure 2.(a)** α 1 chains and **(b)** α 2 chains. Secondary structural elements are assigned based on the crystal structure. Both α 1 and α 2 structures contain β -strands β 1- β 10 and β 1'- β 10' and a 3_{10} helices g1 and g1'. The differences in secondary structures are a 3_{10} helix in α 1 and β -strand β p' in α 2 at the equivalent regions in the two sequences. The partner of β p' strand of α 2 chain is in one of the two α 1 chains. The corresponding region in α 2 and the other α 1 chains are extended structures. These regions marked by boxes. The secondary structures were from PROCHECK(61).

WO 03/012122

PCT/US02/23763

Figure 3. Stereo diagram of deduced NCI hexamer structure. The trimer-trimer interface (“Equatorial Plane”), collagen triple helical junction, and pseudo 3-fold axis or triple helix axis (“Polar Axis”) are identified. The two trimers are related by a 2-fold NCS axis perpendicular to the polar axis and plane of the paper. This figure and Figs. 5, 8, 9 and 10b were made using SETOR (45).

Figure 4. (a) Illustration of $\alpha 1$ monomer structure in the hexamer. Four β -sheet regions are identified as I, II, II' and II and three short 3_{10} helices are also shown.

Figure 5. Topology diagram of NCI trimer depicting interchain and intrachain 3D domain swapping interactions (generic assembly) and chain interfaces with different secondary structural elements (specific assembly). The secondary structural elements are labeled only for $\alpha 1A$ chain. The β -sheets, I & II in the N-subdomain and I' & II' in the C-subdomain are identified. Each subdomain has 10 β -strands ($\beta 1$ - $\beta 10$ and $\beta 1'$ - $\beta 10'$) and two short 3_{10} ($g1$ and $g2'$) helices. Additionally there are distinct secondary structures at the three interfaces—a parallel β -sheet (βp - $\beta p'$) at $\alpha 1B$ - $\alpha 2$ interface and a 3_{10} helix ($g1'$) and extended structure at $\alpha 1A$ - $\alpha 1B$ and $\alpha 2$ - $\alpha 1A$ interfaces.

Figure 6. (a) Generic interactions in the trimer. Six-strand β -sheets formed by interchain and intrachain 3D domain swapping interactions form the major force in the trimer organization. The sheets belonging to subdomains are shown in boxes to highlight such interactions. Central β barrel-like core, shown inside the circle, also plays a role in packing and stabilizing this scaffold. (b) Unique secondary structures and prominent side chain interactions at the three interfaces are shown. The $\alpha 1B$ - $\alpha 2$ interface has more number of hydrogen bonds than the other interfaces.

Figure 7. Trimer-trimer interface. Comparison of essential hydrogen bonding interactions in the interface at “core” (Figure 7A), “outer” (Figure 7B) and major-minor junction (Figure 7C) for $\alpha 1$ - $\alpha 1$ and $\alpha 1$ - $\alpha 2$ dimers (see text for details).

Detailed Description of the Preferred Embodiments

Within this application, unless otherwise stated, the techniques utilized may be found in any of several well-known references such as: *Molecular Cloning: A Laboratory Manual* (Sambrook, et al., 1989, Cold Spring Harbor Laboratory Press), *Gene Expression Technology* (Methods in Enzymology, Vol. 185, edited by D. Goeddel, 1991. Academic Press, San Diego, CA), “Guide to Protein Purification” in *Methods in Enzymology* (M.P. Deutscher, ed., (1990) Academic Press, Inc.); *PCR*

WO 03/012122

PCT/US02/23763

Protocols: A Guide to Methods and Applications (Innis, et al. 1990. Academic Press, San Diego, CA), *Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed.* (R.I. Freshney. 1987. Liss, Inc. New York, NY), and *Gene Transfer and Expression Protocols*, pp. 109-128, ed. E.J. Murray, The Humana Press Inc., Clifton, N.J.).

5 Type IV collagens are synthesized and assembled as heterotrimers inside the cells, which are then secreted extracellularly where hexamer assembly, and subsequent basement membrane (basal lamina) assembly, occurs.

The present work has elucidated the structure of the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer. Knowledge of this structure has utility in the design of
10 compounds that can inhibit assembly of type IV collagen heterotrimers and hexamers, and thus are beneficial in the inhibition of angiogenesis, angiogenesis-mediated disorders, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and basal lamina assembly.

Knowledge of the structure of the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer
15 structure provided by the present invention also has utility in the design of compounds that promote heterotrimer and hexamer assembly by providing tools and reagents for increasing the understanding of type IV collagen assembly, and thus also of basal lamina/basement membrane structure and function in general.

In one aspect, the present invention is directed to the three-dimensional
20 structure of an isolated and purified type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 domain hexamer ("hexamer"), such that the three dimensional structure of the crystallized type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer can be determined to a resolution of 3.0 Å or better, preferably 2.2 Å or better, and most preferably 2.0 Å or better, and wherein the crystals are of space group P2₁, with an approximate a=129.41 Å; approximately b=143.87 Å; approximately c=162.92 Å; and approximate $\beta=91.3^\circ$ at room
25 temperature and 4 hexamers in the asymmetric unit. Alternatively, the crystal has an approximate a=127.16 Å; approximately b=139.57 Å; approximately c=160.20 Å; and approximate $\beta=91.3^\circ$ and 4 hexamers in the asymmetric unit. In a further alternative, the crystals may have an approximate a=79.79 Å; approximately b= 137.20 Å;
30 approximately c= 126.69 Å; and approximate $\beta=90.3^\circ$ at room temperature and 2 hexamers in the asymmetric unit.

In another aspect, the invention provides a method for crystallizing a type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer to a resolution of less than about 3.0 Å or better, preferably 2.2 Å or better, and most preferably 2.0 Å or better, wherein the type IV

WO 03/012122

PCT/US02/23763

collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer is present at a concentration of about 0.5 mg/ml to about 50 mg/ml, more preferably from about 1 mg/ml to about 15 mg/ml and most preferably about 10 mg/ml, and the crystallization takes place at 4°C to 32°C, more preferably from 10°C to 26°C, even more preferably at about 16°C to 24°C, and even more preferably 20°C, to thereby obtain crystals of space group P2₁. The crystals may have an approximate a=129.41 Å; approximate b=143.87 Å; approximate c=162.92 Å; and approximate $\beta=91.3^\circ$ at room temperature and 4 hexamers in the asymmetric unit. Alternatively, cryocooling of the crystals may yield a crystal with an approximate a=127.16 Å; approximate b=139.57 Å; approximate c=160.20 Å; and approximate $\beta=91.3^\circ$ and 4 hexamers in the asymmetric unit. In a further alternative, the crystals may have an approximate a=79.79 Å; approximate b= 137.20 Å; approximate c= 126.69 Å; and approximate $\beta=90.3^\circ$ at room temperature and 2 hexamers in the asymmetric unit.

The crystallization, in one embodiment, may occur using hanging drops and the vapor diffusion method over 10% (w/v) PEG 20K. Alternatively, other crystallization methods may be used. For instance, a temperature variation may be used to produce crystals, or crystallization in space may be used to improve resolution. The crystallization, in another embodiment, may occur over 20% PEG 3350. In addition, other chemicals can be used in the place of PEG 20K or 3350. For instance, organic chemicals (e.g. isopropanol), inorganic chemicals (e.g. $(\text{NH}_4)_2\text{SO}_4$, NaH_2PO_4), and other molecular weight PEG may be used. Further details of the method are as described below.

In a further aspect, the present invention provides a method for determining the three dimensional structure of the crystallized type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer, comprising the steps of crystallizing the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer as described above, and then analyzing the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer to determine its three dimensional structure. In a preferred embodiment, the analyzing is by x-ray diffraction. Data sets generated from the diffraction analysis can be analyzed using any appropriate software, including but not limited to the DENZO and SCALEPACK programs of the HKL2000 suite (39), the SOLVE program (40), the RESOLVE (41) program, and/or the FFT program of CCP4 suite (42). Tracing of the polypeptides from the resulting analysis can be accomplished using any suitable software, including but not limited to the TOM FRODO graphics program (43). The final structure analysis can be accomplished

WO 03/012122

PCT/US02/23763

using any appropriate software, including but not limited to SETOR(45), GRASP(46), and SURFNET(47) graphics software packages, various utility programs in the CCP4 suite, and HBPLUS(48) and protein-protein interaction web server (<http://www.biochem.ucl.ac.uk/bsm/PP/server/>).

5 By analyzing the three-dimensional structure of the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ hexamer, one of skill in the art can determine the critical sites for type IV collagen NC1 domain heterotrimer and hexamer assembly, as described below.

Another aspect of the invention is to use the three-dimensional structure of the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ hexamer to solve the three-dimensional structure of a different type IV collagen NC1 domain hexamer crystal, or crystal of a mutant, homologue or co-complex of type IV collagen NC1 domain hexamer.

A further aspect of this invention is to use the three-dimensional structure of type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ hexamer to design inhibitors of the assembly of heterotrimers and hexamers of type IV collagen, including the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer. These inhibitors may be used as therapeutics to inhibit undesired angiogenesis, angiogenesis-mediated disorders, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and basal lamina assembly. This embodiment comprises:

(a) obtaining crystals of an NC1 hexamer of type IV collagen, wherein the crystal comprises an $[(\alpha 1)_2, \alpha 2]_2$ NC1 hexamer of type IV collagen, wherein the crystal consists of space groups $P2_1$ with approximate $a =$ between 127.16 Å and 129.41 Å, $b =$ between 139.57 Å and 143.87 Å; $c =$ between 160.20 Å and 162.92 Å; $\beta = 91.3^\circ$, such that the three-dimensional structure of the crystallized NC1 domain hexamer can be determined to a resolution of 3 Å or better;

25 (b) analyzing the three-dimensional structure of the crystallized NC1 domain hexamer of type IV collagen; and

(c) designing a potential inhibitor of type IV collagen assembly that targets one or more regions of a type IV collagen NC1 α chain selected from the group consisting of:

- 30 (i) Inter-chain domain swapping region;
- (ii) Intra-chain domain swapping region;
- (iii) Specificity region;
- (iv) Specificity region partner;

WO 03/012122

PCT/US02/23763

- (v) Hexamer interface;
- (vi) Monomer-monomer interface; and
- (vii) Hypervariable region.

As used herein "target" or "targeting" refers to compounds that will interact with this region, via covalent or non-covalent means. The definitions of the various regions are discussed below.

As discussed above, the NC1 domains drive the selection process for type IV collagen chain assembly, and thus analysis of NC1 domain assembly correlates with type IV collagen assembly. Furthermore, given the high degree of homology of the different NC1 domains, analysis of the $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer crystal structure provides insights into the structure of other hexamer types, as well as inhibitors of such assembly.

As used herein, "inhibiting assembly of heterotrimers and hexamers of type IV collagen" means to inhibit initial assembly of such heterotrimers and/or hexamers, or to disrupt the assembly of already assembled heterotrimers and hexamers of type IV collagen NC1 domains. In a highly preferred embodiment, the therapeutic compounds identified herein inhibit the initial assembly of such heterotrimers and/or hexamers of type IV collagen NC1 domains.

The inhibitors can comprise peptides, or antibodies directed against peptides derived from the critical regions that would be expected to interfere with type IV collagen heterotrimer and/or hexamer assembly. Alternatively, small molecules that are identified based on their potential to inhibit such assembly. Electronic screening of large, structurally diverse compound libraries, such as the Available Chemical Directory (ACD) can identify new structural classes of such modulators that would be expected to interact with the identified critical regions. Additionally, knowledge of the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer structure permits "de novo design" of compounds to inhibit assembly of any type IV collagen NC1 domain heterotrimers and/or hexamers.

Potential inhibitors can be examined *in silico* through the use of computer modeling, using a docking program such as GRAM, DOCK, or AUTODOCK [Dunbrack et al., 1997, supra]. These procedures can include computer fitting of candidate compounds to the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer to predict how the shape and chemical structure of the candidate compound will interfere with assembly of the type IV collagen heterotrimer and/or hexamer. Computer programs

WO 03/012122

PCT/US02/23763

can also be used to estimate the attraction, repulsion, and steric hindrance of the candidate compound to the relevant binding site on the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ hexamer. Generally the tighter the fit (e.g., the lower the steric hindrance, and/or the greater the attractive force), the more potent the candidate compound will be, and the less likely that the candidate compound will induce significant side effects due to unwanted interactions with other proteins.

Potential small molecule inhibitors can be obtained, for example, by screening random peptide libraries produced, for example, in recombinant bacteriophage (Scott and Smith, *Science*, 249:386-390 (1990); Cwirla et al., *Proc. Natl. Acad. Sci.*, 87:6378-6382 (1990); Devlin et al., *Science*, 249:404-406 (1990)), or a combinatorial chemical library. Candidate compounds selected in this manner can be systematically modified by computer modeling programs until one or more promising candidate compounds are identified. Such analysis has been shown to be effective, for example, in the development of HIV protease inhibitors (Lam et al., *Science* 263:380-384 (1994); Wlodawer et al., *Ann. Rev. Biochem.* 62:543-585 (1993); Appelt, *Perspectives in Drug Discovery and Design* 1:23-48 (1993); Erickson, *Perspectives in Drug Discovery and Design* 1:109-128 (1993)).

Such computer modeling allows the selection of a finite number of rational chemical modifications, as opposed to the countless number of essentially random chemical modifications that could be made. Thus, the use of the three-dimensional structure disclosed herein, in conjunction with computer modeling, enables rapid screening *in silico*, which dramatically increases screening speed and efficiency.

Once such candidate compounds are identified, they are chemically synthesized, and their biological activity is assayed, as discussed below. For those compounds that show activity, they can be complexed with the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer crystal for further X-ray diffraction analysis to map the interactions of the compound with the crystal structure. The three-dimensional structure of the supplemental crystal can be determined by Molecular Replacement Analysis, which involves using a known three-dimensional structure as a search model to determine the structure of a closely related molecule or protein-ligand complex in a new crystal form. The measured X-ray diffraction properties of the new crystal are compared with the search model structure to compute the position and orientation of the protein in the new crystal. Using this approach, it is possible to use the structure of the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer disclosed herein to

WO 03/012122

PCT/US02/23763

solve the three-dimensional structures of any such type IV collagen hexamer or co-complex.

Functional Assays

5 Any assay that can be used to test the effect of the candidate compounds on the in vitro or in vivo assembly of type IV collagen heterotrimers and/or hexamers can be used to verify the efficacy of the candidate compounds identified by the methods of the invention. Furthermore, any assay that can be used to test the effect of the candidate compounds on angiogenesis, tumor growth, tumor metastasis, and
10 endothelial cell adhesion and/or motility can be used to verify their inhibitory activity. Such assays include, but are not limited to, the following.

Assembly assay

In one example, the methods employed are as described in Boutaud et al., JBC
15 275 (39):30716-30724 (2000). Native GBM hexamers are isolated by standard methods and dissociated by dilution (<50 $\mu\text{g/ml}$) into a solution of 50 mM formic acid buffered at pH 3.0 with Tris base. Under these conditions, complete dissociation to NC1 monomers and dimers occurs, as can be verified by HPLC or FPLC gel
20 filtration. The absence of salt from the buffer is optimal for complete hexamer dissociation. Reassembly of the dissociated NC1 domains is performed by changing the buffer to Tris-buffered saline (50mM Tris, pH 7.4, 150mM NaCl) by repeated dilution-concentration cycles. After incubating the NC1 domains at a concentration of about 1 mg/ml for 24 hours at room temperature, in the presence or absence of the
25 candidate compounds at a desired concentration(s), the reaction products are separated according to their molecular weights using gel filtration chromatography. Quantification of the relative amounts of the various species in the mixture is done by peak area analysis from the HPLC profiles. Hexamer assembly from purified $\alpha 1$ - $\alpha 6$ NC1 domains is carried out similarly.

In all experiments, the ratio of the NC1 domains in the association mixture is
30 preferably kept at 1:1. The isolated NC1 hexamers can subsequently be analyzed for composition by immunoprecipitation followed by Western blotting; for overall appearance (size and shape) by electron microscopy; and for molecular weight by sedimentation equilibrium ultracentrifugation.

WO 03/012122

PCT/US02/23763

In Vitro Effect on Angiogenesis

With modifications, the procedures of Nicosia and Ottinetti, (1990, Lab. Invest., 63, 115) and Nicosia, et. al. (1994, Exp. Biology, 164, 197-206) are utilized for experiments designed to test the effect of the drug candidates on angiogenesis under *in vitro* conditions. The model has been used to study the effects of growth factors and extracellular matrix molecules on the angiogenic response, and employs aortic ring cultures in three-dimensional collagen gels under serum-free conditions.

Experiments are performed with 1-3 month old Swiss Webster male mice. Following anesthesia, the thoracic aorta is excised under aseptic conditions and transferred to sterile MCDB 131 sterile growth medium (Clonetics, San Diego, CA) containing antibiotics. Fat is dissected away from the aorta and approximately six to eight 1 mm thoracic segments are obtained from each specimen. Segments are transferred to 48 well tissue culture plates. The wells of these plates are layered with 100 microliters of Matrigel™ (EHS basement membrane, Collaborative Biomedical Products, Bedford, MA) prior to transfer of the aortic segments. The Matrigel™ is diluted 1:1 with MCDB 131 growth medium prior to use. The segments are centered in the wells and an additional 100 microliters of Matrigel™ is then placed over the specimens. The aortic segments are therefore embedded in the basement membrane matrix. Each well then receives 300 microliters of MCDB 131 growth medium. The plates are placed in an incubator maintained at 37° C with 5% CO₂. Specimens are observed daily over a 7 day period. Newly growing microvessels are counted using an inverted phase microscope at various times during the culture period. To test for the effect of drug candidates on angiogenesis, the drug candidates are mixed with the Matrigel™ and with the MCDB 131 growth medium, and the growth of microvessels from the cultured tissue into the matrix is analyzed.

Subcutaneous fibrin implant angiogenesis

The drug candidates are injected intravenously into rats containing fibrin implants surgically placed subcutaneously, a modified version of the method described by Dvorak et al. (Lab. Invest. 57(6):673-686 (1987)). For example, rats are given tail vein injections of either control, or various concentrations of the drug candidates. The implants are then removed at appropriate times, and directly analyzed

WO 03/012122

PCT/US02/23763

using an inverted microscope. The analysis involved counting the number of blood vessels per implant that grow into the fibrin in the control and experimental group.

Chick embryo CAM angiogenesis assay

- 5 Angiogenesis is induced in the CAMs of 10 day old chick embryos with bFGF as described (Brooks et al., Cell 92:391-400 (1998)). Twenty four hours later, the embryos are systemically treated with various concentrations of the drug candidates, in a total volume of 100 μ l of sterile phosphate buffered saline (PBS). Two days later, the embryos are sacrificed and the filter discs and CAM tissues removed.
- 10 Angiogenesis is quantitated by counting the number of angiogenic blood vessel branch points in the confined area of the filter disc. The Angiogenic Index is defined as the number of branch points from experimental treatment minus control treatment.

Chick embryo tumor growth assay

- 15 Briefly, single cell suspensions of distinct tumor types are applied to the CAM of 10 day old chick embryos. The tumors may include, for example, CS-1 Melanoma cells, HT1080 human fibrosarcoma cells, and Hep-3 human epidermoid carcinoma cells. The embryos are injected systemically with varying concentrations of the drug candidates 24 hours later. The embryos are allowed to incubate for a total of 7 days,
- 20 at which time they are sacrificed. The resulting tumors are resected and wet weights determined compared to control.

Immobilized NCI domains support human endothelial cell adhesion

- In order for new blood vessels to form, endothelial cells must have the capacity to adhere and migrate through the ECM. Moreover, this endothelial cell-ECM interaction may facilitate signal transduction events required for new blood vessel formation. Therefore, the ability of drug candidates to support endothelial cell attachment can be assessed.
- 25

- Microtiter plates are coated with varying amounts of the drug candidates, followed by incubation with 1% bovine serum albumin (BSA) to block non-specific interactions. Endothelial cells, such as human ECV304 cells, are then allowed to attach to the immobilized polypeptides for varying time periods. Non-adherent cells are removed by washing and attached cells are quantified by measuring the optical density of crystal violet eluted from attached cells.
- 30

WO 03/012122

PCT/US02/23763

In vitro Endothelial Cell Migration

Invasive cellular processes, such as angiogenesis and tumor metastasis, also require cellular motility. Thus, the ability of the drug candidates to support human endothelial cell migration can be tested in vitro. These experiments are conducted essentially according to the methods in Brooks et al., J. Clin. Invest. 99:1390-1398 (1997).

In vivo Endothelial Cell Migration

The ability of the drug candidates to support human endothelial cell migration can be tested in vivo. For example, drug candidates can be tested in the metastatic Lewis lung mouse tumor model using a standard protocol which is considered to be a good model of both metastasis and angiogenesis of lung tumors. (See for example, Teicher et al., Anticancer Res. 18:2567-2573 (1998); Guibaud et al., Anticancer Drugs 8:276-282 (1997); Anderson et al., Cancer Res. 56:715-718 (1996)).

Drug candidates are administered intravenously once every 2 days for a desired number of doses starting one day after tumor inoculation. All animals are weighed twice a week throughout the study. Starting one day after the last treatment, 1 or more mice are periodically sacrificed from each control group to measure pulmonary tumor burden. The experiment is terminated when the lungs of control animals have sufficient tumor mass to provide meaningful evaluation. At that time, the lungs of all remaining animals are excised, weighed, and the number of tumor foci greater than 2 mm in diameter counted.

In another aspect, the present invention provides an inhibitor of type IV collagen assembly identified by any of the methods described above.

In another aspect, the present invention provides an inhibitor of one or more process selected from the group consisting of angiogenesis, tumor growth, tumor metastasis, endothelial cell adhesion, endothelial cell proliferation, and basal lamina assembly, identified by any of the methods described above.

In another aspect, the present invention provides novel polypeptides that can be used to inhibit or disrupt type IV collagen assembly, and thus are useful to inhibit angiogenesis, angiogenesis-mediated disorders, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and basal lamina assembly.

WO 03/012122

PCT/US02/23763

The term "polypeptide" is used in its broadest sense to refer to a compound of two or more subunit amino acids, amino acid analogs, or peptidomimetics. The subunits are linked by peptide bonds. The polypeptides described herein may be chemically synthesized or recombinantly expressed.

5 Preferably, the polypeptides of the present invention are chemically synthesized. Synthetic polypeptides, prepared using the well known techniques of solid phase, liquid phase, or peptide condensation techniques, or any combination thereof, can include natural and unnatural amino acids. Amino acids used for peptide synthesis may be standard Boc (N α -amino protected N α -t-butyloxycarbonyl) amino
10 acid resin with the standard deprotecting, neutralization, coupling and wash protocols of the original solid phase procedure of Merrifield (1963, J. Am. Chem. Soc. 85:2149-2154), or the base-labile N α -amino protected 9-fluorenylmethoxycarbonyl (Fmoc) amino acids first described by Carpino and Han (1972, J. Org. Chem. 37:3403-3409). Both Fmoc and Boc N α -amino protected amino acids can be obtained from Sigma,
15 Cambridge Research Biochemical, or other chemical companies familiar to those skilled in the art. In addition, the polypeptides can be synthesized with other N α -protecting groups that are familiar to those skilled in this art.

Solid phase peptide synthesis may be accomplished by techniques familiar to those in the art and provided, for example, in Stewart and Young, 1984, Solid Phase
20 Synthesis, Second Edition, Pierce Chemical Co., Rockford, Ill.; Fields and Noble, 1990, Int. J. Pept. Protein Res. 35:161-214, or using automated synthesizers. The polypeptides of the invention may comprise D-amino acids (which are resistant to L-amino acid-specific proteases in vivo), a combination of D- and L-amino acids, and various "designer" amino acids (e.g., β -methyl amino acids, C α -methyl amino acids,
25 and N α -methyl amino acids, etc.) to convey special properties. Synthetic amino acids include ornithine for lysine, fluorophenylalanine for phenylalanine, and norleucine for leucine or isoleucine.

In addition, the polypeptides can have peptidomimetic bonds, such as ester bonds, to prepare peptides with novel properties. For example, a peptide may be
30 generated that incorporates a reduced peptide bond, i.e., R₁-CH₂-NH-R₂, where R₁ and R₂ are amino acid residues or sequences. A reduced peptide bond may be introduced as a dipeptide subunit. Such a polypeptide would be resistant to protease activity, and would possess an extended half-life in vivo.

WO 03/012122

PCT/US02/23763

As discussed above, type IV collagens are synthesized and assembled as heterotrimers inside the cells, which are then secreted extracellularly where hexamer assembly, and subsequent basement membrane assembly, occurs. The polypeptides disclosed herein can work intra-cellularly to prevent heterotrimer assembly, which also necessarily inhibits hexamer assembly, and provide the desired therapeutic result. Alternatively (or additionally), the polypeptides disclosed herein can work extracellularly, to inhibit hexamer assembly, and/or to disrupt assembled hexamers, providing the desired therapeutic result.

Such polypeptides can be selected based on their utility in inhibiting generic heterotrimer assembly (ie: not α chain specific); specific heterotrimer assembly (ie: α chain specific); generic hexamer assembly (ie: not α chain specific); and/or specific hexamer assembly (ie: not α chain specific). Without knowledge of the type IV collagen $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer structure described herein, the design of inhibitors with such desired properties would not be available to those skilled in the art.

The single letter abbreviation for amino acids is used herein; "norL" refers to nor leucine.

In one embodiment, the polypeptides consist of at least 8 contiguous amino acids of general formula I:

PF(R1)(R2)CN(R3)(R4)(R5)VC(R6)(R7)A (SEQ ID NO:1)

R1 is selected from the group consisting of L, M, A, V, norL, and I;

R2 is selected from the group consisting of F and Y;

R3 is selected from the group consisting of I, V, L, norL, A, and P;

R4 is selected from the group consisting of N, G, and H;

R5 is selected from the group consisting of N, D, Q, and E;

R6 is selected from the group consisting of N, Y, and H; and

R7 is selected from the group consisting of F and Y.

This general formula I is derived from a consensus sequences of type IV collagen NC1 $\alpha 1$ - $\alpha 6$ domains at the inter-chain domain swapping region ("Inter-CDSR") that includes the $\beta 6$ - $\beta 7$ strands in the crystal structure, as further described below. This region is involved in interchain interactions within the heterotrimer, and a substantial portion of the sequence is also present at the hexamer interface, and thus is involved in hexamer assembly/stabilization. As such, peptides of general formula I

WO 03/012122

PCT/US02/23763

are useful for inhibiting appropriate interchain interactions, and thus for disrupting optimal heterotrimer and hexamer assembly.

In various further embodiments, the polypeptides consists of at least 9, 10, 11, 12, 13, or 14 amino acids of general formula I. In a preferred embodiment, the polypeptide consists of 14 amino acids of general formula I.

In a preferred embodiment, the polypeptides consist at least 8 contiguous amino acids of general formula II, with the further limitation that R2 is F; R4 is N; R5 is selected from the group consisting of N and D; R6 is N; and R7 is F. Polypeptides of this embodiment are derived from a consensus sequences of type IV collagen NC1 $\alpha 1$, $\alpha 3$, and $\alpha 5$ domains at the Inter-CDSR.

In a further preferred embodiment, the polypeptides consist at least 8 contiguous amino acids of general formula I, with the further limitation that R2 is Y; R3 is selected from the group consisting of P and I; R5 is selected from the group consisting of D, Q, and E; R6 is selected from the group consisting of Y and H; and R7 is Y. Polypeptides of this embodiment are derived from a consensus sequences of type IV collagen NC1 $\alpha 2$, $\alpha 4$, and $\alpha 6$ domains at the Inter-CDSR.

In a further preferred embodiment, the polypeptides according to formula I consist of at least 8 contiguous amino acids of a sequence selected from the group consisting of PFLFCNINNVCFNA ($\alpha 1$) (SEQ ID NO:2); PFLFCNVNDVCFNA ($\alpha 3$) (SEQ ID NO:3); PFMFCNINNVCFNA ($\alpha 5$) (SEQ ID NO:4); PFLYCNPGDVCYYA ($\alpha 2$) (SEQ ID NO:5); PFAYCNHQQVCHYA ($\alpha 4$) (SEQ ID NO:6); and PFTYCNINEVCHYA ($\alpha 6$) (SEQ ID NO:7). These sequences represent the Inter-CDSR sequences from the individual type IV collagen $\alpha 1$ - $\alpha 6$ NC1 domains. In various further embodiments, the polypeptides consist of at least 9, 10, 11, 12, 13, or 14 amino acids of one of the recited sequences. In a preferred embodiment, the polypeptide consists of 14 amino acids of one of the recited sequences.

In another embodiment, the polypeptides of the present invention consist of at least 7 contiguous amino acids of general formula II:

PF(R1)EC(R2)G(R3)(R4)GTC(R5) (SEQ ID NO:8)

R1 is selected from the group consisting of L, A, V, norL, and I;

R2 is selected from the group consisting of H, N, Q, and S;

R3 is selected from the group consisting of G, R, A, or is absent;

R4 is selected from the group consisting of R and Q; and

R5 is selected from the group consisting of N and H.

WO 03/012122

PCT/US02/23763

This general formula is derived from a consensus sequences of type IV collagen NC1 $\alpha 1$ - $\alpha 6$ domains at the intra-chain domain swapping region ("Intra-CDSR") that includes the $\beta 6'$ - $\beta 7'$ strands in the crystal structure, as further described below. This region is involved in monomer-monomer interactions within the heterotrimer, and a substantial portion of the sequence is also present at the hexamer interface, and thus is involved in hexamer assembly/stabilization. As such, peptides of this general formula are useful for inhibiting both heterotrimer and hexamer interactions of type IV collagen.

In various further embodiments, the polypeptides consists of at least 8, 9, 10, 11, 12, or 13 amino acids of general formula II. In a preferred embodiment, the polypeptide consists of 13 amino acids of general formula II.

In a preferred embodiment, the polypeptides consist at least 7 contiguous amino acids of general formula II, with the further limitation that R2 is H; R3 is R; R4 is G; and R5 is N. Polypeptides of this embodiment are derived from a consensus sequence of the intra-CDSR sequences of the type IV collagen $\alpha 1$, $\alpha 3$, and $\alpha 5$ NC1 domains.

In a further preferred embodiment, the polypeptides consist at least 7 contiguous amino acids of general formula II, with the further limitation that R2 is selected from the group consisting of N, Q, and S; R3 is selected from the group consisting of G, R, and A; R4 is selected from the group consisting of R and Q; and R5 is H. Polypeptides of this embodiment are derived from a consensus sequence of the intra-CDSR sequences of the type IV collagen $\alpha 2$, $\alpha 4$, and $\alpha 6$ NC1 domains.

In a further embodiment, the polypeptides according to general formula II consist of at least 7 contiguous amino acids of a sequence selected from the group consisting of PFIECHGRGTCN ($\alpha 1$ and $\alpha 5$) (SEQ ID NO:9); PFLECHGRGTCN ($\alpha 3$) (SEQ ID NO:10); PFIECNGGRGTCH ($\alpha 2$) (SEQ ID NO:11); PFLECQGRQGTCH ($\alpha 4$) (SEQ ID NO:12); and PFIECSGARGTCH ($\alpha 6$) (SEQ ID NO:13). These sequences represent the Intra-CDSR sequences from the individual type IV collagen $\alpha 1$ - $\alpha 6$ NC1 domains. In various further embodiments, the polypeptides of this embodiment consist of at least 8, 9, 10, 11, 12, or 13 amino acids of one of the recited sequences. In a most preferred embodiment, the polypeptides consist of 12 ($\alpha 1$, $\alpha 3$, $\alpha 5$) or 13 ($\alpha 2$, $\alpha 4$, $\alpha 6$) contiguous amino acids of any one the recited sequences.

WO 03/012122

PCT/US02/23763

In a further embodiment, the full length Intra-CDSR polypeptides (e.g.: SEQ ID NO: 9, 10, 11, 12, or 13) may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Thus, the polypeptides of the invention derived from the Intra-CDSR sequence of the $\alpha 1$ -like NC1 chains can thus be selected from the group consisting of at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 amino acids of a sequence selected from the group consisting of:

$\alpha 1$: (E)(F)(R)(S)(A) PFIECHGRGTCN(Y)(Y)(A)(N)(A) (SEQ ID NO:14),

$\alpha 3$: (E)(F)(R)(A)(S)PFIECHGRGTCN(Y)(Y)(S)(N)(S) (SEQ ID NO: 15); and

$\alpha 5$: (E)(F)(R)(S)(A)PFIECHGRGTCN(Y)(Y)(A)(N)(S) (SEQ ID NO: 16);

wherein the residues in parenthesis are the flanking sequences of the Intra-CDSR.

Alternatively, the polypeptides of the invention derived from the Intra-CDSR sequence of the $\alpha 2$ -like NC1 chains can thus be selected from the group consisting of at least 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, or 23 amino acids of a sequence selected from the group consisting of:

$\alpha 2$: (D)(F)(R)(A)(T)PFIECNGGRGTCH(Y)(Y)(A)(N)(K) (SEQ ID NO: 17);

$\alpha 4$: (D)(F)(R)(A)(A)PFLECQGRQGTCH(F)(F)(A)(N)(K) (SEQ ID NO: 18); and

$\alpha 6$: (D)(F)(R)(A)(T)PFIECSGARGTCH(Y)(F)(A)(N)(K) (SEQ ID NO: 19);

wherein the residues in parenthesis are the flanking sequences of the Intra-CDSR.

30

The Inter CDSR sequence, while widely separated in the linear sequence of a given type IV collagen NC1 domain from the Intra-CDSR sequence in the same α chain (separated by approximately 100 amino acids), is present in close spatial proximity (within approximately 2 amino acids) to the Inter-CDSR sequence in the

WO 03/012122

PCT/US02/23763

same α chain based on the derived crystal structure data. Thus, in another embodiment, the present invention provides chimeric polypeptides comprising:

- (a) one or more Inter-CDSR polypeptides of general formula I;
- (b) one or more Intra-CDSR polypeptides of general formula II; and
- 5 (c) a linker polypeptide between the Intra-CDSR polypeptide and the Inter-CDSR polypeptide consisting of between 0-20 amino acids.

In preferred embodiments, the Inter-CDSR and/or the Intra-CDSR portion of the chimeric polypeptides consists of 8, 9, 10, 11, 12, 13, or 14 amino acids of general formula I and 7, 8, 9, 10, 11, 12, 13 amino acids of general formula II, respectively.

- 10 In various other preferred embodiments, the linker polypeptide consists of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids. The optimal length of the spacer depends, at least in part, on the length of the Inter-CDSR and Intra-CDSR, as well as the position of the sequences within the full length Inter-CDSR and Intra-CDSR used to create the chimera. For example, if a full length Inter-CDSR and a full length Intra-CDSR were used, then the spacer is preferably between
- 15 0-5 amino acids in length, more preferably between 1-4 amino acids in length, and most preferably 2-3 amino acids in length. Based on the teachings herein, it will be apparent to one of skill in the art to design further such chimeric polypeptides.

- In a most preferred embodiment of these chimeric polypeptides, the Inter-CDSR polypeptide is selected from the group consisting of PFLFCNINNVCFNA (SEQ ID NO:2), PFLFCNVNDVCFNA (SEQ ID NO:3), PFMFCNINNVCFNA (SEQ ID NO:4), PFLYCNPGDVCYYA (SEQ ID NO:5), PFAYCNIHQVCHYA (SEQ ID NO:6), and PFIYCNINEVCHYA (SEQ ID NO:7); the Intra-CDSR polypeptide is selected from the group consisting of PFIECHGRGTCN (SEQ ID
- 20 NO:9), PFLECHGRGTCN (SEQ ID NO:10), PFIECNGGRGTCH (SEQ ID NO:11), PFLECQGRGTCH (SEQ ID NO:12), and PFIECSGARGTCH (SEQ ID NO:13); and the linker polypeptide is 1, 2, 3, 4, or 5 amino acids; most preferably 2 amino acids.

- In another embodiment, the polypeptides of the present invention consist of a
- 30 sequence of an amino acids of general formula III:

F(R1)T(R2) (SEQ ID NO:20)

wherein R1 is selected from the group consisting of S and T; and

R2 is selected from the group consisting of M and L.

WO 03/012122

PCT/US02/23763

This general formula III is derived from a consensus sequences of type IV collagen NC1 $\alpha 1$ - $\alpha 6$ domains at the specificity region ("SR") between the $\beta 5$ - $\beta 6$ strands in the crystal structure, as further described below. This region is involved in specific recognition between monomers, by recognizing the specificity region partner ("SRP") in the monomer with which the SR of a given α chain interacts. As such, peptides of general formula III are useful for inhibiting both heterotrimer and hexamer interactions of type IV collagen.

In a further embodiment, the SR polypeptides are selected from the group consisting of FSTM ($\alpha 1$, $\alpha 2$, $\alpha 5$, and $\alpha 6$) (SEQ ID NO:21), FTTM ($\alpha 3$) (SEQ ID NO:22) and FTSL ($\alpha 4$) (SEQ ID NO:23).

In a further embodiment, the SR polypeptides (e.g.: SEQ ID NO:21, 22, and 23) may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Thus, according to this embodiment, the polypeptides of the invention derived from the SR sequence of the NC1 α chains can be selected from the group consisting of:

$\alpha 1$ X1-FSTM-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SCLRK (SEQ ID NO: 24), and Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PFLFC (SEQ ID NO: 25) (the full sequence would thus be SCLRKFSTMPFLFC) (SEQ ID NO: 26);

$\alpha 3$: X3-FTTM-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SCLQR (SEQ ID NO: 27), and Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PFLFC (SEQ ID NO: 25) (the full sequence would thus be SCLQRFTTMPFLFC) (SEQ ID NO:28);

$\alpha 5$: X5-FSTM-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SCLRR (SEQ ID NO: 29), and Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PFMFC (SEQ ID NO: 30) (the full sequence would thus be SCLRRFSTMPFMFC) (SEQ ID NO: 31);

$\alpha 2$: X2-FSTM-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SCLAR (SEQ ID NO: 32), and Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PFLYC (SEQ ID NO: 33) (the full sequence would thus be SCLARFSTMPFLYC) (SEQ ID NO: 34);

WO 03/012122

PCT/US02/23763

$\alpha 4$: X4-FSTL-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SCLPV (SEQ ID NO: 35), and Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PFAYC (SEQ ID NO: 36) (the full sequence would thus be SCLPVFSTLPPFAYC) (SEQ ID NO: 37); and

5 $\alpha 6$: X6-FSTM-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SCLPR (SEQ ID NO: 38), and Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PFIYC (SEQ ID NO: 39) (the full sequence would thus be SCLPRFSTMPFIYC) (SEQ ID NO: 40).

10 In another embodiment, the polypeptides of the invention consist of an amino acid sequence of general formula IV:

(R1)MF(R2)K (SEQ ID NO:41)

wherein R1 is selected from the group consisting of E, R, and D; and

R2 is selected from the group consisting of K, R, and S.

15 This general formula IV is derived from a consensus sequences of type IV collagen NC1 $\alpha 1$, $\alpha 3$, and $\alpha 5$ domains at the specificity region partner ("SRP") located between the $\beta 8'$ and $\beta 9'$ strands, as discussed in more detail below. This region is involved in specific recognition between monomers, by recognizing the specificity region ("SR") in the monomer with which the SRP of a given α chain
20 interacts. As such, peptides of general formula IV are useful for inhibiting both heterotrimer and hexamer interactions of type IV collagen.

In a preferred embodiment, the SRP polypeptides according to general formula IV are selected from the group consisting of EMFKK ($\alpha 1$) (SEQ ID NO:42), RMRK ($\alpha 3$) (SEQ ID NO:43), and DMFSK ($\alpha 5$) (SEQ ID NO:44).

25 In a further preferred embodiment, the SRP polypeptides are selected from the group consisting of SFQ (SRP of $\alpha 2$) (SEQ ID NO:45); LQF (SRP of $\alpha 4$) (SEQ ID NO:46), and QQF (SRP of $\alpha 6$) (SEQ ID NO:47). These sequences represent the SRP of the type IV collagen α chain NC1 domains as indicated. This region in the $\alpha 2$ NC1 domain adopts an extended conformation and pairs with the extended
30 structure (Phe57-Thr59) in the adjacent $\alpha 1$ chain to form a short parallel β sheet, which is the only parallel β -sheet in the entire structure, as further discussed below.

In a further embodiment, the SRP polypeptides (e.g.: SEQ ID NOS:42-47) may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide

WO 03/012122

PCT/US02/23763

appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. The SRP-containing polypeptides of this embodiment of the invention can thus be selected from the group consisting of:

5 $\alpha 1$ X1-EMFKK-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TIERS (SEQ ID NO: 48), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PTPST (SEQ ID NO: 49) (the full length sequence would thus be TIERSEMFKKPTPST) (SEQ ID NO: 50);

10 $\alpha 3$: X3-RMFRK-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLNPE (SEQ ID NO: 51), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PIPST (SEQ ID NO: 52) (the full length sequence would thus be SLNPERMFRKPIPST) (SEQ ID NO: 53);

15 $\alpha 5$: X5-DMFSK-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TVDVVS (SEQ ID NO: 54), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PQSET (SEQ ID NO: 55) (the full length sequence would thus be TVDVSDMFSKPQSET) (SEQ ID NO: 56);

$\alpha 2$: X2-SFQ-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TIPEQ (SEQ ID NO: 57), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GSPSA (SEQ ID NO: 58) (the full length sequence would thus be TIPEQSFQGSPSA) (SEQ ID NO: 59);

20 $\alpha 4$: X4-LQF-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TVKAD (SEQ ID NO: 60), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SSAPA (SEQ ID NO: 61) (the full length sequence would thus be TVKADLQFSSAPA) (SEQ ID NO: 62); and

25 $\alpha 6$: X6-QQF-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TVEER (SEQ ID NO: 63), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GELPV (SEQ ID NO: 64) (the full length sequence would thus be TVEERQQFGELPV) (SEQ ID NO: 65).

In another embodiment, the polypeptides of the invention consist of an amino acid sequence of general formula V:

(R1)AH(R2)QD (SEQ ID NO: 66)

wherein R1 is selected from the group consisting of R and K; and

R2 is selected from the group consisting of G and N.

WO 03/012122

PCT/US02/23763

This general formula V is derived from a consensus sequences of type IV collagen NC1 domain β -barrel-like core at the β 4 strand, as discussed in more detail below. This region is involved in generic monomer-monomer interactions. As such, peptides of general formula V are useful for inhibiting both heterotrimer and hexamer interactions of type IV collagen.

In a preferred embodiment, the polypeptides according to general formula V are selected from the group consisting of RAHGQD (α 1, α 3, α 5) (SEQ ID NO:67) and KAHNQD (α 2, α 4, α 6) (SEQ ID NO:68).

In a further preferred embodiment, the β -barrel polypeptides according to general formula V (e.g.: SEQ ID NOS:67-68) may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. The β -barrel-containing polypeptides of this embodiment of the invention can thus be selected from the group consisting of:

α 1: X1-RAHGQD-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VQGNE (SEQ ID NO: 69), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LGTAG (SEQ ID NO: 70) (the full length sequence would thus be VQGN_{ER}RAHGQDDL_{GTA}) (SEQ ID NO: 71);

α 3: X3-RAHGQD-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VQGNQ (SEQ ID NO: 72), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LGTLG (SEQ ID NO: 73) (the full length sequence would thus be VQGN_{QRA}HGQDL_{GTLG}) (SEQ ID NO:74);

α 5: X5-RAHGQD-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VQGNK (SEQ ID NO: 75), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LGTAG (SEQ ID NO: 70) (the full length sequence would thus be VQGN_{KRA}HGQDL_{GTAG}) (SEQ ID NO: 76);

α 2: X2-KAHNQD-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FEGQE (SEQ ID NO: 77), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LGLAG (SEQ ID NO: 78) (the full length sequence would thus be FEGQE_{KAH}NQDL_{GLAG}) (SEQ ID NO: 79);

α 4: X4-KAHNQD-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LEGQE (SEQ ID NO: 80), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids

WO 03/012122

PCT/US02/23763

of the sequence LGLAG (SEQ ID NO: 78) (the full length sequence would thus be LEGQEKAHNQDLGLAG) (SEQ ID NO: 81); and

$\alpha 6$: X6-KAHNQD-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VEGQE (SEQ ID NO: 82), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LGFAG (SEQ ID NO: 83) (the full length sequence would thus be VEGQEKAHNQDLGFAG) (SEQ ID NO: 84).

In another embodiment, the polypeptides of the invention consist of an amino acid sequence of general formula VI:

10 (R1)G(R2)GQ (SEQ ID NO:85)

wherein R1 is selected from the group consisting of E and Q; and

R2 is selected from the group consisting of S, T, and G.

This general formula VI is derived from a consensus sequences of type IV collagen NC1 domain β -barrel-like core at the $\beta 4'$ strand, as discussed in more detail below. This region is involved in generic monomer-monomer interactions. As such, peptides of general formula VI are useful for inhibiting both heterotrimer and hexamer interactions of type IV collagen.

In a preferred embodiment, the polypeptides according to general formula VI are selected from the group consisting of EGSGQ ($\alpha 1$, $\alpha 5$) (SEQ ID NO:86), EGTGQ ($\alpha 3$) (SEQ ID NO:87), EGGGQ ($\alpha 2$, $\alpha 6$) (SEQ ID NO:88) and QGGGQ ($\alpha 4$) (SEQ ID NO:89).

In a further embodiment, the β -barrel polypeptides according to general formula VI (e.g.: SEQ ID NOS:86-89) may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. The β -barrel-containing polypeptides of this embodiment of the invention can thus be selected from the group consisting of:

$\alpha 1$ and $\alpha 5$ X1-EGSGQ-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TSAGA (SEQ ID NO: 90), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ALASP (SEQ ID NO: 91) (the full length sequence would thus be TSAGAEKSGQALASP) (SEQ ID NO: 92);

$\alpha 3$: X3-EGTGQ-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TSAGS (SEQ ID NO: 93), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids

WO 03/012122

PCT/US02/23763

of the sequence ALASP (SEQ ID NO: 91) (the full length sequence would thus be TSAGSEGTGQALASP) (SEQ ID NO:94);

α2: X2-EGGGQ-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAAGD (SEQ ID NO: 95), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLVSP (SEQ ID NO: 96) (the full length sequence would thus be TAAGDEGGGQSLVSP) (SEQ ID NO: 97);

α4: X4-QGGGQ-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TGAGD (SEQ ID NO: 98), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ALMSP (SEQ ID NO: 99) (the full length sequence would thus be TGAGDQGGGQALMSP) (SEQ ID NO: 100); and

α6: X6-EGGGQ-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAAGA (SEQ ID NO: 101), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLVSP (SEQ ID NO: 96) (the full length sequence would thus be TAAGAEGGGQSLVSP) (SEQ ID NO: 102).

15

In another embodiment, the polypeptides comprise sequences present at the hexamer interface, as determined from the deduced crystal structure. Type IV collagens are synthesized and assembled as trimers inside the cells, which are then secreted extracellularly where hexamer assembly, and subsequent basement membrane assembly, occurs. Therapeutics, such as those disclosed herein, can work intra-cellularly to prevent trimer assembly, thus inhibiting hexamer assembly, thus providing the desired therapeutic result. Alternatively (or additionally), therapeutics can work extracellularly, which leaves trimer assembly uninhibited, but targets hexamer assembly.

As such, polypeptides from regions at the hexamer interface can be used to inhibit hexamer formation or disrupt hexamer formation. In this embodiment, the polypeptides of the invention consist of an amino acid sequence of general formula VII:

(R1)G(R2)(R3) (SEQ ID NO:103)

wherein R1 is selected from the group consisting of Q and E;

R2 is selected from the group consisting of N and Q; and

R3 is selected from the group consisting of E, Q, and K.

This general formula VII is derived from a consensus sequences of type IV collagen NC1 α1-α6 domains at the hexamer interface at the end of the β3 strand up to

WO 03/012122

PCT/US02/23763

the beginning of the $\beta 4$ strand, as discussed in more detail below. This region is present at the hexamer interface, and is involved in hexamer assembly and stabilization. As such, peptides of general formula VII are useful for inhibiting hexamer interactions of type IV collagen.

5 In a preferred embodiment, the polypeptides consist of general formula VII, with the further limitation that R1 is Q and R2 is N. In this embodiment, the formula is a consensus of the sequences present in the $\alpha 1/\alpha 3/\alpha 5$ NC1 domains for general formula VII. In a further preferred embodiment, the polypeptides according to general formula VII are selected from the group consisting of QGNE ($\alpha 1$) (**SEQ ID NO:104**), QGNQ ($\alpha 3$) (**SEQ ID NO:105**), and QGNK ($\alpha 5$) (**SEQ ID NO:106**)

10 In a further preferred embodiment, the polypeptides according to general formula VII consist of EGQE (**SEQ ID NO:107**), which is the sequence of the sequences present in the $\alpha 2/\alpha 4/\alpha 6$ NC1 domains in general formula VII.

15 In a further embodiment, the hexamer polypeptides selected from the group consisting of **SEQ ID NOS:104-107** may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Such polypeptides can thus be selected from the group consisting of:

20 $\alpha 1$: X1-QGNE-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYV (**SEQ ID NO: 108**), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RAHGQ (**SEQ ID NO: 109**) (the full length sequence would thus be SLLYVQGNRAHGQ) (**SEQ ID NO: 110**);

25 $\alpha 3$: X3-QGNQ-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SFLFV (**SEQ ID NO: 111**), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RAHGQ (**SEQ ID NO: 109**) (the full length sequence would thus be SFLFVQGNQRAHGQ) (**SEQ ID NO:112**);

30 $\alpha 5$: X5-QGNK-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYV (**SEQ ID NO:108**), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RAHGQ (**SEQ ID NO: 109**) (the full length sequence would thus be SLLYVQGNKRAHGQ) (**SEQ ID NO: 113**);

$\alpha 2$: X2-EGQE-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYF (**SEQ ID NO:114**), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids

WO 03/012122

PCT/US02/23763

of the sequence KAHNQ (SEQ ID NO:115) (the full length sequence would thus be SLLYFEGQEKAHNQ) (SEQ ID NO: 116);

15 $\alpha 4$: X4-EGQE-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYL (SEQ ID NO:117), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KAHNQ (SEQ ID NO:115) (the full length sequence would thus be SLLYLEGQEKAHNQ) (SEQ ID NO: 118); and

20 $\alpha 6$: X6-EGQE-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLFV (SEQ ID NO:119), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KAHNQ (SEQ ID NO:115) (the full length sequence would thus be SLLFVEGQEKAHNQ) (SEQ ID NO: 120).

An especially preferred embodiment of these hexamer interface polypeptides according to general formula VII consists of 1 additional amino acid at both the amino and carboxy terminus of the $\alpha 1$ - $\alpha 6$ hexamer peptides, as follows:

15 $\alpha 1$ VQGNER (SEQ ID NO: 121)
 $\alpha 3$: VQGNQR (SEQ ID NO: 122)
 $\alpha 5$: VQGNKR (SEQ ID NO: 123)
 $\alpha 2$: FEGQEK (SEQ ID NO: 124)
 $\alpha 4$: LEGQEK (SEQ ID NO: 125)
 20 $\alpha 6$: VEGQEK (SEQ ID NO: 126)

In a further embodiment wherein the polypeptides comprise sequences present at the hexamer interface, as determined from the deduced crystal structure, the polypeptides of the invention consist of an amino acid sequence of general formula 25 VIII:

M(R1)M(R2)P (SEQ ID NO:127)

wherein R1 is selected from the group consisting of S, N, or is absent; and R2 is selected from the group consisting of A, Q, or is absent.

This general formula VIII is derived from a consensus sequences of type IV 30 collagen NC1 $\alpha 1$ - $\alpha 6$ domains at the hexamer interface between the $\beta 8$ and $\beta 9$ strands, as discussed in more detail below. This region is present at the hexamer interface, and is involved in hexamer assembly and stabilization. As such, peptides of general formula VIII are useful for inhibiting hexamer interactions of type IV collagen.

WO 03/012122

PCT/US02/23763

In a preferred embodiment, the polypeptides of general formula VIII are selected from the group consisting of MSMAP ($\alpha 1$) (SEQ ID NO:128), MNMAP ($\alpha 3$) (SEQ ID NO:129), MSMQP ($\alpha 5$) (SEQ ID NO:130), and MMP ($\alpha 2$, $\alpha 4$, and $\alpha 6$) (SEQ ID NO: 131).

5 In a further preferred embodiment, the hexamer polypeptides selected from the group consisting of SEQ ID NOS:128-131 may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Such polypeptides can thus be selected
10 from the group consisting of:

$\alpha 1$: X1-MSMAP-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PEPMP (SEQ ID NO: 132), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ITGEN (SEQ ID NO: 133) (the full length sequence would thus be PEPMPMSMAPITGEN) (SEQ ID NO: 134);

15 $\alpha 3$: X3-MNMAP-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PALMP (SEQ ID NO: 135), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ITGRA (SEQ ID NO: 136) (the full length sequence would thus be PALMPMNMAPITGRA) (SEQ ID NO:137);

20 $\alpha 5$: X5-MSMQP-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PEPMP (SEQ ID NO:132), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LKGQS (SEQ ID NO: 138) (the full length sequence would thus be PEPMPMSMQPLKGQS) (SEQ ID NO: 139);

$\alpha 2$: X2-MMP-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAPLP (SEQ ID NO:140), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids
25 of the sequence VAEDE (SEQ ID NO:141) (the full length sequence would thus be TAPLPMMPVAEDE) (SEQ ID NO: 142);

$\alpha 4$: X4-MMP-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AAPLP (SEQ ID NO:143), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LSEEA (SEQ ID NO:144) (the full length sequence would thus be
30 AAPLPMPLSEEA) (SEQ ID NO: 145); and

$\alpha 6$: X6-MMP-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAPIP (SEQ ID NO:146), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VSQTQ (SEQ ID NO:147) (the full length sequence would thus be TAPIPMMPVSQTQ) (SEQ ID NO: 148).

WO 03/012122

PCT/US02/23763

An especially preferred embodiment of these hexamer interface peptides according to general formula VIII consists of 3 additional amino acids at both the amino and carboxy terminus of the $\alpha 1$ - $\alpha 6$ hexamer peptides, as follows:

- 5 $\alpha 1$ PMPMSMAPITG (SEQ ID NO: 149);
 $\alpha 3$: LMPMNMAPITG (SEQ ID NO:150);
 $\alpha 5$: PMPMSMQPLKG (SEQ ID NO: 151);
 $\alpha 2$: PLPMMPVAE (SEQ ID NO: 152);
 $\alpha 4$: PLPMMPLSE (SEQ ID NO: 153); and
10 $\alpha 6$: PIPMMPVSQ (SEQ ID NO: 154).

In a further embodiment wherein the polypeptides comprise sequences present at the hexamer interface, as determined from the deduced crystal structure, the polypeptides of the invention consist of an amino acid sequence of general formula
15 IX:

AG(R1)(R2) (SEQ ID NO:155)

wherein R1 is selected from the group consisting of A, S and D; and

R2 is selected from the group consisting of E and Q.

This general formula IX is derived from a consensus sequences of type IV
20 collagen NC1 $\alpha 1$ - $\alpha 6$ domains between the $\beta 3'$ and $\beta 4'$ strands, as discussed in more detail below. This region is present at the hexamer interface, and is involved in hexamer assembly and stabilization. As such, peptides of general formula IX are useful for inhibiting hexamer interactions of type IV collagen.

In a preferred embodiment, the polypeptides of general formula IX are
25 selected from the group consisting of AGAE ($\alpha 1$, $\alpha 5$, and $\alpha 6$) (SEQ ID NO:156), AGSE ($\alpha 3$) (SEQ ID NO:157), AGDE ($\alpha 2$) (SEQ ID NO:158), and AGDQ ($\alpha 4$) (SEQ ID NO:159).

In a further embodiment, the hexamer polypeptides selected from the group consisting of SEQ ID NOS:156-159 may optionally further include 0-5 amino acids
30 at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Such polypeptides can thus be selected from the group consisting of:

WO 03/012122

PCT/US02/23763

- $\alpha 1$: X1-AGAE-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VMHTS (SEQ ID NO: 160), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GSGQA (SEQ ID NO: 161) (the full length sequence would thus be VMHTSAGAEGSGQA) (SEQ ID NO: 162);
- 5 $\alpha 3$: X3-AGSE-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence IMFTS (SEQ ID NO: 163), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GTGQA (SEQ ID NO: 164) (the full length sequence would thus be IMFTSAGSEGTGQA) (SEQ ID NO: 165);
- $\alpha 5$: X5-AGAE-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence MMHTS (SEQ ID NO: 166), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GSGQA (SEQ ID NO: 161) (the full length sequence would thus be MMHTSAGAEGSGQA) (SEQ ID NO: 167);
- 10 $\alpha 2$: X2-AGDE-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LMHTA (SEQ ID NO: 168), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GGGQS (SEQ ID NO: 169) (the full length sequence would thus be LMHTAAGDEGGQS) (SEQ ID NO: 170);
- $\alpha 4$: X4-AGDQ-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LMHTG (SEQ ID NO: 171), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GGGQA (SEQ ID NO: 172) (the full length sequence would thus be LMHTGAGDQGGQA) (SEQ ID NO: 173); and
- 20 $\alpha 6$: X6-AGAE-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LMHTA (SEQ ID NO: 168), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GGGQS (SEQ ID NO: 169) (the full length sequence would thus be LMHTAAGAEGGGQS) (SEQ ID NO: 174).

25

In a further embodiment wherein the polypeptides comprise sequences present at the hexamer interface, as determined from the deduced crystal structure, the polypeptides of the invention consist of at least 5 amino acids of the sequence of general formula X:

- 30 EC(R1)G(R2)(R3)GTC(R4)(R5)(R6) (SEQ ID NO: 175)
 wherein R1 is selected from the group consisting of H, N, Q, and S;
 R2 is selected from the group consisting of G, R, A, or is absent;
 R3 is selected from the group consisting of R and Q
 R4 is selected from the group consisting of N and H;

WO 03/012122

PCT/US02/23763

R5 is selected from the group consisting of F and Y; and

R6 is selected from the group consisting of F and Y.

In various preferred embodiments, the polypeptide consists of at least 6, 7, 8, 9, 10, 11, or 12 amino acids of general formula X. In a preferred embodiment, the polypeptide consists of 12 amino acids of general formula X. This general formula X extensively overlaps with the Intra-CDSR, discussed above, and is present within the $\beta 6'$ - $\beta 7'$ strands, as discussed in more detail below. This region is present at the hexamer interface, and is involved in hexamer assembly and stabilization. As such, peptides of general formula X are useful for inhibiting hexamer interactions of type IV collagen.

In a further embodiment, the polypeptides are as described above for general formula X, with the exception that R2 is selected from the group consisting of G, R, A; and R4 is H. Polypeptides of this embodiment are derived from the consensus sequence of the $\alpha 2/4/6$ of general formula X.

In a further preferred embodiment, the polypeptides of general formula X are selected from the group consisting of ECHGRGTCNYY ($\alpha 1/3/5$) (SEQ ID NO:176), ECNGGRGTCHYY ($\alpha 2$) (SEQ ID NO:177), ECQGRQGTCHFF ($\alpha 4$) (SEQ ID NO:178), and ECSGARGTCHYP ($\alpha 6$) (SEQ ID NO:179).

In a further preferred embodiment, the polypeptides of the invention consist of an amino acid sequence of general formula XI:

(R1)(R2)T(R3)K (SEQ ID NO:180)

wherein R1 is selected from the group consisting of P, S, and A;

R2 is selected from the group consisting of S, E, and D; and

R3 is selected from the group consisting of L and V.

This general formula XI is present overlapping with the $\beta 9'$ strand in the crystal structure, as discussed in more detail below. This region is present at the hexamer interface, and is involved in hexamer assembly and stabilization. As such, peptides of general formula XI are useful for inhibiting hexamer interactions of type IV collagen.

In a preferred embodiment of general formula XI, R3 is L (as in $\alpha 2/4/6/1/5$).

In a further preferred embodiment of general formula XI, R2 is selected from D and E ($\alpha 2/4/5/6$). In further preferred embodiments, the polypeptide according to general formula XI is selected from the group consisting of PSTLK ($\alpha 1$) (SEQ ID NO:181),

WO 03/012122

PCT/US02/23763

PSTVK ($\alpha 3$) (SEQ ID NO:182), SETLK ($\alpha 5$ and $\alpha 6$) (SEQ ID NO:183), ADTLK ($\alpha 2$) (SEQ ID NO:184), and PDTLK ($\alpha 4$) (SEQ ID NO:185).

In a further embodiment, the hexamer polypeptides selected from the group consisting of SEQ ID NOS:181-185 may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Such polypeptides can thus be selected from the group consisting of:

10 $\alpha 1$: X1-PSTLK-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FKKPT (SEQ ID NO: 186), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AGELR (SEQ ID NO: 187) (the full length sequence would thus be FKKPTPSTLKAGELR) (SEQ ID NO: 188);

15 $\alpha 3$: X3-PSTVK-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FRKPI (SEQ ID NO: 189), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AGELE (SEQ ID NO: 190) (the full length sequence would thus be FRKPIPSTVKAGELE) (SEQ ID NO:191);

20 $\alpha 5$: X5-SETLK-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FSKPQ (SEQ ID NO:192), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AGDLR (SEQ ID NO: 193) (the full length sequence would thus be FSKPQSETLKAGDLR) (SEQ ID NO: 194);

25 $\alpha 2$: X2-ADTLK-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence QGSPS (SEQ ID NO:195), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AGLIR (SEQ ID NO:196) (the full length sequence would thus be QGSPSADTLKAGLIR) (SEQ ID NO: 197);

$\alpha 4$: X4-PDTLK-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SSAPA (SEQ ID NO:198), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ESQAQ (SEQ ID NO:199) (the full length sequence would thus be SSAPAPDTLKESQAQ) (SEQ ID NO: 200); and

30 $\alpha 6$: X6-SETLK-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GELPV (SEQ ID NO:201), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AGQLH (SEQ ID NO:202) (the full length sequence would thus be GELPVSETLKAGQLH) (SEQ ID NO: 203).

WO 03/012122

PCT/US02/23763

In a further preferred embodiment, the polypeptides of the invention consist of an amino acid sequence of general formula XII:

A(R1)RND (SEQ ID NO:204)

wherein R1 is selected from the group consisting of S, Q, and R.

5 This general formula XII is present in the highly conserved loop connecting the $\beta 7$ and $\beta 8$ strands in the crystal structure. This region is present at the hexamer interface, and is involved in hexamer assembly and stabilization. As such, peptides of general formula XII are useful for inhibiting hexamer interactions of type IV collagen.

10 In further preferred embodiments, the polypeptide according to general formula XII is selected from the group consisting of ASRND ($\alpha 1$, $\alpha 3$, $\alpha 5$, $\alpha 2$) (SEQ ID NO:205), AQRND ($\alpha 4$) (SEQ ID NO:206), and ARRND ($\alpha 6$) (SEQ ID NO:207).

15 In a further embodiment, the hexamer polypeptides selected from the group consisting of SEQ ID NOS:205, 206, and 207 may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Such polypeptides can thus be selected from the group consisting of:

20 $\alpha 1$ and $\alpha 5$: X1-ASRND-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence NVCNF (SEQ ID NO: 208), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSYWL (SEQ ID NO: 209) (the full length sequence would thus be NVCNFASRNDYSYWL) (SEQ ID NO: 210);

25 $\alpha 3$: X3-ASRND-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence DVCNF (SEQ ID NO: 211), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSYWL (SEQ ID NO: 209) (the full length sequence would thus be DVCNFASRNDYSYWL) (SEQ ID NO:212);

30 $\alpha 2$: X2-ASRND-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence DVCYY (SEQ ID NO:213), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KSYWL (SEQ ID NO:214) (the full length sequence would thus be DVCYYASRNDKSYWL) (SEQ ID NO: 215);

$\alpha 4$: X4-AQRND-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence QVCHY (SEQ ID NO:216), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino

WO 03/012122

PCT/US02/23763

acids of the sequence RSYWL (SEQ ID NO:217) (the full length sequence would thus be QVCHYAQRNDRSYWL) (SEQ ID NO: 218); and

α6: X6-ARRND-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence EVCHY (SEQ ID NO:219), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KSYWL (SEQ ID NO:214) (the full length sequence would thus be EVCHYARRNDKSYWL) (SEQ ID NO: 220).

In a further preferred embodiment, the polypeptides of the invention consist of an amino acid sequence of general formula XIII:

(R1)(R2)(R3)N(R4) (SEQ ID NO:221)

10 wherein R1 is selected from the group consisting of Y and F;
R2 is selected from the group consisting of Y and F;
R3 is selected from the group consisting of A and S; and
R4 is selected from the group consisting of A, S, and K.

15 This general formula XIII is present in the highly conserved loop connecting the β7' and β8' strands in the crystal structure. This region is present at the hexamer interface, and is involved in hexamer assembly and stabilization. As such, peptides of general formula XIII are useful for inhibiting hexamer interactions of type IV collagen.

20 In further preferred embodiments, the polypeptide according to general formula XIII is selected from the group consisting of YYANA (α1) (SEQ ID NO:222) YYSNS (α3) (SEQ ID NO:223) YYANS (α5) (SEQ ID NO:224) YYANK (α2) (SEQ ID NO:225) FFANK (α4) (SEQ ID NO:226) and YFANK(α6) (SEQ ID NO:227).

25 In a further embodiment, the hexamer polypeptides selected from the group consisting of SEQ ID NOS:222-227 may optionally further include 0-5 amino acids at either or both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Such polypeptides can thus be selected from the group consisting of:

30 α1: X1-YYANA-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO: 228), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO: 229) (the full length sequence would thus be RGTCNYYANAYSFWL) (SEQ ID NO: 230);

WO 03/012122

PCT/US02/23763

- $\alpha 3$: X3-YYSNS-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO: 228), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO: 229) (the full length sequence would thus be RGTCNYYSNSYSFWL) (SEQ ID NO:231);
- 5 $\alpha 5$: X1-YYANS-Z2, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO: 228), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO: 229) (the full length sequence would thus be RGTCNYYANSYSFWL) (SEQ ID NO: 232);
- 10 $\alpha 2$: X2-YYANK-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO:233), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO:229) (the full length sequence would thus be RGTCNYYANKYSFWL) (SEQ ID NO: 234);
- 15 $\alpha 4$: X4-FFANK-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence QGTCH (SEQ ID NO:235), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO:229) (the full length sequence would thus be QGTCHFFANKYSFWL) (SEQ ID NO: 236); and
- 20 $\alpha 6$: X6-YFANK-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO:233), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO:229) (the full length sequence would thus be RGTCNYFANKYSFWL) (SEQ ID NO: 237).

In a further embodiment, the present invention provides novel polypeptides derived from the hypervariable region of the type IV collagen α chain NC1 domain sequences located between the $\beta 8'$ and the $\beta 9'$ strands, which are identified from the crystal structure as being present at the monomer-monomer interface, and which include the SRP and are involved in providing appropriate secondary structure for optimal interactions between the SR and the SRP. In this embodiment, the polypeptides consist of at least 7 amino acids of a sequence selected from the group consisting of IERSEMFKKPT ($\alpha 1$) (SEQ ID NO:238), LNPERMFRKPI ($\alpha 3$) (SEQ ID NO:239), VDVSDMFSPQ ($\alpha 5$) (SEQ ID NO:240), IPEQSFQGSFS ($\alpha 2$) (SEQ ID NO:241), VKADLQFSSAPA ($\alpha 4$) (SEQ ID NO:242), and VEERQQFGELPV ($\alpha 6$) (SEQ ID NO:243). In various embodiments, the polypeptides consist of at least

WO 03/012122

PCT/US02/23763

8, 9, 10, 11, or 12 amino acids of a sequence selected from the group consisting of SEQ ID NO:235-240.

In a further embodiment, the polypeptides selected from the group consisting of **SEQ ID NOS:238-243** may optionally further include 0-5 amino acids at either or
 5 both the amino and carboxyl terminus that are derived from the same α chain, in order to provide appropriate secondary structural characteristics to the polypeptide for optimal inhibitory activity. Such polypeptides can thus be selected from the group consisting of:

$\alpha 1$: X1-IERSEMFKKPT-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids
 10 of the sequence FWLAT (**SEQ ID NO: 244**), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PSTLK (**SEQ ID NO: 181**) (the full length sequence would thus be FWLATIERSEMFKKPTSTLK) (**SEQ ID NO: 245**);

$\alpha 2$: X3-LNPERMFRKPI-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids
 15 of the sequence FWLAS (**SEQ ID NO: 246**), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PSTVK (**SEQ ID NO: 182**) (the full length sequence would thus be FWLASLNPERMFRKPIPSTVK) (**SEQ ID NO:247**);

$\alpha 5$: X1-VDVSDMFSKPQ-Z2, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids
 20 of the sequence FWLAT (**SEQ ID NO: 244**), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SETLK (**SEQ ID NO: 183**) (the full length sequence would thus be FWLATVDVSDMFSKPQSETLK) (**SEQ ID NO: 248**);

$\alpha 2$: X2-IPEQSFQGSFS-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of
 the sequence FWLTT (**SEQ ID NO:249**), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ADTLK (**SEQ ID NO:184**) (the full length sequence would thus be FWLTTIPEQSFQGSFSADTLK) (**SEQ ID NO: 250**);

$\alpha 4$: X4-VKADLQFSSAPA-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino
 25 acids of the sequence FWLTT (**SEQ ID NO:249**), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PDTLK (**SEQ ID NO:185**) (the full length sequence would thus be FWLTTVKADLQFSSAPADTLK) (**SEQ ID NO: 251**); and

$\alpha 6$: X6-VEERQQFGELPV-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino
 30 acids of the sequence FWLTT (**SEQ ID NO:249**), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SETLK (**SEQ ID NO:183**) (the full length sequence would thus be FWLTTVEERQQFGELPVSETLK) (**SEQ ID NO: 252**).

WO 03/012122

PCT/US02/23763

In further embodiments, the present invention provides other polypeptides that include multiple regions identified as being important for inhibiting monomer-monomer interactions (and thus heterotrimer assembly), and/or trimer-trimer interactions (and thus hexamer assembly). Polypeptides according to this aspect of the invention include the following:

SR plus the Inter-CDSR:

$\alpha 1$: FSTMPFLFCNINNVCFNA (SEQ ID NO: 253)
 $\alpha 3$: FTTMPFLFCNVNDVCNFA (SEQ ID NO: 254)
 10 $\alpha 5$: FSTMPFMFCNINNVCFNA (SEQ ID NO: 255)
 $\alpha 2$: FSTMPFLYCNPGDVCYYA (SEQ ID NO: 256)
 $\alpha 4$: FSTLPAAYCNIHQVCHYA (SEQ ID NO: 257)
 $\alpha 6$: FSTMPFTYCNINEVCHYA (SEQ ID NO: 258)

15 Inter-CDSR plus contiguous hexamer interface region:

$\alpha 1$: PFLFCNINNVCFNFA (SEQ ID NO: 259)
 $\alpha 3$: PFLFCNVNDVCNFA (SEQ ID NO: 260)
 $\alpha 5$: PFMFCNINNVCFNFA (SEQ ID NO: 261)
 $\alpha 2$: PFLYCNPGDVCYYA (SEQ ID NO: 262)
 20 $\alpha 4$: PFAAYCNIHQVCHYA (SEQ ID NO: 263)
 $\alpha 6$: PFIYCNINEVCHYARRND (SEQ ID NO: 264)

SR plus the Inter-CDSR plus contiguous hexamer interface region:

$\alpha 1$: FSTMPFLFCNINNVCFNFA (SEQ ID NO: 265)
 25 $\alpha 3$: FTTMPFLFCNVNDVCNFA (SEQ ID NO: 266)
 $\alpha 5$: FSTMPFMFCNINNVCFNFA (SEQ ID NO: 267)
 $\alpha 2$: FSTMPFLYCNPGDVCYYA (SEQ ID NO: 268)
 $\alpha 4$: FSTLPAAYCNIHQVCHYA (SEQ ID NO: 269)
 $\alpha 6$: FSTMPFIYCNINEVCHYARRND (SEQ ID NO: 270)

30

Intra-CDSR plus contiguous hexamer interface region:

$\alpha 1$ and $\alpha 5$: PFIECHGRGTCNYY (SEQ ID NO: 271)
 $\alpha 3$: PFLECHGRGTCNYY (SEQ ID NO: 272)
 $\alpha 2$: PFIECNGGRGTCHYY (SEQ ID NO: 273)

WO 03/012122

PCT/US02/23763

$\alpha 4$: PFLECCGRQGTCHFF (SEQ ID NO: 274)
 $\alpha 6$: PFIECSGARGTCHYF (SEQ ID NO: 275)

SRP/variable region plus contiguous hexamer interface:

5 $\alpha 1$: IERSEMFKKPTSTLKAG (SEQ ID NO: 276)
 $\alpha 3$: LNPERMFRKPIPSTVKAG (SEQ ID NO: 277)
 $\alpha 5$: VDVSDMFSKPQSETLKAG (SEQ ID NO: 278)
 $\alpha 2$: IPEQSFQGSPSADTLKAG (SEQ ID NO: 279)
 $\alpha 4$: VKADLQFSSAPADTLKES (SEQ ID NO: 280)
10 $\alpha 6$: VEERQQFGELPVSETLKAG (SEQ ID NO: 281)

Specific monomer-monomer inhibitor plus SR:

$\alpha 1$: GSCLRFSTM (SEQ ID NO: 282)
 $\alpha 3$: GSCLQRFTTM (SEQ ID NO: 283)
15 $\alpha 5$: GSCLRRFSTM (SEQ ID NO: 284)
 $\alpha 2$: GSCLARFSTM (SEQ ID NO: 285)
 $\alpha 4$: GSCLPVFSTL (SEQ ID NO: 286)
 $\alpha 6$: GSCLPRFSTM (SEQ ID NO: 287)

20 Monomer-monomer inhibitor plus SR plus Inter-CDSR plus hexamer interface

$\alpha 1$: LRKFSTMPFLFCNINNVCF (SEQ ID NO: 288)
 $\alpha 3$: LQRFTTMPFLFCNVNDVCF (SEQ ID NO: 289)
 $\alpha 5$: LRRFSTMPFMFCNINNVCF (SEQ ID NO: 290)
 $\alpha 2$: LARFSTMPFLYCNPGDVCF (SEQ ID NO: 291)
25 $\alpha 4$: LPVFSTLFPAYCNHQVCF (SEQ ID NO: 292)
 $\alpha 6$: LPRFSTMPFIYCNINEVCF (SEQ ID NO: 293)

In another aspect, the present invention provides methods for inhibiting angiogenesis, angiogenesis-mediated disorders, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and basal lamina assembly, comprising administering to a subject in need thereof an amount effective to inhibit angiogenesis, angiogenesis-mediated disorders, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and basal lamina assembly of one or more polypeptides

WO 03/012122

PCT/US02/23763

of the invention, antibodies against such polypeptides, or pharmaceutical compositions thereof.

“Angiogenesis-mediated disorders” refers to diseases and conditions with accompanying undesired angiogenesis, including but not limited to solid and blood-borne tumors, diabetic retinopathy, rheumatoid arthritis, retinal neovascularization, choroidal neovascularization, macular degeneration, corneal neovascularization, retinopathy of prematurity, corneal graft rejection, neovascular glaucoma, retrolental fibroplasia, epidemic keratoconjunctivitis, Vitamin A deficiency, contact lens overwear, atopic keratitis, superior limbic keratitis, pterygium keratitis sicca, soggrens, acne rosacea, phlyctenulosis, syphilis, Mycobacteria infections, lipid degeneration, chemical burns, bacterial ulcers, fungal ulcers, Herpes simplex infections, Herpes zoster infections, protozoan infections, Kaposi's sarcoma, Mooren ulcer, Terrieu's marginal degeneration, marginal keratolysis, trauma, systemic lupus, polyarteritis, Wegeners sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy, sickle cell anemia, sarcoid, pseudoxanthoma elasticum, Pagets disease, vein occlusion, artery occlusion, carotid obstructive disease, chronic uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets disease, myopia, optic pits, Stargarts disease, pars planitis, chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser complications, abnormal proliferation of fibrovascular tissue, hemangiomas, Osler-Weber-Rendu, acquired immune deficiency syndrome, ocular neovascular disease, osteoarthritis, chronic inflammation, Crohn's disease, ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid. (See U.S. Patent No. 5,712,291)

The polypeptides, or antibodies against such polypeptides, may be subjected to conventional pharmaceutical operations such as sterilization and/or may contain conventional adjuvants, such as preservatives, stabilizers, wetting agents, emulsifiers, buffers etc.

For administration, the polypeptides, or antibodies against such polypeptides, are ordinarily combined with one or more adjuvants appropriate for the indicated route of administration. The polypeptides, or antibodies against such polypeptides, may be admixed with lactose, sucrose, starch powder, cellulose esters of alkanolic acids, stearic acid, talc, magnesium stearate, magnesium oxide, sodium and calcium salts of phosphoric and sulphuric acids, acacia, gelatin, sodium alginate, polyvinylpyrrolidone, and/or polyvinyl alcohol, and tableted or encapsulated for conventional administration. Alternatively, the polypeptides, or antibodies against

WO 03/012122

PCT/US02/23763

such polypeptides of this invention may be dissolved in saline, water, polyethylene glycol, propylene glycol, carboxymethyl cellulose colloidal solutions, ethanol, corn oil, peanut oil, cottonseed oil, sesame oil, tragacanth gum, and/or various buffers. Other adjuvants and modes of administration are well known in the pharmaceutical art. The carrier or diluent may include time delay material, such as glyceryl monostearate or glyceryl distearate alone or with a wax, or other materials well known in the art.

In practicing this aspect of the invention, the amount or dosage range of the polypeptides, antibodies against such polypeptides, or pharmaceutical compositions employed is one that effectively inhibits angiogenesis, angiogenesis-mediated disorders, tumor growth, tumor metastasis, and/or endothelial cell-extracellular matrix interactions. An inhibiting amount of the polypeptides that can be employed ranges generally between about 0.01 $\mu\text{g}/\text{kg}$ body weight and about 10 mg/kg body weight, preferably ranging between about 0.05 $\mu\text{g}/\text{kg}$ and about 5 mg/kg body weight.

The polypeptides, antibodies against such polypeptides, or pharmaceutical compositions thereof may be administered by any suitable route, including orally, parentally, by inhalation spray, rectally, or topically in dosage unit formulations containing conventional pharmaceutically acceptable carriers, adjuvants, and vehicles. The term parenteral as used herein includes, subcutaneous, intravenous, intra-arterial, intramuscular, intrasternal, intratendinous, intraspinal, intracranial, intrathoracic, infusion techniques or intraperitoneally. In preferred embodiments, the polypeptides are administered intravenously or subcutaneously.

The polypeptides, antibodies against such polypeptides, or pharmaceutical compositions thereof may be made up in a solid form (including granules, powders or suppositories) or in a liquid form (*e.g.*, solutions, suspensions, or emulsions). The polypeptides and antibodies against such polypeptides of the invention may be applied in a variety of solutions. Suitable solutions for use in accordance with the invention are sterile, dissolve sufficient amounts of the polypeptides, and are not harmful for the proposed application.

In a preferred embodiment, one or more of the disclosed polypeptides, antibodies against such polypeptides, or pharmaceutical compositions thereof, are used so as to target more than one region of type IV collagen for inhibition of assembly. For example, peptides that target different hexamer regions can be used in combination to increase their inhibitory effect. Alternatively, or additionally,

WO 03/012122

PCT/US02/23763

combining a peptide targeting monomer-monomer interactions with a peptide that targets hexamer assembly can provide an additive inhibitory effect. Other combinations are well within the knowledge of one of skill in the art, based on the teachings herein.

5

EXAMPLES

Protein Purification and Crystallization. The $[(\alpha 1)_2(\alpha 2)]_2$ NC1 hexamer was isolated from bovine eye lenses purchased from Pel-Freeze Biologicals (Rogers, AR) (37). Briefly, LBM was prepared by sonication of the lenses in the presence of 1 M NaCl and protease inhibitors (38). To cleave the NC1 domain from the full-length type IV collagen, the LBM preparation was digested with bacterial collagenase at 37° C. The NC1 hexamer was purified by using DE-52 and S-300 column chromatography.

Initial crystallization screening with commercial sparse matrix kits (Hampton Research, Laguna Niguel, CA) was carried out using concentrated protein (10 mg/ml) and hanging drop vapor diffusion method. LBM NC1 crystals grow as small clusters overnight in 10% (w/v) PEG 20K, 0.1 M Bicine buffer (pH 9.0) at room temperature. Diffraction quality crystals were grown using microseeding procedures under similar conditions with lower protein concentration. The crystals belong to monoclinic $P2_1$ space group with unit cell dimensions $a = 129.41 \text{ \AA}$, $b = 143.87 \text{ \AA}$, $c = 162.92 \text{ \AA}$, and $\beta = 91.3^\circ$ at room temperature and four hexamers in the asymmetric unit. Cryocooling of the crystals in 25% 2,4-methyl pentanediol (MPD) or glycerol results in the shrinkage of the unit cell ($a = 127.16 \text{ \AA}$, $b = 139.57 \text{ \AA}$; $c = 160.20 \text{ \AA}$; $\beta = 91.3^\circ$).

Structure Determination and Refinement. Initial heavy atom soaks were carried out at the crystallization pH and later switched to neutral pH with phosphate buffer. NC1 crystals soaked in synthetic mother liquor containing 2mM LuCl_3 or K_2PtCl_6 transform the lattice to a smaller unit cell of dimensions $a = 79.79 \text{ \AA}$, $b = 137.20 \text{ \AA}$, $c = 126.69 \text{ \AA}$, $\beta = 90.3^\circ$ and two hexamers in the asymmetric unit. The crystals were routinely transformed into new form by soaking in 2 mM LuCl_3 overnight and used for further heavy atom soakings. Multiwavelength anomalous diffraction (MAD) data sets were collected at peak, inflection and two remote wavelengths using a single crystal soaked in 0.5 M KBr for 1 min and flash-frozen in cold N_2 stream (Table 1).

WO 03/012122

PCT/US02/23763

The heavy atom soak screens were carried out at beamlines 1-5 and 9-2 of Stanford Synchrotron Radiation Laboratory (SSRL) and beamline X8C of National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. The Br-MAD data sets used in this study were collected at SSRL and processed using DENZO and

5 SCALEPACK programs of HKL2000 suite (39). The Br⁻ sites were located using SOLVE program (40) and 37 highest peaks ($> 6\sigma$) were used for phasing the reflections at 2.2 Å resolution. The resulting phases were improved by solvent flattening using RESOLVE (41) and the electron density map was calculated using FFT program of CCP4 suite (42). Polypeptides of two $\alpha 1$ chains and one $\alpha 2$ chain

10 (chains A-C) were traced using the TOM FRODO graphics program (43). The complete asymmetric unit was generated using non-crystallographic symmetry ("NCS") relations obtained from Br⁻ sites—first the second trimer (chains D-F) was generated to complete one hexamer and then the second hexamer (chains G-L) was generated from the first hexamer. The 2.0 Å data set collected at 0.8856 Å (λ_4) was

15 used for model refinement using CNS program (44) and 5% of the data were set aside for monitoring R_{free} . The initial model was subjected to rigid body refinement using reflections in the 30.0-3.0 Å resolution range ($R = 0.361$ and $R_{free} = 0.364$) followed by simulated annealing refinement in the 10.0-2.5 Å resolution range ($R = 0.326$ and $R_{free} = 0.287$). Resolution was slowly extended to 2.0 Å in several iterative cycles of

20 model building and refinement of positional and thermal parameters. During the final rounds of refinement, solvent molecules (water and glycerol) and Br⁻ ions were added in steps using $2F_o-F_c$ and F_o-F_c maps and hydrogen bonding criteria. Multiple conformers of a few sidechains were modeled in the final round. The structure was analyzed using SETOR(45), GRASP(46), and SURFNET(47) graphics software

25 packages and various utility programs in CCP4 suite. The hexamer interface was analyzed using HBPLUS(48) and protein-protein interaction web server (<http://www.biochem.ucl.ac.uk/bsm/PP/server/>).

WO 03/012122

PCT/US02/23763

Table 1. Summary of Crystallographic Analysis

Data Collection Dataset	Peak	Inflection	Remote1	Remote2
Wavelength (Å)	0.9195	0.9197	0.9537	0.8856
Resolution (Å)	2.1	2.1	2.15	2.0
Measured reflections	602,172	603,309	568,640	686,286
Unique reflections	159,617	159,667	149,817	184,445
Completeness (%) [*]	98.3 (90.9)	98.2 (90.5)	98.7 (95.1)	97.9 (87.8)
R _{sym} (%) [†]	4.0 (7.7)	3.0 (6.7)	2.4 (4.9)	3.4 (8.6)
I/σ(I)	29.2(15.0)	33.0 (18.2)	37.6 (26)	30.5 (13.1)
Phasing Statistics				
Resolution range (Å)	50.0 – 2.2			
Number of Br sites	33			
Overall Z-score	127			
Figure of Merit SOLVE / RESOLVE	0.67 / 0.76			
Refinement Statistics				
Resolution range (Å)	8.0 – 2.0			
Number of reflections (σ>2) working / test	166,448 / 8,789			
R _{wp} / R _{int} (%) [‡]	17.0 / 19.6			
RMS deviation				
Bond lengths (Å)	0.0051			
Bond angles (°)	1.29			

^{*}The overall completeness is given, with the completeness in the highest resolution shell shown in the parentheses. Similar convention is followed for R_{sym} and I/σ(I) also. [†]R_{sym} = $\sum_h \sum_l |\langle I(h) \rangle - I(h)| / \sum_h \sum_l I(h)$. [‡] 5% of the data were excluded from refinement and were used to determine the R_{int}. The R_{wp} does not include these reflections. In both cases $R = \sum (|F_o| - k|F_c|) / \sum F_o$, with an appropriate choice of reflections for the summation.

RESULTS AND DISCUSSION

Structure Determination and Overview. The bovine LBM NC1 hexamer, composed of α1 and α2 chains, crystallizes in monoclinic space group P2₁ (A-form) with four hexamers per asymmetric unit. This is different from the crystal forms reported for mouse EHS tumor NC1 (49) and human placenta NC1 hexamers (50), which

WO 03/012122

PCT/US02/23763

crystallized with two hexamers and one hexamer in the asymmetric unit respectively. The intensity statistics of the preliminary diffraction data suggested the presence of pseudo-translation symmetry along the *c* axis in LBM NC1 crystals. An extensive search for heavy atom derivatives using soaking experiments was not successful.

5 However, crystals soaked in LuCl_3 at pH 7.0 transformed the lattice to a smaller unit cell as a result of pseudo-translation symmetry becoming crystallographic translation in the same space group with only two hexamers in the asymmetric unit (B-form). MAD data of the crystals soaked in LuCl_3 did not provide useful phase information, probably due to a single weak binding site that was responsible for lattice

10 transformation. However, we took advantage of the smaller unit cell for further heavy atom screening, including the newly suggested short-soaking strategy with halides (51,52). The LuCl_3 -soaked B-form crystal structure was determined at 2.0 Å resolution by the MAD method using Br^- as the anomalous scatterer combined with solvent flattening. The data collection, phasing and refinement statistics are shown in

15 **Table 1.**

The map was fitted with human NC1 $\alpha 1$ and $\alpha 2$ sequences (Fig. 2) since neither of the bovine sequences is available. Four each of the $\alpha 1$ and $\alpha 2$ sequences of other mammalian species are known, which share more than 95% sequence identity among them. More than 95% of the residues of the human sequences fit experimental

20 electron density map. Differences between the human sequences and the map were found for residues Ile15Thr, Ser22Pro, Pro129Gln in $\alpha 1$ chain and Asp96Glu, Glu97Asp, and Gly176Ala in $\alpha 2$ chain. The sequences are numbered so that the residue after the last Gly-Xaa-Yaa repeat of the collagenous region is counted as the first residue in both α chains. The 12 chains in two hexamers have been assigned

25 chain IDs A-L in the order of $\alpha 1$, $\alpha 1$ and $\alpha 2$ in each trimer. The map shows disorder for 5-6 residues at N- and two residues at C-termini of all the chains. The final model includes two hexamers, 36 Br^- ions, 48 glycerol molecules and 1139 water molecules. The final R-factor and R_{free} of the refinement are 0.168 and 0.197 respectively. More than 90% of the residues are within the most favorable regions in Ramachandran map

30 and Arg76 and Ser148 of the first $\alpha 1$ chain, Ser148 of the second $\alpha 1$ chain and Arg75, Glu95 and Ala145 of $\alpha 2$ chain in each trimer lie in the disallowed region. Only a handful of residues are in multiple conformations. The two hexamers in the

WO 03/012122

PCT/US02/23763

asymmetric unit are similar with no apparent differences due to crystal contacts. The hexamer comprising chains A-F is used to describe the model.

The overall structure of the hexamer is illustrated in Fig. 3. The two trimers in the hexamer are related by a 2-fold NCS axis at the interface ("equatorial plane") and the monomers within a trimer are related by a pseudo 3-fold symmetry coinciding with the triple helix axis ("polar axis").

Monomer Topology: The NC1 monomer folds into a novel tertiary structure with predominantly β -strands as predicted by our earlier study using multiple sequence alignment (22)(Fig. 4). The two $\alpha 1$ chains in the trimer are identical and the $\alpha 2$ chain has a similar overall structure. The C_{α} atoms of 214 matching residues in one of the $\alpha 1$ chains and the $\alpha 2$ chain superimpose with an RMS deviation of 0.9 Å. Each chain can be divided into two homologous subdomains, N- and C-subdomains. The two subdomains fold in a similar topology and C_{α} atoms of 96 matching residues of two subdomains of $\alpha 1$ chain superimpose with an RMS deviation of 1.0 Å. The 12 invariant cysteine residues form six disulfides, three in each subdomain, at conserved positions (Fig. 2 and 5). The major difference between the two subdomains occurs at the regions encompassing Pro86-Pro95 in the N-subdomain and Ile196-Thr209 in the C-subdomain, which are least conserved in the six human sequences (Fig. 1). Each subdomain has two β -sheets—a three-strand anti-parallel sheet (I & I') close to the triple helical junction and a six-strand anti-parallel sheet (II & II') close to the hexamer interface, which consists of the regions of interactions between the two trimers that make up the hexamer (Fig. 4 and 5). The β -sheet I is formed by the three non-contiguous strands ($\beta 1$, $\beta 10$ and $\beta 2$) of the sequence belonging to the first half of the polypeptide. However, in the β -sheet II, only four strands ($\beta 4$, $\beta 3$, $\beta 8$, and $\beta 9$) belong to the first half of the sequence and the remaining two strands ($\beta 6'$ and $\beta 7'$) form a part of the second half of the sequence. Thus, a β -hairpin structure from the second half of the sequence (the "intra-chain domain swapping region", or "Intra-CDSR") swaps into the N-subdomain to form a six-strand β -sheet. The two halves of the polypeptide being topologically similar, the region in the C-subdomain corresponding to the six-strand β -sheet in the N-subdomain lacks two strands to form a similar β -sheet in the isolated monomer structure. Similarly, $\beta 6$ - $\beta 7$ hairpin in the N-terminal half, which corresponds to the $\beta 6'$ - $\beta 7'$ hairpin in the C-terminal involved

WO 03/012122

PCT/US02/23763

in the domain swapping interaction, extends out in the monomer structure. These two features form the basis for the trimer organization described in the next section.

Trimer Organization: Two chains of the $\alpha 1$ NC1 domain and one chain of the $\alpha 2$ NC1 domain form the trimer structure with a pseudo 3-fold molecular symmetry. Since each chain is made up of topologically similar subdomains, there is even a pseudo 6-fold symmetry. The topology diagram of the trimer is shown in Fig. 5. The trimer structure is approximately cone-shaped with a base diameter of about 65 Å and a hollow core of about 12-14.0 Å inner diameter. This is about the same of as the diameter of the collagen triple helix, with N-termini of all three chains coming together at the vertex of the cone where the triple helical collagenous domain links with the NC1 domain. The trimer is tightly packed through several interchain hydrophobic and hydrogen bonding interactions (Table 2). Residues of five segments in the N-subdomain of one chain make contact with those of seven segments in the C-subdomain of the second chain, and constitute the “monomer-monomer interface”, which consists of the regions of monomer-monomer interaction within the trimer. The most important interactions are confined to one N-subdomain segment and two C-subdomain segments (Fig. 1). There are two levels of monomer-monomer interactions, one essential for the “generic trimer” assembly and the other dictating the α NC1 chain specificity of the monomer-monomer interactions within the trimer.

Table 2. Comparison of monomer-monomer interfaces in the trimer.

Interface Parameter	$\alpha 1A-\alpha 1B$		$\alpha 1B-\alpha 2$		$\alpha 2-\alpha 1A$	
	$\alpha 1A$	$\alpha 1B$	$\alpha 1B$	$\alpha 2$	$\alpha 2$	$\alpha 1A$
Number of segments	5	7	5	7	5	8
Number of residues	49	60	51	65	49	59
Δ ASA (Å ²)	2137	2182	2087	2066	1985	2044
Polar/non-polar atoms (%)	40.1/59.9	24.5/75.5	44.3/55.7	32.5/67.5	39.9/60.1	24.8/75.3
Hydrogen bonds	9/8/5		11/8/12		9/9/3	

Δ ASA, interface solvent accessible area; M, main chain; S, side chain

25

WO 03/012122

PCT/US02/23763

Within the trimer, the following monomer-monomer interfaces exist: $\alpha 1A$ - $\alpha 2C$; $\alpha 1B$ - $\alpha 2C$; and $\alpha 1A$ - $\alpha 1B$. The hexamer contains two such trimers; the monomer-monomer interfaces in the second trimer are $\alpha 1D$ - $\alpha 2F$; $\alpha 1E$ - $\alpha 2F$; and $\alpha 1D$ - $\alpha 1E$.

- 5 **Generic Trimer:** At the first level, the monomers intertwine with each other to form the trimer through 3D domain swapping interactions (**Fig. 5 and 6a**) (53). A six-strand β -sheet (Π') is formed in the C-subdomain from strands of two different α chains similar to the β -sheet Π in the N-subdomain formed from the strands in two halves of the same chain. These β -sheets are indistinguishable in $\alpha 1$ and $\alpha 2$ chains.
- 10 Thus, there are six β -sheets (Π/Π'), one in each of the six subdomains, forming the close-ended 3D domain swapping interactions in the NC1 trimer structure. Each of these six-strand β -sheets is formed by four strands ($\beta 4/4'$, $\beta 3/3'$, $\beta 8/8'$, $\beta 9/9'$) in one half of the sequence and the remaining two strands ($\beta 6/6'$, $\beta 7/7'$) are contributed by the other half of the same chain ($\beta 6/\beta 7$; the Inter-CDSR) or adjacent chain ($\beta 6'/\beta 7'$;
- 15 the "Intra-chain domain swapping region", or "Intra-CDSR"). The amino acid sequences of all the strands with the exception of $\beta 9$, are highly conserved in α chains within and across the species. The six topologically similar β -sheets formed in cyclical fashion give the pseudo 6-fold symmetry appearance for the trimer (**Fig. 6a**). In each of the β -sheets, the outermost strand ($\beta 9/\beta 9'$) lies on the surface parallel to the equatorial plane of the hexamer interface forming a part of the outer ring and the innermost strand ($\beta 4/\beta 4'$) runs nearly parallel to the polar axis or pseudo 3-fold axis in the core. The angle between these two strands within each sheet is about 75° giving it a right-handed twist. The $\beta 4/\beta 4'$ strands from all the six β -sheets form a parallel β barrel-like core of about 14 Å diameter even though there are no backbone
- 25 hydrogen bonds between them (**Fig. 6a**). However, these core strands are stabilized by backbone-side chain hydrogen bonds either directly or mediated through solvent molecules. The $\beta 4/4'$ strands have a mixture of hydrophobic and hydrophilic residues, with the former pointing to the core and the latter pointing towards the adjacent strand. Interestingly, the $\beta 4$ strands contain long chain hydrophilic amino
- 30 acids so that they form more direct hydrogen bonds with the backbone atoms of the $\beta 4'$ strand of the neighboring chain indicating stronger interchain interactions. The interactions between $\beta 4'$ and $\beta 4$ within a chain are mainly mediated through solvent

WO 03/012122

PCT/US02/23763

molecules. Thus, the six-strand β -sheets are essential structural components in the organization of the generic trimer structure through 3D domain swapping interactions and the compact β barrel-like core structure. However, they may play only a limited role in the chain specific assembly of the trimer. Therefore, compounds that target the

5 Intra-CDSR, the Inter-CDSR, and the $\beta 4/\beta 4'$ based β barrel-like core, such as peptides derived from these regions, can be used to inhibit generic monomer-monomer interactions, and thus to inhibit trimer assembly.

Chain Specificity in the Trimer Structure: The sequence of the loop connecting the

10 $\beta 8'$ and $\beta 9'$ strands is the most variable region in all the six human α chains (referred to as the "hypervariable region"). This hypervariability in the primary sequences manifests itself as different secondary structures in the $\alpha 1$ and $\alpha 2$ chains in the crystal structure. Whereas it forms a short 3_{10} helix ($g2'$) in all the $\alpha 1$ -like chains (the "specificity region partner" or "SRP"; Glu200-Lys204 (EMFKK)), the corresponding

15 region in $\alpha 2$ chain (Ser198-Gln200; SFQ) adopts an extended conformation ($\beta p'$) and pairs with the extended structure (the "specificity region", or "SR"; βp , Phe57-Met60; FSTM) in the adjacent $\alpha 1B$ chain to form a short parallel β -sheet (Fig. 8b). It should be noted that the sequence of the SR from $\alpha 1$ and $\alpha 2$ is identical (FSTM). This is the only parallel β -sheet in the entire structure, which is predominantly made up of

20 β -strands. The sequence of the βp is highly conserved in all the six α chains and forms the same extended structure in $\alpha 2$ chain also, even though it doesn't have a partner in $\alpha 1A$ chain to form the parallel β -sheet. Thus, these additional main chain hydrogen bond interactions between the two chains are found only at the $\alpha 1B$ - $\alpha 2$ interface (i.e.: which includes the interaction of the SR of $\alpha 1$ and the SRP of $\alpha 2$), but

25 not in $\alpha 2$ - $\alpha 1A$ (i.e.: which includes the interaction of the SRP of $\alpha 1$ and the SR of $\alpha 2$) or $\alpha 1A$ - $\alpha 1B$ (i.e.: which includes the interaction of the SR of $\alpha 1$ and the SRP of $\alpha 1$) interfaces, due to the presence of the 3_{10} helical structure in $\alpha 1$ chains rather than the extended structure present in $\alpha 2$ chain. Besides this difference in the secondary structural elements in the three interfaces, there are also differences in the main

30 chain-side chain and side chain-side chain interactions at the monomer-monomer interface (Fig. 6b). This is also reflected in different ratios of polar to non-polar atoms at the three interfaces (Table 2). Therefore, compounds that target the SR, the SRP,

WO 03/012122

PCT/US02/23763

or the hyper-variability region, such as peptides derived from these regions, can be used to inhibit specific monomer-monomer interactions, and thus inhibit trimer assembly.

5 Furthermore, given the composition of the individual interfaces within the monomer-monomer interface, a preferred inhibitor of specific trimer assembly would target the SR, which is identical in $\alpha 1$ and $\alpha 2$, and thus such an inhibitor would be expected to interfere with interactions at each interface within the monomer-monomer interface, and thus to inhibit trimer assembly. Also preferred would be an inhibitor that targets the $\alpha 2$ SRP, which is required for the additional H-bonding interactions
10 seen at the $\alpha 1$ B- $\alpha 2$ interface.

The side chain of Lys56($\alpha 1$ B) is sandwiched between the backbone of the loop preceding the parallel β -sheet in $\alpha 2$ chain and the contiguous bonds of backbone and side chain of Gln120($\alpha 2$). In this tightly locked position, Lys56($\alpha 1$ B) assumes a linear conformation to form two strong hydrogen bonds with the carbonyl of
15 Ile194($\alpha 2$) and the carboxyl of Asp121($\alpha 2$), and two more weak interactions with the carbonyls of Gln120($\alpha 2$) and Glu196($\alpha 2$). The $\alpha 1$ -like (ie: $\alpha 1/3/5$ family) region corresponding to the parallel β -sheet of $\alpha 2$ chain is the 3_{10} helix, which spans a longer sequence. Hence, in the $\alpha 1$ A- $\alpha 1$ B interface, Lys56($\alpha 1$ A) is not quite parallel to the backbone bonds, which provides more room for this lysine to adopt a different
20 rotamer conformation to form only weak hydrogen bond with the carbonyl oxygen of Ile196($\alpha 1$ B). This may also be influenced by the presence of hydrophobic Thr124 in $\alpha 1$ chains in place of hydrophilic Asp121 in $\alpha 2$. At the $\alpha 2$ - $\alpha 1$ A interface Arg55($\alpha 2$) is docked in similar position as Lys56 of $\alpha 1$ chains in other two interfaces with one strong hydrogen bond interaction with carbonyl of Ile196($\alpha 1$ A). Other differences in
25 amino acid sequences including Arg55/Ala54 and Gly98/Glu95 make differences in hydrogen bonding patterns at the interfaces. Thus, the Arg55($\alpha 2$)/Lys56($\alpha 1$) is an important residue for optimal $\alpha 1$ - $\alpha 2$ monomer-monomer interactions, and compounds targeting this region, such as peptides including LRKF (SEQ ID NO:294) ($\alpha 1$) or LARF (SEQ ID NO:295) ($\alpha 2$), can be used to inhibit the assembly of specific
30 monomer-monomer interactions. Since this region precedes the SR, this region can be combined with the SR to form a longer peptide that will interfere with multiple aspects of specific monomer-monomer interactions, and thus be even more effective at inhibiting trimer assembly.

WO 03/012122

PCT/US02/23763

Furthermore, the regions Ile194-Glu196 ($\alpha 2$), Ile196 ($\alpha 1$) and Gln120-Asp121($\alpha 2$) also are involved in optimal $\alpha 1$ - $\alpha 2$ monomer-monomer interactions, and compounds targeting these region, such as peptides including IPE (SEQ ID NO:294) ($\alpha 2$ 184-196), IER (SEQ ID NO:295) ($\alpha 1$ 196-198) or QD (SEQ ID NO:296) ($\alpha 2$ 120-121), can be used to inhibit the assembly of specific monomer-monomer interactions, and thus to inhibit trimer assembly.

The $\alpha 1$ B- $\alpha 2$ interface (i.e.: which includes the interaction of the SR of $\alpha 1$ and the SRP of $\alpha 2$) has the maximum number of contact residues, the highest proportion of hydrophilic atoms, and contains more hydrogen bonds than the other monomer-monomer interfaces (Table 2). On the other hand, the buried surface area is largest for $\alpha 1$ A- $\alpha 1$ B interface (i.e.: which includes the interaction of the SR of $\alpha 1$ and the SRP of $\alpha 1$). From these observations, it is evident that the $\alpha 1$ B- $\alpha 2$ interface is formed predominantly through hydrogen bonding interactions and the $\alpha 1$ A- $\alpha 1$ B interface is stabilized by more hydrophobic forces.

In addition to the specific interactions at the interfaces, packing considerations may also play an important role in determining chain stoichiometry in the trimer. Even though the $\alpha 1$ and $\alpha 2$ NC1 chains fold in a similar tertiary structure with a low RMS deviation, the relative orientation of the two subdomains in each NC1 chain is different near the triple helical junction. The region encompassing Thr13-Tyr30 of the N-subdomain in the $\alpha 2$ chain is farther from its equivalent region Asp121-Tyr138 of the C-subdomain in the $\alpha 2$ chain compared to the relative orientations of similar regions in the $\alpha 1$ structure. The larger width of the $\alpha 2$ structure near the triple helical junction results in serious steric clashes when packed into a hypothetical $\alpha 2$ -homotrimer. However, it is possible to accommodate three $\alpha 1$ chains in a hypothetical homotrimer, albeit with weaker interactions.

It is preferred that peptides designed to interfere with monomer-monomer interactions are preferably delivered into the cell, where such monomer-monomer assembly occurs. Alternatively, the peptides can be used to disrupt assembled trimers that have been secreted by the cell.

Hexamer Assembly: The type IV collagen trimer, once formed in the endoplasmic lumen, is secreted into the extracellular space where it assembles into the hexamer, and then into a supramolecular network through N- and C-terminal associations. The

WO 03/012122

PCT/US02/23763

NC1 domains play the dominant role in this assembly, by determining the C-terminal dimeric association, leading to hexamer assembly. In this section we describe the forces that influence such hexamer assembly as observed in the crystal structure, and provide a rationale for the specificity in the type IV collagen network assembly.

5 The foot-ball shaped hexamer is made up of two identical trimers, each containing two $\alpha 1$ chains and one $\alpha 2$ chain as described in the previous section. Each protomer (ie: the complete type IV collagen trimer, including NC1 domains) formed by the tightly intertwined trimer is considered as a single entity so that the hexamer can be analyzed relative to other homodimeric protein complexes (43). We have
10 determined several parameters defining the hexamer interface to evaluate the strength of interactions between the two trimers and analyze hexamer assembly in the type IV collagen network (Table 3).

Table 3. Comparison of interface parameters defining the trimer-trimer interaction in the NC1 hexamer and observed mean for 32 homodimer complexes.

Interface Parameter	NC1 Hexamer	Observed Mean (43) (32 Homodimers)
Δ ASA (\AA^2)	4173.1	1685.03
Planarity	1.91	3.46
Circularity	0.87	0.71
Segmentation	18	5.22
Hydrogen bonds per 100 \AA^2	1.2	0.70
Gap Index	1.24	2.2

15 Percentage of polar and non-polar atoms are 45.5 and 54.5 respectively.

Like most homodimers, the two NC1 trimers are related by a 2-fold NCS axis in lying the equatorial plane and perpendicular to the pseduo 3-fold axis of symmetry within an individual trimer (Fig. 4). This symmetry constraint may be partly
20 influenced by a few differences in the interface residues of $\alpha 1$ like and $\alpha 2$ like sequences in addition to more efficient packing. The hexamer interface is formed by the nearly flat surfaces of the two trimers, with an RMS deviation of 1.9 \AA for all the hexamer interface atoms from the mean plane (Fig. 9a). This is significantly lower
25 than the average planarity value of 3.5 \AA for 32 homodimers discussed in a recent review (43). The hexamer interface formed by six segments each of the three

WO 03/012122

PCT/US02/23763

monomers, with a total of 109 residues per trimer, is nearly circular, with the major and minor axial lengths of the mean plane measuring approximately 69 and 61 Å respectively. This flat circular hexamer interface covers about 4400 Å² of solvent accessible area per trimer, which correlates with the observation of larger molecules
5 having larger interfaces (54). Such a large interface facilitates strong interaction between the trimers, involving both hydrophobic and hydrophilic residues. The polar (45.5%) and non-polar atoms (54.5%) in the hexamer interface are nearly in equal proportions, underscoring the importance of both types of interactions in hexamer stabilization.

10 The discussion thus far focused on the overall nature of the hexamer interface. Next, the interactions between the individual chains at the hexamer interface are analyzed in more detail. Each monomer of one trimer makes contact with two monomers of the other trimer, designated as the "major" and "minor" contacts based on the extent of the contact area and number of hydrogen bonds. The two monomers
15 making major contact is referred to as "dimer" in a similar sense as the term used in the denaturation experiments of hexamers (55). The 2-fold NCS between the two trimers results in only one "homodimer" formed by two $\alpha 1$ chains (Figure 7A), with the remaining two "heterodimers" formed by $\alpha 1$ and $\alpha 2$ chains (Figure 7A-B).

A 120° rotation of one trimer with respect to the other about the pseudo 3-fold
20 axis will result in an "all homodimers" structure. Why such an arrangement is not possible can be explained mainly on symmetry consideration: breaking the symmetry results in less efficient packing with possibly fewer interactions and some unfavorable contacts. In order to understand the complex hydrogen bonding interactions at the interface, it is essential to look into the interactions of each monomer with its "major"
25 and "minor" interacting partners. The complexity presented even at this level may be simplified further by breaking down the interactions to three regions in the structure: "core" and "outer" regions of "major" contact and the "major-minor junction".

Core regions of major contact: The two 6-strand β -sheets, II and II', formed by the 3D domain swapping interactions play as crucial role in the formation of
30 hexamer assembly as in the case of trimer organization. The hexamer interface is populated with β -turns connecting $\beta 3$ - $\beta 4$ and $\beta 3'$ - $\beta 4'$ in the core. These turns along with the remaining strands of the β -sheets II/II' position a large number of conserved residues for extensive hydrogen bonding interactions at the hexamer interface. The

WO 03/012122

PCT/US02/23763

core β -turns (two per monomer contributed by the two equivalent subdomains) in the two trimers pack in staggered configuration such that each turn in one trimer contacts with two turns in the other trimer. The turns in the N-subdomains are of type I'/III' containing hydrophilic amino acids in the second (Asn39/Gln38) and third positions
5 (Glu40/39). The C-subdomain turns are of type II in α 1 chains and type II' in α 2 chains with small hydrophobic amino acids, Ala149/146-Gly150/147-Ala151/Asp148, with Ala149 α 1 or Asp148 of α 2 introducing a β -bulge. Thus, the hydrophilic side chains of turns in the N-subdomain participate in hydrogen bonds and hydrophobic residues of turns in C-subdomain pack through hydrophobic interaction as well as
10 stacking interaction of peptide planes (Fig. 7A). Whereas the Asn39(Gln38) side chain in the N-subdomain forms a hydrogen bond with the backbone amide in C-subdomain turn, the conserved Glu40(39) penetrates between the N- and C-subdomains of a monomer chain in the other trimer to form a hydrogen bond with the side chain of the conserved Gln37(36). The Glu40 residues in the α 1- α 1 dimer form a
15 strong hydrogen bond with each other that is missing in α 1- α 2 dimers. The packing of the turns and side chains appear to be tight at the core interface in CPK models indicating strong van der Waals interactions in additions to the obvious hydrogen bonding interactions. Therefore, compounds that target the core regions of major contact at the hexamer interface, such as peptides derived from these regions, can be
20 used to inhibit hexamer assembly. For example, peptides including the β 3- β 4 connecting region or the β 3'- β 4' connecting region, can be used to inhibit hexamer assembly at the core region of major contact.

Outer regions of major contact: The sequence variability preceding
25 Arg179(177), influences the number of potential H bonds at the α 1- α 2 (hexamer) interface. The interactions in the outer region involve the highly conserved loop connecting the β 7 and β 8, and β 7'- β 8' sheets. In the α 1- α 1 major interface of the hexamer, five contiguous carbonyl oxygens of highly conserved Ala74-Asp78 in one chain form hydrogen bonds with side chains Asn77, Arg179, and Tyr185 of the other
30 chain in symmetrical sets (Fig. 9c). These side chains are also conserved in both α 1 and α 2 chains. However, insertion of Gly176 and substitution of Asn174 in α 2 sequence alters the orientation conserved Asn78 and Arg177 residues, which results in the few hydrogen bonds in the α 1- α 2 interface. Therefore, compounds that target

WO 03/012122

PCT/US02/23763

the outer regions of major contact at the hexamer interface, such as peptides derived from these regions, can be used to inhibit hexamer assembly. For example, peptides including the sequence ASRND (SEQ ID NO:201) ($\alpha 1$) or YYANA (SEQ ID NO:218) ($\alpha 1$), or the corresponding sequences in the other alpha chains, can be used to inhibit hexamer assembly at the outer region of major contact.

Major-minor junction: The major-minor junction is the area of the hexamer interface where two chains from one trimer contact two chains of the other trimer. There are two types of junctions, one involving three $\alpha 1$ and one $\alpha 2$ chains, and the other involving two each of $\alpha 1$ and $\alpha 2$ chains. The hydrogen bonding pattern in the two junctions is highly conserved (Figure 7C). Both $\alpha 1$ - $\alpha 1$ and $\alpha 2$ - $\alpha 2$ form a Asn187(185)-Tyr189(188) (NYX) (SEQ ID NO:297) hydrogen bond pairs in the interface. In addition to this, Asn187(185) forms a pair of hydrogen bonds with Arg76(75) of another chain (within the outer region of major contact discussed above) from the opposite trimer. The multiple hydrogen bonds formed by Asn187(185) involving residues from two different chains is probably one of the major factors stabilizing the trimer-trimer interface. Therefore, compounds that target major-minor junction at the hexamer interface, such as peptides derived from these regions, can be used to inhibit hexamer assembly. For example, peptides including the sequence NYX (SEQ ID NO:288) ($\alpha 1$) (such as ECHGRGTCNYX (SEQ ID NO:172)), or corresponding sequences in the other α chains, all of which is present at the hexamer interface (and which includes a large portion of the Intra-CDSR), or ASRND (SEQ ID NO:201) ($\alpha 1$) (which includes the ARG76(75) residue), or corresponding sequences in the other α chains, can be used to inhibit hexamer assembly at the major-minor junction. Thus, peptides containing the sequence ASRND (SEQ ID NO:201) can interfere with hexamer assembly by interfering with interactions at both the outer region of major contact and the major-minor junction. Similarly, peptides that target the Intra-CDSR and extend to contain the 2 additional Y residues from the sequence "NYX" (SEQ ID NO:288) can be used to inhibit trimer assembly, as well as hexamer assembly.

Other residues that are located at the hexamer interface, and that are believed to be important for hexamer assembly, include (1) MSMAP (SEQ ID NO:129)

WO 03/012122

PCT/US02/23763

(residues 91-95 $\alpha 1$)/MMP (SEQ ID NO:132) ($\alpha 2$), and corresponding sequences in the other α chains; (2) PSTLK (SEQ ID NO:177) (residues 208-212 in $\alpha 1$; $\beta 9$ ' strand; ADTLK in $\alpha 2$ (SEQ ID NO:180)), and corresponding sequences in the other α chains; (3) FCNINNVCFNA (SEQ ID NO:289) ($\alpha 1$ AND $\alpha 5$ -co-extensive with the Inter-
 5 CDSR), and corresponding sequences in the other α chains:
 $\alpha 3$: FCNVNDVCFN (SEQ ID NO:298)
 $\alpha 2$: YCNPGDVCFY (SEQ ID NO:299)
 $\alpha 4$: YCNHQQVCFY (SEQ ID NO:300)
 $\alpha 6$: YCNINEVCFY (SEQ ID NO:301)
 10 Thus, peptides containing these sequences, or portions thereof, can be used to inhibit hexamer assembly.

Disulfide bonds: Interchain or Intrachain?

Disulfide cross-linking is a recurring theme in collagen assembly and is
 15 believed to play an important role in the stabilization of the trimeric structure (11). Fibrillar procollagens are believed to form interchain disulfide bonds catalyzed by protein disulfide isomerase in either the C-telopeptide or C-propeptide (56, Kiovu, 1987 #343). Interchain disulfides have been proposed to form both in the collagenous and NC1 domains of type IV collagen. Whereas the interchain disulfides in the
 20 collagenous domains are formed within a protomer to stabilize the collagen triple helix, those in the NC1 domains are believed to occur between the protomers to stabilize the network at the C-terminus. Disulfide exchange between NC1 domains of similar α chains from two different protomers was proposed as one of the major stabilizing forces in the **hexamer** assembly (57). Under denaturing conditions, the
 25 human placenta derived NC1 hexamer dissociated as dimers and monomers. The dimers were shown to be crosslinked predominantly by disulfide bridges. However, a later study by Langeveld et al (55) comparing the NC1 hexmers isolated from several BMs revealed rather complex results. Whereas the results of placenta BM and kidney glomerular BM NC1 hexamers agreed with the previous observations, dissociating as
 30 dimers upon denaturation, the LBM NC1 hexamer dissociated predominantly as monomers implying the absence of disulfide cross-linking. The crystal structure of LBM NC1 hexamer reveals just that—all the cysteines are involved in intrachain disulfides.

WO 03/012122

PCT/US02/23763

Siebold et al (57) proposed disulfide exchanges involving Cys20(20')-Cys111'(111) and Cys53(53')-Cys108'(108) pairs in N-subdomain (and those in similar positions in C-subdomain) in $\alpha 1$ chain resulting in a total of four disulfide crosslinkings in each subdomain based on the cynogen bromide. The topological arrangement of disulfides observed in the crystal structure suggests the possibility for such a rearrangement is extremely remote (Figure 7A). The disulfides in the NC1 monomer are arranged in three tiers with Cys20-Cys111 and Cys130-Cys225 are close to the triple helical junction, Cys65-Cys71 and Cys176-Cys182 are close to the interface and Cys53-Cys108 and Cys164-Cys222 lies in between. The disulfide pairs Cys20-Cys111 and Cys53-Cys108 in the monomers of $\alpha 1A$ - $\alpha 1D$ dimer are about 70 Å and 50 Å apart respectively. Thus the possibility for disulfide exchange, if any, exists only for the Cys65-Cys71 and Cys176-Cys182 pairs. However, the staggered arrangement of the two trimers brings Cys65-Cys71 pair of $\alpha 1A$ closer to its C-subdomain equivalent Cys176'-Cys182' pair of $\alpha 1D$ chain rather than its counterpart Cys65'-Cys71' in the N-subdomain. These two closest disulfide pairs in $\alpha 1A$ - $\alpha 1D$ dimer are about 16 Å from each other. Even more importantly, these intrachain disulfides are located in the 3D domain-swapped β -hairpin regions. If the disulfide exchanges were indeed possible between these pairs it would involve major conformational alterations. Such a movement of the β -hairpins containing the "exchangeable" cysteine residues would break both the interchain and intrachain 3D domain swapping interactions, thus destabilizing the trimer structure. From these arguments, it is difficult to envisage disulfide cross-linking between the monomers belonging to two protomers in the present structure. We also examined the possibility of intra-protomer disulfides, which would also require major conformational changes and potentially move the N-terminii of the three chains severely affecting collagen-NC1 linkage. An alternative conformation must exist for the NC1 domains from all other BMs to account for the inter-protomer disulfide cross-linkings.

Biological Significance. There is very little crystallographic data available on non-collagenous domains. The only available structures of non-collagenous domains are those of endostatins (58,59), which are homologous fragments of single chains from types XVIII and XV collagens.

WO 03/012122

PCT/US02/23763

The present work provides the first unambiguous structural basis for the chain stoichiometry of the type IV collagen $\alpha 1. \alpha 2$ network, as well as the structural basis for chain specific assembly of type IV collagen. The NC1 monomer folds into a novel tertiary structure and the close ended-trimer of $(\alpha 1)_2. \alpha 2$ is organized through unique
5 3D domain swapping interactions. These features must be conserved in all type IV collagen networks, from all species, due to overall sequence similarity and very high sequence identity of the regions participating in domain swapping. The chain specificity is determined by the differences in the primary sequences of the hypervariable regions of the NC1 domains of the constituent chains, which manifest
10 as different secondary structures at the monomer-monomer interfaces. The hexamer structure is stabilized by the extensive hydrophobic and hydrophilic interactions at the trimer-trimer interface without a need for disulfide cross-linking. The crystal structure of LBM NC1 hexamer and the denaturation studies of NC1 hexamers from several BMs suggest an alternative conformation must exist in hexamers that are
15 cross-linked by interchain disulfides. Some hitherto unknown enzymatic process might be responsible for folding the same amino acid sequences into different conformations in different tissues.

20

WO 03/012122

PCT/US02/23763

References

1. Timpl, R., and Brown, J. C. (1996) *Bioessays* 18(2), 123-131
2. Weber, M. (1992) *Kidney International* 41, 620-628
3. Piilajaniemi, T. (1996) in *Molecular Pathology and Genetics of Alport Syndrome* (Trygvasson, K., ed) Vol. 117, pp. 46-79, Karger, Basel
- 5 4. Miner, J. (1999) *Kidney International* 56, 2016-2024
5. Prockop, D. J., and Kivirikko, K. I. (1995) *Ann. Rev. Biochem.* 64, 403-34
6. Myllyharju, J., and Kivirikko, K. I. (2001) *Ann Med* 33, 7-21
7. Kadler, K. (1994)
- 10 8. Bachinger, H.-P., Bruckner, P., Timpl, R., Prockop, D. J., and Engel, J. (1980) *Eur. J. Biochem.* 106, 619-632
9. Bachinger, H.-P., Fessler, L. I., Timpl, R., and Fessler, J. H. (1981) *J. Biol. Chem.* 256, 13193-13199
10. Dolz, R., Engel, J., and Kuhn, K. (1988) *Eur J Biochem* 178(2), 357-66
- 15 11. McLaughlin, S. H., and Bulleid, N. J. (1998) *Matrix Biology* 16, 369-377
12. Lees, J. F., Tasab, M., and Bulleid, N. J. (1997) *EMBO J.* 16(5), 908-916
13. Dion, A. S., and Myers, J. C. (1987) *J Mol Biol* 193(1), 127-43
14. Rosenbloom, J., Endo, R., and Harsch, M. (1976) *J. Biol. Chem.* 251, 2070-2076
- 20 15. Schofield, D. J., Uitto, J., and Prockop, D. J. (1974) *Biochemistry* 13, 1801-1806
16. Uitto, V., Uitto, J., and Prockop, D. J. (1981) *Arch. Biochem. Biophys.* 210, 445-454
17. Boutaud, A., Borza, D.-B., Bondar, O., Gunwar, S., Netzer, K.-O., Singh, N.,
- 25 Ninomiya, Y., Sado, Y., Noelken, M. E., and Hudson, B. G. (2000) *J. Biol. Chem.* 275, 30716-30724
18. Borza, D. B., Bondar, O., Ninomiya, Y., Sado, Y., Naito, I., Todd, P., and Hudson, B. G. (2001) *J Biol Chem* 276(30), 28532-40.
19. Hudson, B. G., Reeders, S. T., and Trygvasson, K. (1993) *J Biol Chem* 268(35), 26033-6
- 30 20. Timpl, R., Wiedemann, H., van Delden, V., Furthmayr, H., and Kuhn, K. (1981) *Eur J Biochem* 120(2), 203-11
21. Zhou, J., Ding, M., Zhao, Z., and Reeders, S. T. (1994) *J. Biol. Chem.* 269, 13193-13199

WO 03/012122

PCT/US02/23763

22. Netzer, K. O., Suzuki, K., Itoh, Y., Hudson, B. G., and Khalifah, R. G. (1998) *Protein Sci* 7(6), 1340-51
23. Fowler, S. J., Jose, S., Zhang, X., Deutzmann, R., Sarra, M. P., Jr., and Boot-Handford, R. P. (2000) *J Biol Chem* 275(50), 39589-99.
- 5 24. Boute, N., Exposito, J. Y., Boury-Esnault, N., Vacelet, J., Noro, N., Miyazaki, K., Yoshizato, K., and Garrone, R. (1996) *Biol Cell* 88(1-2), 37-44
25. Guo, X. D., and Kramer, J. M. (1989) *J Biol Chem* 264(29), 17574-82.
26. Sibley, M. H., Johnson, J. J., Mello, C. C., and Kramer, J. M. (1993) *J Cell Biol* 123(1), 255-64.
- 10 27. Blumberg, B., MacKrell, A. J., and Fessler, J. H. (1988) *J Biol Chem* 263(34), 18328-37.
28. Exposito, J. Y., D'Alessio, M., Di Liberto, M., and Ramirez, F. (1993) *J Biol Chem* 268(7), 5249-54.
29. Gunwar, S., Ballester, F., Noelken, M. E., Sado, Y., Ninomiya, Y., and Hudson, B. G. (1998) *J Biol Chem* 273(15), 8767-75
- 15 30. Zhang, X., Hudson, B. G., and Sarra, M. P., Jr. (1994) *Dev Biol* 164(1), 10-23
31. Guo, X., Johnson, J. J., and Kramer, J. M. (1991) *Nature* 349, 707-709
32. Sibley, M. H., Graham, P. L., von Mende, N., and Kramer, J. M. (1994) *EMBO J.* 13, 3278-3285
- 20 33. Kashtan, C. E., and Michael, A. F. (1993) *Am. J. Kid. Dis.* 22, 627-640
34. Kashtan, C. E., and Michael, A. F. (1996) *Kidney Int* 50, 1445-1463
35. Cosgrove, D., Meehan, D. T., Grunkemeyer, J. A., Kornak, J. M., Sayers, R., Hunter, W. J., and Samuelson, G. C. (1996) *Genes Dev* 10(23), 2981-92.
36. Miner, J. H., and Sanes, J. R. (1996) *J Cell Biol* 135(5), 1403-13.
- 25 37. Gunwar, S., Noelken, M. E., and Hudson, B. G. (1991) *J Biol Chem* 266(21), 14088-94
38. Peczon, B. D., McCarthy, C. A., and Merrit, R. B. (1982) *Exp. Eye. Res.* 35, 643-651
39. Otwinowski, Z., and Minor, W. (1997) *Methods in Enzymology* 276, 307-326
- 30 40. Terwilliger, T. C., and Berendzen, J. (1997) *Acta Crystallogr.* D55, 849-861
41. Terwilliger, T. C. (2000) *Acta Crystallogr D* 56(Pt 8), 965-72.
42. Dodson, E. J., Winn, M., and Ralph, A. (1997) *Methods in Enzymology* 277, 620-633
43. Jones, S., and Thornton, J. M. (1996) *Proceedings of The National*

WO 03/012122

PCT/US02/23763

Academy of Science (U.S.A) 93, 13-20

44. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Pannu, N. S., and al., e. (1998) *Acta Crystallogr.* D54, 905-921
- 5 45. Evans, S. V. (1993) *J. Mol. Graphics* 11, 134-138
46. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* 11, 281-296
47. Laskowski, R. A. (1995) *J. Mol. Graph.* 13., 323-330
48. McDonald, I. K., and Thornton, J. M. (1994) *J. Mol. Biol.* 238, 777-793.
49. Timpl, R., Oberbaumer, I., von der Mark, H., Bode, W., Wick, G., Weber, S.,
10 and Engel, J. (1985) *Ann NY Acad Sci* 460, 58-72
50. Stubbs, M., Summers, L., Mayr, I., Schneider, M., Bode, W., Huber, R., Ries, A., and Kuhn, K. (1990) *J Mol Biol* 211, 683-684
51. Dauter, Z., and Dauter, M. (1999) *J. Mol. Biol.* 289, 93-101
52. Dauter, Z., Dauter, M., and Rajashankar, K. R. (2000) *Acta Crystallogr.* D56, 232-237
- 15 53. Schlunegger, M. P., Bennett, M. J., and Eisenberg, D. (1997) *Advances in Protein Science* 50, 61-132
54. Jones, T. A. (1978) *J. Appl. Crystallogr.* 11, 268-272
55. Langeveld, J. P., Wieslander, J., Timoneda, J., McKinney, P., Butkowski, R. J., Wisdom, B. J., Jr., and Hudson, B. G. (1988) *J Biol Chem* 263(21), 10481-8
- 20 56. Uitto, J., and Prockop, D. J. (1973) *Biochem. Biophys. Res. Commun.* 55, 904-911
57. Siebold, B., Deutzmann, R., and Kuhn, K. (1988) *Eur J Biochem* 176(3), 617-24
- 25 58. Hohenester, E., Sasaki, T., Olsen, B. R., and Timpl, R. (1998) *Embo J.* 17, 1656-1664
59. Sasaki, T., Larsson, H., Tisi, D., Claesson-Welsh, L., Hohenester, E., and Timpl, R. (2000) *J. Mol. Biol.* 301, 1179-1190
60. Petitclerc, E., Boutaud, A., Prestayko, A., Xu, J., Sado, Y., Ninomiya, Y., Sarras, M. P., Jr., Hudson, B. G., and Brooks, P. C. (2000) *J Biol Chem* 275(11), 8051-61
- 30 61. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* 26, 283-291
62. Barton, G. J. (1993) *Prot. Eng.* 6, 37-40

WO 03/012122

PCT/US02/23763

WO 03/012122

PCT/US02/23763

1. A polypeptide consisting of at least 8 contiguous amino acids of general
5 formula I:
PF(R1)(R2)CN(R3)(R4)(R5)VC(R6)(R7)A (SEQ ID NO:1)
R1 is selected from the group consisting of L, M, A, V, norL, and I;
R2 is selected from the group consisting of F and Y;
R3 is selected from the group consisting of I, V, L, norL, A, and P;
10 R4 is selected from the group consisting of N, G, and H;
R5 is selected from the group consisting of N, D, Q, and E;
R6 is selected from the group consisting of N, Y, and H; and
R7 is selected from the group consisting of F and Y.
2. The polypeptide of claim 1 consisting of the amino acid sequence of general
15 formula I.
3. The polypeptide of claim 1, wherein
R2 is F;
R4 is N;
R5 is selected from the group consisting of N and D;
20 R6 is N; and
R7 is F.
4. The polypeptide of claim 1, wherein
R2 is Y;
R3 is selected from the group consisting of P and I;
25 R5 is selected from the group consisting of D, Q, and E;
R6 is selected from the group consisting of Y and H; and
R7 is Y.
5. The polypeptide of claim 1 wherein the polypeptide is selected from the group
consisting of PFLFCNINNVCNFA (SEQ ID NO:2); PFLFCNVNDVCNFA (SEQ
30 ID NO:3); PFMFCNINNVCNFA (SEQ ID NO:4); PFLYCNPGDVCYYA (SEQ ID
NO:5); PFAYCNIHQVCHYA (SEQ ID NO:6); and PFYCNINEVCHYA (SEQ ID
NO:7).

WO 03/012122

PCT/US02/23763

6. A polypeptide consisting of at least 7 contiguous amino acids of general formula II:
PF(R1)EC(R2)G(R3)(R4)GTC(R5) (SEQ ID NO:8)
R1 is selected from the group consisting of L, A, V, norL, and I;
R2 is selected from the group consisting of H, N, Q, and S;
R3 is selected from the group consisting of G, R, A, or is absent;
R4 is selected from the group consisting of R and Q; and
R5 is selected from the group consisting of N and H.
7. The polypeptide of claim 6 consisting of the amino acid sequence of general formula II.
8. The polypeptide of claim 6 wherein
R2 is H;
R3 is R;
R4 is G; and
R5 is N.
9. The polypeptide of claims 6 wherein
R2 is selected from the group consisting of N, Q, and S;
R3 is selected from the group consisting of G, R, and A;
R4 is selected from the group consisting of R and Q; and
R5 is H.
10. The polypeptide of claim 6 wherein the polypeptide is selected from the group consisting of PFIECHGRGTCN (SEQ ID NO:9); PFLECHGRGTCN (SEQ ID NO:10); PFIECNGGRGTCH (SEQ ID NO:11); PFLECQGRQGTCH (SEQ ID NO:12); and PFIECSGARGTCH (SEQ ID NO:13).
11. A polypeptide consisting of at least 13 amino acids selected from the group consisting of:
(a) EFRSAPFIECHGRGTCNYYANA (SEQ ID NO:14),
(b) EFRASPFLIECHGRGTCNYYSNS (SEQ ID NO: 15);
(c) EFRSAPFIECHGRGTCNYYANS (SEQ ID NO: 16);
(d) DFRATPFIECNGGRGTCHYYA)NK (SEQ ID NO: 17);
(e) DFRAAPFLECQGRQGTCHFFANK (SEQ ID NO: 18); and
(f) DFRATPFIECSGARGTCHYFANK (SEQ ID NO: 19)

WO 03/012122

PCT/US02/23763

12. A chimeric polypeptide consisting of:
- (a) a polypeptide according to claim 1;
 - (b) a polypeptide according to claim 6; and
 - (c) a polypeptide linker consisting of between 0 and 20 amino acids.
- 5
13. The chimeric polypeptide of claim 12 consisting of
- (a) a polypeptide according to claim 5;
 - (b) a polypeptide according to claim 10; and
 - (c) a polypeptide linker consisting of 2 amino acids.
- 10
14. A polypeptide consisting of a sequence of general formula III:
F(R1)T(R2) (SEQ ID NO:20)
wherein R1 is selected from the group consisting of S and T; and
R2 is selected from the group consisting of M and L.
- 15
15. The polypeptide of claim 14, wherein the polypeptide is selected from the
group consisting of FSTM (SEQ ID NO:21), FTTM (SEQ ID NO:22) and FTSL
(SEQ ID NO:23).
16. A polypeptide selected from the group consisting of
- (a) X1-FSTM-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence SCLRK (SEQ ID NO: 24), and Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the
20 sequence PFLFC (SEQ ID NO: 25);
 - (b) X3-FTTM-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence SCLQR (SEQ ID NO: 27), and Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence PFLFC (SEQ ID NO: 25);
 - (c) X5-FSTM-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the
25 sequence SCLRR (SEQ ID NO: 29), and Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence PFMFC (SEQ ID NO: 30);
 - (d) X2-FSTM-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence SCLAR (SEQ ID NO: 32), and Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence PFLYC (SEQ ID NO: 33);
 - (e) X4-FSTL-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the
30 sequence SCLPV (SEQ ID NO: 35), and Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence PFAYC (SEQ ID NO: 36); and

WO 03/012122

PCT/US02/23763

- (f) X6-FSTM-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SCLPR (SEQ ID NO: 38), and Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PFIYC (SEQ ID NO: 39).
- 5 17. A polypeptide consisting of a sequence of general formula IV:
(R1)MF(R2)K (SEQ ID NO:41)
wherein R1 is selected from the group consisting of E, R, and D; and
R2 is selected from the group consisting of K, R, and S.
- 10 18. The polypeptide of claim 17, wherein the polypeptide is selected from the group consisting of EMFKK (SEQ ID NO:42), RMFRK (SEQ ID NO:43), and DMFSK (SEQ ID NO:44).
19. A polypeptide selected from the group consisting of
- 15 (a) X1-EMFKK-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TERS (SEQ ID NO: 48), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PTPST (SEQ ID NO: 49);
- (b) X3-RMFRK-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLNPE (SEQ ID NO: 51), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids
- 20 of the sequence PIPST (SEQ ID NO: 52);
- (c) X5-DMFSK-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TVDVS (SEQ ID NO: 54), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PQSET (SEQ ID NO: 55);
- 25 20. A polypeptide selected from the group consisting of SFQ (SEQ ID NO:45); LQF (SEQ ID NO:46), and QQF (SEQ ID NO:47).
21. A polypeptide selected from the group consisting of
- (a) X2-SFQ-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the
- 30 sequence TIPEQ (SEQ ID NO: 57), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GSPSA (SEQ ID NO: 58);
- (b) X4-LQF-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TVKAD (SEQ ID NO: 60), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SSAPA (SEQ ID NO: 61); and

WO 03/012122

PCT/US02/23763

- (c) X6-QQF-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TVEER (SEQ ID NO: 63), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GELPV (SEQ ID NO: 64).
- 5 22. A polypeptide consisting of a sequence of general formula V:
(R1)AH(R2)QD (SEQ ID NO:66)
wherein R1 is selected from the group consisting of R and K; and
R2 is selected from the group consisting of G and N.
23. The polypeptide of claim 22 wherein the polypeptide consists of a sequence
10 selected from the group consisting of RAHGQD (SEQ ID NO:67) and KAHNQD
(SEQ ID NO:68)..
24. A polypeptide selected from the group consisting of
- (a) X1-RAHGQD-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the
15 sequence VQGNE (SEQ ID NO: 69), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids
of the sequence LGTAG (SEQ ID NO: 70);
- (b) X3-RAHGQD-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence VQGNQ (SEQ ID NO: 72), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino
acids of the sequence LGTLG (SEQ ID NO: 73);
- 20 (c) X5-RAHGQD-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence VQGNK (SEQ ID NO: 75), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino
acids of the sequence LGTAG (SEQ ID NO: 70);
- (d) X2-KAHNQD-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the
25 sequence FEGQE (SEQ ID NO: 77), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids
of the sequence LGLAG (SEQ ID NO: 78);
- (e) X4-KAHNQD-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence LEGQE (SEQ ID NO: 80), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids
of the sequence LGLAG (SEQ ID NO: 78); and
- (f) X6-KAHNQD-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the
30 sequence VEGQE (SEQ ID NO: 82), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids
of the sequence LGFAG (SEQ ID NO: 83).
25. A polypeptide consisting of a sequence of general formula VI:
(R1)G(R2)GQ (SEQ ID NO:85)

WO 03/012122

PCT/US02/23763

- wherein R1 is selected from the group consisting of E and Q; and
R2 is selected from the group consisting of S, T, and G.
26. The polypeptide of claim 25, wherein the polypeptide is selected from the group consisting of EGSGQ (SEQ ID NO:86), EGTGQ (SEQ ID NO:87), EGGGQ (SEQ ID NO:88) and QGGGQ (SEQ ID NO:89)
27. A polypeptide selected from the group consisting of
- (a) X1-EGSGQ-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TSAGA (SEQ ID NO: 90), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ALASP (SEQ ID NO: 91);
- (b) X3-EGTGQ-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TSAGS (SEQ ID NO: 93), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ALASP (SEQ ID NO: 91);
- (c) X2-EGGGQ-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAAGD (SEQ ID NO: 95), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLVSP (SEQ ID NO: 96);
- (d) X4-QGGGQ-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TGAGD (SEQ ID NO: 98), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ALMSP (SEQ ID NO: 99); and
- (e) X6-EGGGQ-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAAGA (SEQ ID NO: 101), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLVSP (SEQ ID NO: 96).
28. A polypeptide consisting of a sequence of general formula VII:
(R1)G(R2)(R3) (SEQ ID NO:103)
wherein R1 is selected from the group consisting of Q and E;
R2 is selected from the group consisting of N and Q; and
R3 is selected from the group consisting of E, Q, and K.
29. The polypeptide of claim 28, wherein R1 is Q and R2 is N.
30. The polypeptide of claim 28, wherein the polypeptide is selected from the group consisting of QGNE (SEQ ID NO:104), QGNQ (SEQ ID NO:105), QGNK (SEQ ID NO:106), and EGQE (SEQ ID NO:107)

WO 03/012122

PCT/US02/23763

31. A polypeptide selected from the group consisting of
- (a) X1-QGNE-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYV (SEQ ID NO: 108), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RAHGQ (SEQ ID NO: 109);
- 5 (b) X3-QGNQ-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLFV (SEQ ID NO: 111), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RAHGQ (SEQ ID NO: 109);
- (c) X5-QGNK-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYV (SEQ ID NO:108), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RAHGQ (SEQ ID NO: 109);
- 10 (d) X2-EGQE-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYF (SEQ ID NO:114), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KAHNQ (SEQ ID NO:115);
- (e) X4-EGQE-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLYL (SEQ ID NO:117), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KAHNQ (SEQ ID NO:115); and
- 15 (f) X6-EGQE-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SLLFV (SEQ ID NO:119), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KAHNQ (SEQ ID NO:115).
- 20
32. The polypeptide of claim 31, wherein the polypeptide is selected from the group consisting of VQGNER (SEQ ID NO: 121), VQGNQR (SEQ ID NO: 122), VQGNKR (SEQ ID NO: 123), FEGQEK (SEQ ID NO: 124), LEGQEK (SEQ ID NO: 125), and VEGQEK (SEQ ID NO: 126).
- 25
33. A polypeptide consisting of a sequence of general formula VIII:
M(R1)M(R2)P (SEQ ID NO:127)
wherein R1 is selected from the group consisting of S, N, or is absent; and
R2 is selected from the group consisting of A, Q, or is absent.
- 30 34. The polypeptide of claim 33 wherein the polypeptide is selected from the group consisting of MSMAP (SEQ ID NO:128), MNMAP (SEQ ID NO:129), MSMQP (SEQ ID NO:130), and MMP (SEQ ID NO: 131).
35. A polypeptide selected from the group consisting of

WO 03/012122

PCT/US02/23763

- (a) X1-MSMAP-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PEPMP (SEQ ID NO: 132), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ITGEN (SEQ ID NO: 133);
- (b) X3-MNMAP-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PALMP (SEQ ID NO: 135), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ITGRA (SEQ ID NO: 136);
- (c) X5-MSMQP-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PEPMP (SEQ ID NO:132), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LKGQS (SEQ ID NO: 138);
- (d) X2-MMP-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAPLP (SEQ ID NO:140), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VAEDE (SEQ ID NO:141);
- (e) X4-MMP-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AAPLP (SEQ ID NO:143), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LSEEA (SEQ ID NO:144); and
- (f) X6-MMP-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence TAPIP (SEQ ID NO:146), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VSQTQ (SEQ ID NO:147).
36. A polypeptide selected from the group consisting of PMPMSMAPITG (SEQ ID NO: 149); LMPMNMAPITG (SEQ ID NO:150); PMPMSMQPLKG (SEQ ID NO: 151); PLPMMPVAE (SEQ ID NO: 152); PLPMMPLSE (SEQ ID NO: 153); and PIPMMPVSQ (SEQ ID NO: 154).
37. A polypeptide consisting of a sequence of general formula IX:
 AG(R1)(R2) (SEQ ID NO:155)
 wherein R1 is selected from the group consisting of A, S and D; and
 R2 is selected from the group consisting of E and Q.
38. The polypeptide of claim 37 wherein the polypeptide is selected from the group consisting of AGAE (SEQ ID NO:156), AGSE (SEQ ID NO:157), AGDE (SEQ ID NO:158), and AGDQ (SEQ ID NO:159).
39. A polypeptide selected from the group consisting of

WO 03/012122

PCT/US02/23763

- (a) X1-AGAE-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence VMHTS (SEQ ID NO: 160), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GSGQA (SEQ ID NO: 161);
- (b) X3-AGSE-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence IMFTS (SEQ ID NO: 163), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GTGQA (SEQ ID NO: 164);
- (c) X5-AGAE-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence MMHTS (SEQ ID NO:166), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GSGQA (SEQ ID NO: 161);
- (d) X2-AGDE-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LMHTA (SEQ ID NO:168), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GGGQS (SEQ ID NO:169);
- (e) X4-AGDQ-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LMHTG (SEQ ID NO:171), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GGGQA (SEQ ID NO:172); and
- (f) X6-AGAE-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence LMHTA (SEQ ID NO:168), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GGGQS (SEQ ID NO:169).
40. A polypeptide consisting of at least 5 contiguous amino acids of general formula X:
 EC(R1)G(R2)(R3)GTC(R4)(R5)(R6) (SEQ ID NO:175)
 wherein R1 is selected from the group consisting of H, N, Q, and S;
 R2 is selected from the group consisting of G, R, A, or is absent;
 R3 is selected from the group consisting of R and Q
 R4 is selected from the group consisting of N and H;
 R5 is selected from the group consisting of F and Y; and
 R6 is selected from the group consisting of F and Y.
41. The polypeptide of claim 40, wherein the polypeptide consists of the amino acid sequence of general formula X.
42. The polypeptide of claim 40, wherein
 R2 is selected from the group consisting of G, R, A; and
 R4 is H.

WO 03/012122

PCT/US02/23763

43. The polypeptide of claims 40 wherein the polypeptide is selected from the group consisting of ECHGRGTCNYY (SEQ ID NO:176), ECNGGRGTCHYY (SEQ ID NO:177), ECQGRQGTCHFF (SEQ ID NO:178), and ECSGARGTCHYF (SEQ ID NO:179).
- 5
44. A polypeptide consisting of an amino acid sequence of general formula XI:
(R1)(R2)T(R3)K (SEQ ID NO:180)
wherein R1 is selected from the group consisting of P, S, and A;
R2 is selected from the group consisting of S, E, and D; and
10 R3 is selected from the group consisting of L and V.
45. The polypeptide of claim 44 wherein R3 is L.
46. The polypeptide of claim 44 wherein R2 is selected from D and E.
47. The polypeptide of claim 44 wherein the polypeptide is selected from the
15 group consisting of PSTLK (SEQ ID NO:181), PSTVK (SEQ ID NO:182), SETLK
(SEQ ID NO:183), ADTLK (SEQ ID NO:184), and PDTLK (SEQ ID NO:185).
48. A polypeptide selected from the group consisting of
- (a) X1-PSTLK-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the
20 sequence FKKPT (SEQ ID NO: 186), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino
acids of the sequence AGE LR (SEQ ID NO: 187);
- (b) X3-PSTVK-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence FRKPI (SEQ ID NO: 189), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids
of the sequence AGELE (SEQ ID NO: 190);
- 25 (c) X5-SETLK-Z5, wherein X5 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence FSKPQ (SEQ ID NO:192), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids
of the sequence AGDLR (SEQ ID NO: 193);
- (d) X2-ADTLK-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence QGSPS (SEQ ID NO:195), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids
30 of the sequence AGLIR (SEQ ID NO:196);
- (e) X4-PDTLK-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the
sequence SSAPA (SEQ ID NO:198), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids
of the sequence ESQAQ (SEQ ID NO:199); and

WO 03/012122

PCT/US02/23763

(f) X6-SETLK-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence GELPV (SEQ ID NO:201), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence AGQLH (SEQ ID NO:202).

- 5 49. A polypeptide consisting of an amino acid sequence of general formula XII:
A(R1)RND (SEQ ID NO:204)
wherein R1 is selected from the group consisting of S, Q, and R.
50. The polypeptide of claim 49, wherein the polypeptide sequence is selected from the group consisting of ASRND (SEQ ID NO:205), AQRND (SEQ ID NO:206), and ARRND (SEQ ID NO:207).
- 10 51. A polypeptide selected from the group consisting of
- (a) X1-ASRND-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence NVCNF (SEQ ID NO: 208), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSYWL (SEQ ID NO: 209);
- 15 (b) X3-ASRND-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence DVCNF (SEQ ID NO: 211), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSYWL (SEQ ID NO: 209);
- (c) X2-ASRND-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence DVCYY (SEQ ID NO:213), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KSYWL (SEQ ID NO:214);
- 20 (d) X4-AQRND-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence QVCHY (SEQ ID NO:216), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RSYWL (SEQ ID NO:217); and
- (e) X6-ARRND-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence EVCHY (SEQ ID NO:219), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence KSYWL (SEQ ID NO:214).
- 25 52. A polypeptide consisting of an amino acid sequence of general formula XIII:
(R1)(R2)(R3)N(R4) (SEQ ID NO:221)
wherein R1 is selected from the group consisting of Y and F;
R2 is selected from the group consisting of Y and F;
R3 is selected from the group consisting of A and S; and
R4 is selected from the group consisting of A, S, and K.
- 30

WO 03/012122

PCT/US02/23763

53. The polypeptide of claim 49, wherein the polypeptide sequence is selected from the group consisting of YYANA (SEQ ID NO:222) YYSNS (SEQ ID NO:223) YYANS (SEQ ID NO:224) YYANK (SEQ ID NO:225) FFANK (SEQ ID NO:226) and YFANK (SEQ ID NO:227).
- 5
54. A polypeptide selected from the group consisting of
- (a) X1-YYANA-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO: 228), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO: 229);
- 10 (b) X3-YYSNS-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO: 228), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO: 229);
- (c) X1-YYANS-Z2, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCN (SEQ ID NO: 228), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO: 229);
- 15 (d) X2-YYANK-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCH (SEQ ID NO:233), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO:229);
- (e) X4-FFANK-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence QGTCH (SEQ ID NO:235), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO:229); and
- 20 (f) X6-YFANK-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence RGTCH (SEQ ID NO:233), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence YSFWL (SEQ ID NO:229).
- 25
55. A polypeptide selected from the group consisting of IERSEMFKKPT (SEQ ID NO:238), LNPERMFRKPI (SEQ ID NO:239), VDVSDMFSKPQ (SEQ ID NO:240), IPEQSFQGPS (SEQ ID NO:241), VKADLQFSSAPA (SEQ ID NO:242), and VEERQQFGELPV (SEQ ID NO:243).
- 30
56. A polypeptide selected from the group consisting of
- (a) X1-IERSEMFKKPT-Z1, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FWLAT (SEQ ID NO: 244), and wherein Z1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PSTLK (SEQ ID NO: 181);

WO 03/012122

PCT/US02/23763

- (b) X3-LNPERMFRKPI-Z3, wherein X3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FWLAS (SEQ ID NO: 246), and wherein Z3 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PSTVK (SEQ ID NO: 182);
- (c) X1-VDVSDMFSKPQ-Z2, wherein X1 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FWLAT (SEQ ID NO: 244), and wherein Z5 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SETLK (SEQ ID NO: 183);
- (d) X2-IPEQSFQGPS-Z2, wherein X2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FWLTT (SEQ ID NO: 249), and wherein Z2 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence ADTLK (SEQ ID NO: 184);
- (e) X4-VKADLQFSSAPA-Z4, wherein X4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FWLTT (SEQ ID NO: 249), and wherein Z4 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence PDTLK (SEQ ID NO: 185); and
- (f) X6-VEERQQFGELPV-Z6, wherein X6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence FWLTT (SEQ ID NO: 249), and wherein Z6 is 0, 1, 2, 3, 4, or 5 amino acids of the sequence SETLK (SEQ ID NO: 183).

57. A polypeptide selected from the group consisting of
 FSTMPFLFCNINNVCFNA (SEQ ID NO: 253), FTTMPFLFCNVNDVCFNA (SEQ ID NO: 254), FSTMPFMFCNINNVCFNA (SEQ ID NO: 255),
 20 FSTMPFLFCNPGDVCYYA (SEQ ID NO: 256), FSTLPFAYCNIHQVCHYA (SEQ ID NO: 257), FSTMPFIYCNINEVCHYA (SEQ ID NO: 258),
 PFLFCNINNVCFNFRND (SEQ ID NO: 259), PFLFCNVNDVCFNFRND (SEQ ID NO: 260), PFMFCNINNVCFNFRND (SEQ ID NO: 261),
 PFLFCNPGDVCYYASRND (SEQ ID NO: 262), PFAYCNIHQVCHYAQRND (SEQ ID NO: 263),
 25 PFYCNINEVCHYARRND (SEQ ID NO: 264), FSTMPFLFCNINNVCFNFRND (SEQ ID NO: 265),
 FTTMPFLFCNVNDVCFNFRND (SEQ ID NO: 266), FSTMPFMFCNINNVCFNFRND (SEQ ID NO: 267),
 FSTMPFLFCNPGDVCYYASRND (SEQ ID NO: 268),
 30 FSTLPFAYCNIHQVCHYAQRND (SEQ ID NO: 269), FSTMPFIYCNINEVCHYARRND (SEQ ID NO: 270), PFIECHGRGTCNYY (SEQ ID NO: 271),
 PFLECHGRGTCNYY (SEQ ID NO: 272), PFIECNGGRGTCNYY (SEQ ID NO: 273), PFLECQGRQGTCHFF (SEQ ID NO: 274),
 PFIECSGARGTCHYF (SEQ ID NO: 275), IERSEMFKKTPSTLKAG (SEQ ID NO: 276)

WO 03/012122

PCT/US02/23763

NO: 276), LNPERMFRKPIPVSTVKAG (SEQ ID NO:277),
 VDVSDFMSKPSQSETLKAG (SEQ ID NO: 278), IPEQSFQGPSADTLKAG (SEQ
 ID NO: 279), VKADLQFSSAPADTLKES (SEQ ID NO: 280),
 VEERQQFGBELPVSETLKAG (SEQ ID NO: 281), GSCLRFSTM (SEQ ID NO:
 5 282), GSCLQRFTTM (SEQ ID NO:283), GSCLRRFSTM (SEQ ID NO: 284),
 GSCLARFSTM (SEQ ID NO: 285), GSCLPVFSTL (SEQ ID NO: 286),
 GSCLPRFSTM (SEQ ID NO: 287), LRFSTMPFLFCNINNVCF (SEQ ID NO:
 288), LQRFTTMPFLFCNVNDVCF (SEQ ID NO:289),
 LRRFSTMPFMFCNINNVCF (SEQ ID NO: 290),
 10 LARFSTMPFLYCNPGDVCYY (SEQ ID NO: 291),
 LPVFSTLFFAYCNIHQVCHY (SEQ ID NO: 292), LPRFSTMPFIYCNINEVCHY
 (SEQ ID NO: 293), IPE (SEQ ID NO:294), IER (SEQ ID NO:295), QD (SEQ ID
 NO:296), NYX (SEQ ID NO:297), FCNVNDVCF (SEQ ID NO:298),
 YCNPGDVCYY (SEQ ID NO:299), YCNIHQVCHY (SEQ ID NO:300), and
 15 YCNINEVCHY (SEQ ID NO:301)

58. A pharmaceutical composition comprising:
 (a) the polypeptide of any one of claims 1-57; and
 (b) a pharmaceutically acceptable carrier.

20

59. A method for inhibiting angiogenesis in tissue comprising contacting said
 tissue with an effective inhibiting amount of the polypeptide of any one of claims 1-
 57.

25

60. A method for inhibiting angiogenesis in tissue comprising contacting said
 tissue with an effective inhibiting amount of the pharmaceutical composition of claim
 58.

- 30 61. The method of claim 59 or 60 wherein the angiogenesis is tumor-induced.

62. A method for treating an angiogenesis-mediated disease or condition in a
 mammal, comprising administering to a mammal with an angiogenesis-mediated

WO 03/012122

PCT/US02/23763

disease or condition an amount effective to inhibit angiogenesis of the polypeptide of any one of claims 1-57.

63. A method for treating an angiogenesis-mediated disease or condition in a mammal, comprising administering to a mammal with an angiogenesis-mediated disease or condition an amount effective to inhibit angiogenesis of the pharmaceutical composition of claim 58

64. The method of claim 62 or 63 wherein the angiogenesis-mediated disease or condition is selected from the group consisting of solid and blood-borne tumors, diabetic retinopathy, rheumatoid arthritis, retinal neovascularization, choroidal neovascularization, macular degeneration, corneal neovascularization, retinopathy of prematurity, corneal graft rejection, neovascular glaucoma, retrolental fibroplasia, epidemic keratoconjunctivitis, Vitamin A deficiency, contact lens overwear, atopic keratitis, superior limbic keratitis, pterygium keratitis sicca, sogrens, acne rosacea, phlyctenulosis, syphilis, Mycobacteria infections, lipid degeneration, chemical burns, bacterial ulcers, fungal ulcers, Herpes simplex infections, Herpes zoster infections, protozoan infections, Kaposi's sarcoma, Mooren ulcer, Terrien's marginal degeneration, marginal keratolysis, trauma, systemic lupus, polyarteritis, Wegeners sarcoidosis, scleritis, Steven's Johnson disease, radial keratotomy, sickle cell anemia, sarcoid, pseudoxanthoma elasticum, Pagets disease, vein occlusion, artery occlusion, carotid obstructive disease, chronic uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets disease, myopia, optic pits, Stargarts disease, pars planitis, chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis, post-laser complications, abnormal proliferation of fibrovascular tissue, hemangiomas, Osler-Weber-Rendu, acquired immune deficiency syndrome, ocular neovascular disease, osteoarthritis, chronic inflammation, Crohn's disease, ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid.

65. A method for inhibiting tumor metastasis, comprising contacting a tumor or tissue with an amount effective to inhibit tumor metastasis of the polypeptide of any one of claims 1-57.

WO 03/012122

PCT/US02/23763

66. A method for inhibiting tumor metastasis, comprising contacting a tumor or tissue with an amount effective to inhibit tumor metastasis of the pharmaceutical composition of claim 58.
67. A method for inhibiting tumor growth, comprising contacting a tumor or tissue with an amount effective to inhibit tumor growth of the polypeptide of any one of claims 1-57.
68. A method for inhibiting tumor growth, comprising contacting a tumor or tissue with an amount effective to inhibit tumor growth of the pharmaceutical composition of claim 58.
69. A method for inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue comprising contacting the tumor or animal tissue with an amount effective to inhibit endothelial cell interaction with the extracellular matrix of the polypeptide of any one of claims 1-57.
70. A method for inhibiting endothelial cell interaction with the extracellular matrix in an animal tissue comprising contacting the tumor or animal tissue with an amount effective to inhibit endothelial cell interaction with the extracellular matrix of the pharmaceutical composition of claim 58.
71. A method for inhibiting basal lamina membrane formation in cell or tissue development comprising contacting the cell or tissue with an amount effective to inhibit basal lamina membrane formation of the polypeptide of any one of claims 1-57.
72. A method for inhibiting basal lamina membrane formation in cell or tissue development comprising contacting the cell or tissue with an amount effective to inhibit basal lamina membrane formation of the pharmaceutical composition of claim 58.

WO 03/012122

PCT/US02/23763

73. A crystal of an NC1 domain hexamer of type IV collagen, wherein the crystal comprises an $[(\alpha 1)_2 \alpha 2]_2$ NC1 hexamer of type IV collagen, wherein the crystal consists of space groups $P2_1$ with approximate $a =$ between 127.16 Å and 129.41 Å,
5 $b =$ between 139.57 Å and 143.87 Å; $c =$ between 160.20 Å and 162.92 Å; $\beta = 91.3^\circ$, such that the three-dimensional structure of the crystallized NC1 domain hexamer can be determined to a resolution of 3 Å or better.

74. The crystal of claim 73, wherein the three-dimensional structure of the
10 crystallized NC1 domain hexamer can be determined to a resolution of 2.2 Å or better.

75. The crystal of claim 73, wherein the three-dimensional structure of the
15 crystallized NC1 domain hexamer can be determined to a resolution of 2 Å or better.

76. A method for identifying inhibitors of type IV collagen assembly, comprising:
(a) obtaining crystals of an NC1 hexamer of type IV collagen, wherein the
crystal comprises an $[(\alpha 1)_2 \alpha 2]_2$ NC1 hexamer of type IV collagen, wherein the
20 crystal consists of space groups $P2_1$ with approximate $a =$ between 127.16 Å and 129.41 Å, $b =$ between 139.57 Å and 143.87 Å; $c =$ between 160.20 Å and 162.92 Å; $\beta = 91.3^\circ$, such that the three-dimensional structure of the crystallized NC1 domain hexamer can be determined to a resolution of 3 Å or better.

(b) analyzing the three-dimensional structure of the crystallized NC1
25 domain hexamer of type IV collagen of claim 75; and

(b) designing a potential inhibitor of type IV collagen assembly that
targets one or more regions of a type IV collagen NC1 α chain selected from the
group consisting of:

- 30 (i) Intor-chain domain swapping region;
- (ii) Intra-chain domain swapping region;
- (iii) Specificity region;
- (iv) Specificity region partner;
- (v) Hexamer interface;
- (vi) Monomer-monomer interface; and

WO 03/012122

PCT/US02/23763

(vii) Hypervariable region.

77. The method of claim 76, further comprising:
5 (a) synthesizing the potential inhibitor; and
(b) determining whether the potential inhibitor inhibits the assembly of type IV collagen.
- 78 The method of claim 76, further comprising:
10 (a) synthesizing the potential inhibitor; and
(b) conducting an assay to determine whether the potential inhibitor inhibits one or more of angiogenesis, tumor growth, tumor metastasis, endothelial cell adhesion, endothelial cell proliferation, and basal lamina assembly.
- 15 79. An inhibitor of type IV collagen assembly identified by the method of any one of claims 76-78.
80. An inhibitor of one or more process selected from the group consisting of angiogenesis, tumor growth, tumor metastasis, endothelial cell adhesion, endothelial
20 cell proliferation, and basal lamina assembly, identified by the method of any one of claims 76-78.

WO 03/012122

PCT/US02/23763

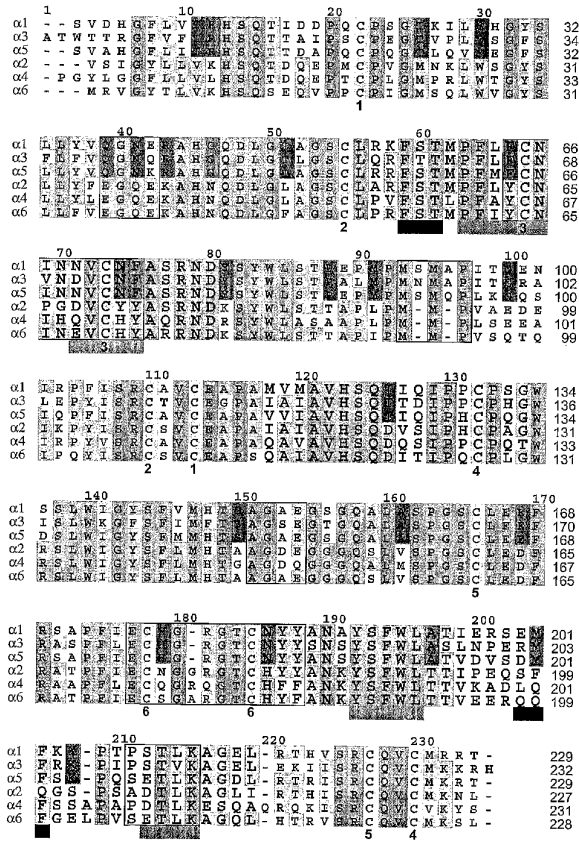


FIGURE 1

WO 03/012122

PCT/US02/23763

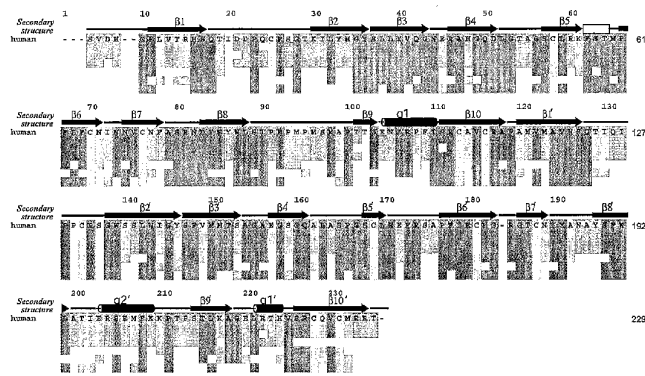


FIGURE 2a

WO 03/012122

PCT/US02/23763

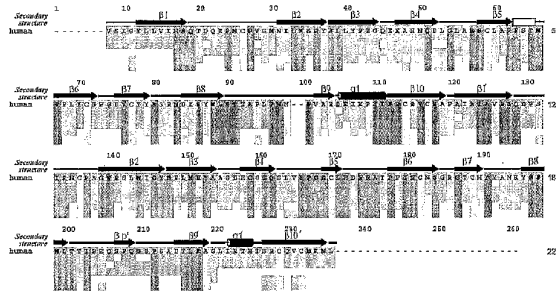


FIGURE 2b

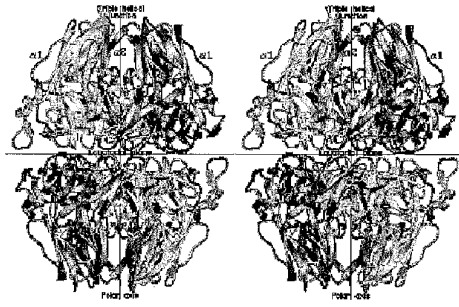


FIGURE 3

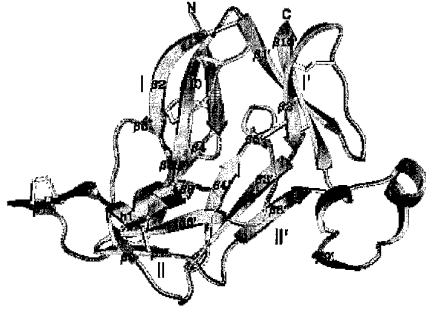


FIGURE 4

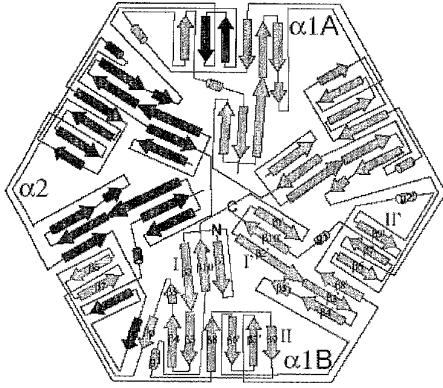


FIGURE 5

WO 03/012122

PCT/US02/23763

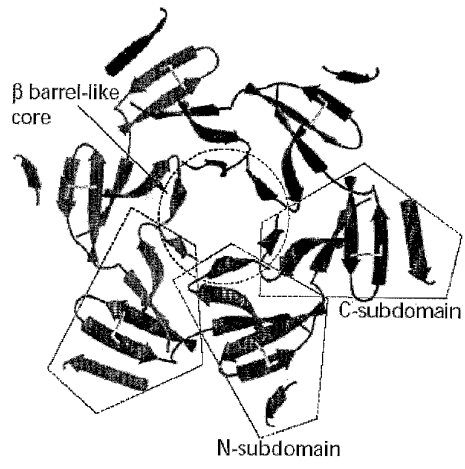


FIGURE 6a

WO 03/012122

PCT/US02/23763

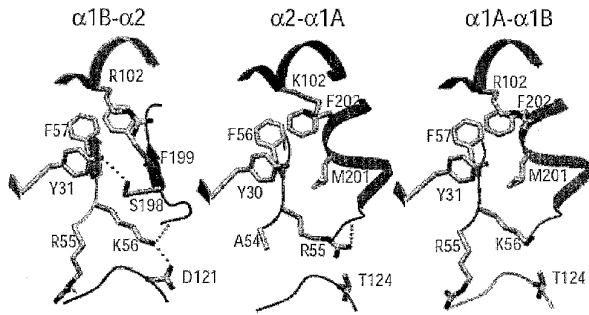


FIGURE 6b

WO 03/012122

PCT/US02/23763

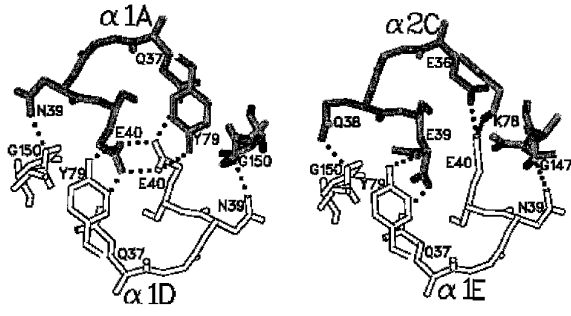


FIGURE 7a

WO 03/012122

PCT/US02/23763

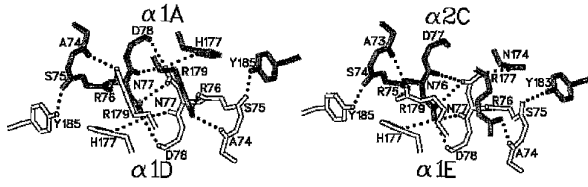


FIGURE 7b

WO 03/012122

PCT/US02/23763

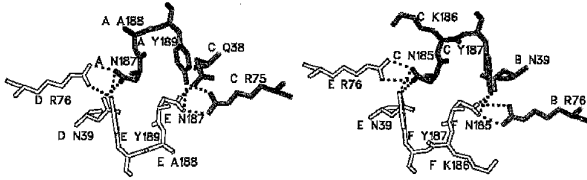


FIGURE 7c

WO 03/012122

1/75

PCT/US02/23763

SEQUENCE LISTING

<110> Sundaramoorthy, M.
Hudson, B.

<120> Crystallized structure of Type IV Collagen NCI Domain Hexamer

<130> MBHB 01-1017-PCT

<150> US 60/308,523
<151> 2001-07-27

<150> US 60/351,289
<151> 2001-10-29

<150> US 60/366,854
<151> 2002-03-22

<150> US 60/385,362
<151> 2002-06-03

<160> 307

<170> PatentIn version 3.1

<210> 1
<211> 14
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X stands for leucine, methionine, alanine, valine, norleucine, or
isoleucine.

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X stands for phenylalanine or tyrosine.

<220>
<221> MISC_FEATURE
<222> (7)..(7)
<223> X stands for isoleucine, valine, leucine, norleucine, alanine, or
proline.

<220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> X stands for asparagine, glycine, or histidine.

<220>

WO 03/012122

2/75

PCT/US02/23763

<221> MISC_FEATURE
<222> (9)..(9)
<223> X stands for glutamine, aspartate, asparagine, and glutamate.

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> X stands for asparagine, tyrosine, or histidine.

<220>
<221> MISC_FEATURE
<222> (13)..(13)
<223> X stands for phenylalanine or tyrosine.

<400> 1
Pro Phe Xaa Xaa Cys Asn Xaa Xaa Xaa Val Cys Xaa Xaa Ala
1 5 10

<210> 2
<211> 14
<212> PRT
<213> Homo sapiens

<400> 2
Pro Phe Leu Phe Cys Asn Ile Asn Asn Val Cys Asn Phe Ala
1 5 10

<210> 3
<211> 14
<212> PRT
<213> Homo sapiens

<400> 3
Pro Phe Leu Phe Cys Asn Val Asn Asp Val Cys Asn Phe Ala
1 5 10

<210> 4
<211> 14
<212> PRT
<213> Homo sapiens

<400> 4
Pro Phe Met Phe Cys Asn Ile Asn Asn Val Cys Asn Phe Ala
1 5 10

<210> 5

WO 03/012122

3/75

PCT/US02/23763

<211> 14
<212> PRT
<213> Homo sapiens

<400> 5

Pro Phe Leu Tyr Cys Asn Pro Gly Asp Val Cys Tyr Tyr Ala
1 5 10

<210> 6
<211> 14
<212> PRT
<213> Homo sapiens

<400> 6

Pro Phe Ala Tyr Cys Asn Ile His Gln Val Cys His Tyr Ala
1 5 10

<210> 7
<211> 14
<212> PRT
<213> Homo sapiens

<400> 7

Pro Phe Ile Tyr Cys Asn Ile Asn Glu Val Cys His Tyr Ala
1 5 10

<210> 8
<211> 13
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X stands for leucine, alanine, valine, norleucine, or isoleucine.

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> X stands for histidine, asparagine, glutamine, or serine.

<220>
<221> MISC_FEATURE
<222> (8)..(8)
<223> X stands for glycine, arginine, alanine, or is absent.

<220>
<221> MISC_FEATURE

WO 03/012122

4/75

PCT/US02/23763

<222> (9)..(9)

<223> X stands for arginine or glutamine.

<220>

<221> MISC_FEATURE

<222> (13)..(13)

<223> X stands for asparagine or histidine.

<400> 8

Pro Phe Xaa Glu Cys Xaa Gly Xaa Xaa Gly Thr Cys Xaa
1 5 10

<210> 9

<211> 12

<212> PRT

<213> Homo sapiens

<400> 9

Pro Phe Ile Glu Cys His Gly Arg Gly Thr Cys Asn
1 5 10

<210> 10

<211> 12

<212> PRT

<213> Homo sapiens

<400> 10

Pro Phe Leu Glu Cys His Gly Arg Gly Thr Cys Asn
1 5 10

<210> 11

<211> 13

<212> PRT

<213> Homo sapiens

<400> 11

Pro Phe Ile Glu Cys Asn Gly Gly Arg Gly Thr Cys His
1 5 10

<210> 12

<211> 13

<212> PRT

<213> Homo sapiens

<400> 12

Pro Phe Leu Glu Cys Gln Gly Arg Gln Gly Thr Cys His

WO 03/012122

5/75

PCT/US02/23763

1 5 10

<210> 13
<211> 13
<212> PRT
<213> Homo sapiens

<400> 13

Pro Phe Ile Glu Cys Ser Gly Ala Arg Gly Thr Cys His
1 5 10

<210> 14
<211> 22
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Amino acids at positions 1-5 are optionally absent, such that if
5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.

<220>
<221> MISC_FEATURE
<222> (18)..(22)
<223> Amino acids at positions 18-22 are optionally absent, such that if
18 is absent, 19-22 are absent, if 19 is absent, 20-22 are absent, etc.

<400> 14

Glu Phe Arg Ser Ala Pro Phe Ile Glu Cys His Gly Arg Gly Thr Cys
1 5 10 15

Asn Tyr Tyr Ala Asn Ala
20

<210> 15
<211> 22
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Amino acids at positions 1-5 are optionally absent, such that if
5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.

<220>
<221> MISC_FEATURE

WO 03/012122

6/75

PCT/US02/23763

<222> (18)..(22)
<223> Amino acids at positions 18-22 are optionally absent, such that if
18 is absent, 19-22 are absent, if 19 is absent, 20-22 are absent, etc.

<400> 15

Glu Phe Arg Ala Ser Pro Phe Leu Glu Cys His Gly Arg Gly Thr Cys
1 5 10 15

Asn Tyr Tyr Ser Asn Ser
20

<210> 16

<211> 22

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (1)..(5)

<223> Amino acids at positions 1-5 are optionally absent, such that if
5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.

<220>

<221> MISC_FEATURE

<222> (18)..(22)

<223> Amino acids at positions 18-22 are optionally absent, such that if
18 is absent, 19-22 are absent, if 19 is absent, 20-22 are absent, etc.

<400> 16

Glu Phe Arg Ser Ala Pro Phe Ile Glu Cys His Gly Arg Gly Thr Cys
1 5 10 15

Asn Tyr Tyr Ala Asn Ser
20

<210> 17

<211> 23

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (1)..(5)

<223> Amino acids at positions 1-5 are optionally absent, such that if
5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.

<220>

WO 03/012122

7/75

PCT/US02/23763

<221> MISC_FEATURE
<222> (19)..(23)
<223> Amino acids at positions 19-23 are optionally absent, such that if
19 is absent, 20-23 are absent, if 20 is absent, 21-23 are absent, etc.

<400> 17

Asp Phe Arg Ala Thr Pro Phe Ile Glu Cys Asn Gly Gly Arg Gly Thr
1 5 10 15

Cys His Tyr Tyr Ala Asn Lys
20

<210> 18
<211> 23
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Amino acids at positions 1-5 are optionally absent, such that if
5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.

<220>
<221> MISC_FEATURE
<222> (19)..(23)
<223> Amino acids at positions 19-23 are optionally absent, such that if
19 is absent, 20-23 are absent, if 20 is absent, 21-23 are absent, etc.

<400> 18

Asp Phe Arg Ala Ala Pro Phe Leu Glu Cys Gln Gly Arg Gln Gly Thr
1 5 10 15

Cys His Phe Phe Ala Asn Lys
20

<210> 19
<211> 23
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(5)
<223> Amino acids at positions 1-5 are optionally absent, such that if
5 is absent, 1-4 are absent, if 4 is absent, 1-3 are absent, etc.

WO 03/012122

8/75

PCT/US02/23763

<220>
<221> MISC_FEATURE
<222> (19)..(23)
<223> Amino acids at positions 19-23 are optionally absent, such that if
19 is absent, 20-23 are absent, if 20 is absent, 21-23 are absent, etc.

<400> 19

Asp Phe Arg Ala Thr Pro Phe Ile Glu Cys Ser Gly Ala Arg Gly Thr
1 5 10 15

Cys His Tyr Phe Ala Asn Lys
20

<210> 20
<211> 4
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> X stands for serine or threonine.

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X stands for methionine and leucine.

<400> 20

Phe Xaa Thr Xaa
1

<210> 21
<211> 4
<212> PRT
<213> Homo sapiens

<400> 21

Phe Ser Thr Met
1

<210> 22
<211> 4
<212> PRT
<213> Homo sapiens

<400> 22

WO 03/012122

9/75

PCT/US02/23763

Phe Thr Thr Met
1

<210> 23
<211> 4
<212> PRT
<213> Homo sapiens

<400> 23

Phe Thr Ser Leu
1

<210> 24
<211> 5
<212> PRT
<213> Homo sapiens

<400> 24

Ser Cys Leu Arg Lys
1 5

<210> 25
<211> 5
<212> PRT
<213> Homo sapiens

<400> 25

Pro Phe Leu Phe Cys
1 5

<210> 26
<211> 14
<212> PRT
<213> Homo sapiens

<400> 26

Ser Cys Leu Arg Lys Phe Ser Thr Met Pro Phe Leu Phe Cys
1 5 10

<210> 27
<211> 5
<212> PRT
<213> Homo sapiens

<400> 27

Ser Cys Leu Gln Arg

WO 03/012122

10/75

PCT/US02/23763

1 5

<210> 28
<211> 14
<212> PRT
<213> Homo sapiens

<400> 28
Ser Cys Leu Gln Arg Phe Thr Thr Met Pro Phe Leu Phe Cys
1 5 10

<210> 29
<211> 5
<212> PRT
<213> Homo sapiens

<400> 29
Ser Cys Leu Arg Arg
1 5

<210> 30
<211> 5
<212> PRT
<213> Homo sapiens

<400> 30
Pro Phe Met Phe Cys
1 5

<210> 31
<211> 14
<212> PRT
<213> Homo sapiens

<400> 31
Ser Cys Leu Arg Arg Phe Ser Thr Met Pro Phe Met Phe Cys
1 5 10

<210> 32
<211> 5
<212> PRT
<213> Homo sapiens

<400> 32
Ser Cys Leu Ala Arg
1 5

WO 03/012122

11/75

PCT/US02/23763

<210> 33
<211> 5
<212> PRT
<213> Homo sapiens

<400> 33

Pro Phe Leu Tyr Cys
1 5

<210> 34
<211> 14
<212> PRT
<213> Homo sapiens

<400> 34

Ser Cys Leu Ala Arg Phe Ser Thr Met Pro Phe Leu Tyr Cys
1 5 10

<210> 35
<211> 5
<212> PRT
<213> Homo sapiens

<400> 35

Ser Cys Leu Pro Val
1 5

<210> 36
<211> 5
<212> PRT
<213> Homo sapiens

<400> 36

Pro Phe Ala Tyr Cys
1 5

<210> 37
<211> 14
<212> PRT
<213> Homo sapiens

<400> 37

Ser Cys Leu Pro Val Phe Ser Thr Leu Pro Phe Ala Tyr Cys
1 5 10

<210> 38

WO 03/012122

12/75

PCT/US02/23763

<211> 5
<212> PRT
<213> Homo sapiens

<400> 38

Ser Cys Leu Pro Arg
1 5

<210> 39
<211> 5
<212> PRT
<213> Homo sapiens

<400> 39

Pro Phe Ile Tyr Cys
1 5

<210> 40
<211> 14
<212> PRT
<213> Homo sapiens

<400> 40

Ser Cys Leu Pro Arg Phe Ser Thr Met Pro Phe Ile Tyr Cys
1 5 10

<210> 41
<211> 5
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> X stands for glutamate, arginine, or aspartate.

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X stands for lysine, arginine, or serine.

<400> 41

Xaa Met Phe Xaa Lys
1 5

<210> 42
<211> 5

WO 03/012122

13/75

PCT/US02/23763

<212> PRT
<213> Homo sapiens

<400> 42

Glu Met Phe Lys Lys
1 5

<210> 43
<211> 5
<212> PRT
<213> Homo sapiens

<400> 43

Arg Met Phe Arg Lys
1 5

<210> 44
<211> 5
<212> PRT
<213> Homo sapiens

<400> 44

Asp Met Phe Ser Lys
1 5

<210> 45
<211> 3
<212> PRT
<213> Homo sapiens

<400> 45

Ser Phe Gln
1

<210> 46
<211> 3
<212> PRT
<213> Homo sapiens

<400> 46

Leu Gln Phe
1

<210> 47
<211> 3
<212> PRT
<213> Homo sapiens

WO 03/012122

14/75

PCT/US02/23763

<400> 47

Gln Gln Phe
1

<210> 48

<211> 5

<212> PRT

<213> Homo sapiens

<400> 48

Thr Ile Glu Arg Ser
1 5

<210> 49

<211> 5

<212> PRT

<213> Homo sapiens

<400> 49

Pro Thr Pro Ser Thr
1 5

<210> 50

<211> 15

<212> PRT

<213> Homo sapiens

<400> 50

Thr Ile Glu Arg Ser Glu Met Phe Lys Lys Pro Thr Pro Ser Thr
1 5 10 15

<210> 51

<211> 5

<212> PRT

<213> Homo sapiens

<400> 51

Ser Leu Asn Pro Glu
1 5

<210> 52

<211> 5

<212> PRT

<213> Homo sapiens

<400> 52

WO 03/012122

15/75

PCT/US02/23763

Pro Ile Pro Ser Thr
1 5

<210> 53
<211> 15
<212> PRT
<213> Homo sapiens

<400> 53

Ser Leu Asn Pro Glu Arg Met Phe Arg Lys Pro Ile Pro Ser Thr
1 5 10 15

<210> 54
<211> 5
<212> PRT
<213> Homo sapiens

<400> 54

Thr Val Asp Val Ser
1 5

<210> 55
<211> 5
<212> PRT
<213> Homo sapiens

<400> 55

Pro Gln Ser Glu Thr
1 5

<210> 56
<211> 15
<212> PRT
<213> Homo sapiens

<400> 56

Thr Val Asp Val Ser Asp Met Phe Ser Lys Pro Gln Ser Glu Thr
1 5 10 15

<210> 57
<211> 5
<212> PRT
<213> Homo sapiens

<400> 57

Thr Ile Pro Glu Gln

WO 03/012122

16/75

PCT/US02/23763

1 5

<210> 58
<211> 5
<212> PRT
<213> Homo sapiens

<400> 58

Gly Ser Pro Ser Ala
1 5

<210> 59
<211> 13
<212> PRT
<213> Homo sapiens

<400> 59

Thr Ile Pro Glu Gln Ser Phe Gln Gly Ser Pro Ser Ala
1 5 10

<210> 60
<211> 5
<212> PRT
<213> Homo sapiens

<400> 60

Thr Val Lys Ala Asp
1 5

<210> 61
<211> 5
<212> PRT
<213> Homo sapiens

<400> 61

Ser Ser Ala Pro Ala
1 5

<210> 62
<211> 13
<212> PRT
<213> Homo sapiens

<400> 62

Thr Val Lys Ala Asp Leu Gln Phe Ser Ser Ala Pro Ala
1 5 10

WO 03/012122

17/75

PCT/US02/23763

<210> 63
<211> 5
<212> PRT
<213> Homo sapiens

<400> 63

Thr Val Glu Glu Arg
1 5

<210> 64
<211> 5
<212> PRT
<213> Homo sapiens

<400> 64

Gly Glu Leu Pro Val
1 5

<210> 65
<211> 13
<212> PRT
<213> Homo sapiens

<400> 65

Thr Val Glu Glu Arg Gln Gln Phe Gly Glu Leu Pro Val
1 5 10

<210> 66
<211> 6
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> X stands for arginine or lysine.

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X stands for glycine or asparagine.

<400> 66

Xaa Ala His Xaa Gln Asp
1 5

WO 03/012122

18/75

PCT/US02/23763

<210> 67
<211> 5
<212> PRT
<213> Homo sapiens

<400> 67

Arg Ala His Gly Gln
1 5

<210> 68
<211> 6
<212> PRT
<213> Homo sapiens

<400> 68

Lys Ala His Asn Gln Asp
1 5

<210> 69
<211> 5
<212> PRT
<213> Homo sapiens

<400> 69

Val Gln Gly Asn Glu
1 5

<210> 70
<211> 5
<212> PRT
<213> Homo sapiens

<400> 70

Leu Gly Thr Ala Gly
1 5

<210> 71
<211> 16
<212> PRT
<213> Homo sapiens

<400> 71

Val Gln Gly Asn Glu Arg Ala His Gly Gln Asp Asp Leu Gly Thr Ala
1 5 10 15

<210> 72
<211> 5

WO 03/012122

19/75

PCT/US02/23763

<212> PRT
<213> Homo sapiens

<400> 72

Val Gln Gly Asn Gln
1 5

<210> 73
<211> 5
<212> PRT
<213> Homo sapiens

<400> 73

Leu Gly Thr Leu Gly
1 5

<210> 74
<211> 16
<212> PRT
<213> Homo sapiens

<400> 74

Val Gln Gly Asn Gln Arg Ala His Gly Gln Asp Leu Gly Thr Leu Gly
1 5 10 15

<210> 75
<211> 5
<212> PRT
<213> Homo sapiens

<400> 75

Val Gln Gly Asn Lys
1 5

<210> 76
<211> 16
<212> PRT
<213> Homo sapiens

<400> 76

Val Gln Gly Asn Lys Arg Ala His Gly Gln Asp Leu Gly Thr Ala Gly
1 5 10 15

<210> 77
<211> 5
<212> PRT
<213> Homo sapiens

WO 03/012122

20/75

PCT/US02/23763

<400> 77

Phe Glu Gly Gln Glu
1 5

<210> 78

<211> 5

<212> PRT

<213> Homo sapiens

<400> 78

Leu Gly Leu Ala Gly
1 5

<210> 79

<211> 16

<212> PRT

<213> Homo sapiens

<400> 79

Phe Glu Gly Gln Glu Lys Ala His Asn Gln Asp Leu Gly Leu Ala Gly
1 5 10 15

<210> 80

<211> 5

<212> PRT

<213> Homo sapiens

<400> 80

Leu Glu Gly Gln Glu
1 5

<210> 81

<211> 16

<212> PRT

<213> Homo sapiens

<400> 81

Leu Glu Gly Gln Glu Lys Ala His Asn Gln Asp Leu Gly Leu Ala Gly
1 5 10 15

<210> 82

<211> 5

<212> PRT

<213> Homo sapiens

<400> 82

WO 03/012122

21/75

PCT/US02/23763

Val Glu Gly Gln Glu
1 5

<210> 83
<211> 5
<212> PRT
<213> Homo sapiens
<400> 83

Leu Gly Phe Ala Gly
1 5

<210> 84
<211> 16
<212> PRT
<213> Homo sapiens
<400> 84

Val Glu Gly Gln Glu Lys Ala His Asn Gln Asp Leu Gly Phe Ala Gly
1 5 10 15

<210> 85
<211> 5
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> X stands for glutamate or glutamine.

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X stands for serine, threonine, or glycine.

<400> 85

Xaa Gly Xaa Gly Gln
1 5

<210> 86
<211> 5
<212> PRT
<213> Homo sapiens

<400> 86

WO 03/012122

22/75

PCT/US02/23763

Glu Gly Ser Gly Gln
1 5

<210> 87
<211> 5
<212> PRT
<213> Homo sapiens

<400> 87

Glu Gly Thr Gly Gln
1 5

<210> 88
<211> 5
<212> PRT
<213> Homo sapiens

<400> 88

Glu Gly Gly Gly Gln
1 5

<210> 89
<211> 5
<212> PRT
<213> Homo sapiens

<400> 89

Gln Gly Gly Gly Gln
1 5

<210> 90
<211> 5
<212> PRT
<213> Homo sapiens

<400> 90

Thr Ser Ala Gly Ala
1 5

<210> 91
<211> 5
<212> PRT
<213> Homo sapiens

<400> 91

Ala Leu Ala Ser Pro
1 5

WO 03/012122

23/75

PCT/US02/23763

<210> 92
<211> 15
<212> PRT
<213> Homo sapiens

<400> 92

Thr Ser Ala Gly Ala Glu Gly Ser Gly Gln Ala Leu Ala Ser Pro
1 5 10 15

<210> 93
<211> 5
<212> PRT
<213> Homo sapiens

<400> 93

Thr Ser Ala Gly Ser
1 5

<210> 94
<211> 15
<212> PRT
<213> Homo sapiens

<400> 94

Thr Ser Ala Gly Ser Glu Gly Thr Gly Gln Ala Leu Ala Ser Pro
1 5 10 15

<210> 95
<211> 5
<212> PRT
<213> Homo sapiens

<400> 95

Thr Ala Ala Gly Asp
1 5

<210> 96
<211> 5
<212> PRT
<213> Homo sapiens

<400> 96

Ser Leu Val Ser Pro
1 5

WO 03/012122

24/75

PCT/US02/23763

<210> 97
<211> 15
<212> PRT
<213> Homo sapiens

<400> 97

Thr Ala Ala Gly Asp Glu Gly Gly Gly Gln Ser Leu Val Ser Pro
1 5 10 15

<210> 98
<211> 5
<212> PRT
<213> Homo sapiens

<400> 98

Thr Gly Ala Gly Asp
1 5

<210> 99
<211> 5
<212> PRT
<213> Homo sapiens

<400> 99

Ala Leu Met Ser Pro
1 5

<210> 100
<211> 15
<212> PRT
<213> Homo sapiens

<400> 100

Thr Gly Ala Gly Asp Gln Gly Gly Gly Gln Ala Leu Met Ser Pro
1 5 10 15

<210> 101
<211> 5
<212> PRT
<213> Homo sapiens

<400> 101

Thr Ala Ala Gly Ala
1 5

<210> 102
<211> 15

WO 03/012122

25/75

PCT/US02/23763

<212> PRT
<213> Homo sapiens

<400> 102

Thr Ala Ala Gly Ala Glu Gly Gly Gly Gln Ser Leu Val Ser Pro
1 5 10 15

<210> 103
<211> 4
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> X stands for glutamine or glutamate.

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X stands for asparagine or glutamine.

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X stands for glutamate, glutamine, or lysine.

<400> 103

Xaa Gly Xaa Xaa
1

<210> 104
<211> 4
<212> PRT
<213> Homo sapiens

<400> 104

Gln Gly Asn Glu
1

<210> 105
<211> 4
<212> PRT
<213> Homo sapiens

<400> 105

Gln Gly Asn Gln

WO 03/012122

26/75

PCT/US02/23763

1

<210> 106
<211> 4
<212> PRT
<213> Homo sapiens

<400> 106

Gln Gly Asn Lys
1

<210> 107
<211> 4
<212> PRT
<213> Homo sapiens

<400> 107

Glu Gly Gln Glu
1

<210> 108
<211> 5
<212> PRT
<213> Homo sapiens

<400> 108

Ser Leu Leu Tyr Val
1 5

<210> 109
<211> 5
<212> PRT
<213> Homo sapiens

<400> 109

Arg Ala His Gly Gln
1 5

<210> 110
<211> 14
<212> PRT
<213> Homo sapiens

<400> 110

Ser Leu Leu Tyr Val Gln Gly Asn Glu Arg Ala His Gly Gln
1 5 10

WO 03/012122

27/75

PCT/US02/23763

<210> 111
<211> 5
<212> PRT
<213> Homo sapiens

<400> 111

Ser Phe Leu Phe Val
1 5

<210> 112
<211> 14
<212> PRT
<213> Homo sapiens

<400> 112

Ser Phe Leu Phe Val Gln Gly Asn Gln Arg Ala His Gly Gln
1 5 10

<210> 113
<211> 14
<212> PRT
<213> Homo sapiens

<400> 113

Ser Leu Leu Tyr Val Gln Gly Asn Lys Arg Ala His Gly Gln
1 5 10

<210> 114
<211> 5
<212> PRT
<213> Homo sapiens

<400> 114

Ser Leu Leu Tyr Phe
1 5

<210> 115
<211> 5
<212> PRT
<213> Homo sapiens

<400> 115

Lys Ala His Asn Gln
1 5

<210> 116

WO 03/012122

28/75

PCT/US02/23763

<211> 14
<212> PRT
<213> Homo sapiens

<400> 116

Ser Leu Leu Tyr Phe Glu Gly Gln Glu Lys Ala His Asn Gln
1 5 10

<210> 117
<211> 5
<212> PRT
<213> Homo sapiens

<400> 117

Ser Leu Leu Tyr Leu
1 5

<210> 118
<211> 14
<212> PRT
<213> Homo sapiens

<400> 118

Ser Leu Leu Tyr Leu Glu Gly Gln Glu Lys Ala His Asn Gln
1 5 10

<210> 119
<211> 5
<212> PRT
<213> Homo sapiens

<400> 119

Ser Leu Leu Phe Val
1 5

<210> 120
<211> 14
<212> PRT
<213> Homo sapiens

<400> 120

Ser Leu Leu Phe Val Glu Gly Gln Glu Lys Ala His Asn Gln
1 5 10

<210> 121
<211> 6
<212> PRT

WO 03/012122

29/75

PCT/US02/23763

<213> Homo sapiens

<400> 121

Val Gln Gly Asn Glu Arg
1 5

<210> 122

<211> 6

<212> PRT

<213> Homo sapiens

<400> 122

Val Gln Gly Asn Gln Arg
1 5

<210> 123

<211> 6

<212> PRT

<213> Homo sapiens

<400> 123

Val Gln Gly Asn Lys Arg
1 5

<210> 124

<211> 6

<212> PRT

<213> Homo sapiens

<400> 124

Phe Glu Gly Gln Glu Lys
1 5

<210> 125

<211> 6

<212> PRT

<213> Homo sapiens

<400> 125

Leu Glu Gly Gln Glu Lys
1 5

<210> 126

<211> 6

<212> PRT

<213> Homo sapiens

WO 03/012122

30/75

PCT/US02/23763

<400> 126

Val Glu Gly Gln Glu Lys
1 5

<210> 127

<211> 5
<212> PRT
<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> X stands for serine, asparagine, or is absent.

<220>

<221> MISC_FEATURE

<222> (4)..(4)

<223> X stands for alanine, glutamine, or is absent.

<400> 127

Met Xaa Met Xaa Pro
1 5

<210> 128

<211> 5
<212> PRT
<213> Homo sapiens

<400> 128

Met Ser Met Ala Pro
1 5

<210> 129

<211> 5
<212> PRT
<213> Homo sapiens

<400> 129

Met Asn Met Ala Pro
1 5

<210> 130

<211> 5
<212> PRT
<213> Homo sapiens

<400> 130

WO 03/012122

31/75

PCT/US02/23763

Met Ser Met Gln Pro
1 5

<210> 131
<211> 3
<212> PRT
<213> Homo sapiens

<400> 131

Met Met Pro
1

<210> 132
<211> 5
<212> PRT
<213> Homo sapiens

<400> 132

Pro Glu Pro Met Pro
1 5

<210> 133
<211> 5
<212> PRT
<213> Homo sapiens

<400> 133

Ile Thr Gly Glu Asn
1 5

<210> 134
<211> 15
<212> PRT
<213> Homo sapiens

<400> 134

Pro Glu Pro Met Pro Met Ser Met Ala Pro Ile Thr Gly Glu Asn
1 5 10 15

<210> 135
<211> 5
<212> PRT
<213> Homo sapiens

<400> 135

Pro Ala Leu Met Pro

WO 03/012122

32/75

PCT/US02/23763

1 5
<210> 136
<211> 5
<212> PRT
<213> Homo sapiens

<400> 136
Ile Thr Gly Arg Ala
1 5

<210> 137
<211> 15
<212> PRT
<213> Homo sapiens

<400> 137
Pro Ala Leu Met Pro Met Asn Met Ala Pro Ile Thr Gly Arg Ala
1 5 10 15

<210> 138
<211> 5
<212> PRT
<213> Homo sapiens

<400> 138
Leu Lys Gly Gln Ser
1 5

<210> 139
<211> 15
<212> PRT
<213> Homo sapiens

<400> 139
Pro Glu Pro Met Pro Met Ser Met Gln Pro Leu Lys Gly Gln Ser
1 5 10 15

<210> 140
<211> 5
<212> PRT
<213> Homo sapiens

<400> 140
Thr Ala Pro Leu Pro
1 5

WO 03/012122

33/75

PCT/US02/23763

<210> 141
<211> 5
<212> PRT
<213> Homo sapiens

<400> 141

Val Ala Glu Asp Glu
1 5

<210> 142
<211> 13
<212> PRT
<213> Homo sapiens

<400> 142

Thr Ala Pro Leu Pro Met Met Pro Val Ala Glu Asp Glu
1 5 10

<210> 143
<211> 5
<212> PRT
<213> Homo sapiens

<400> 143

Ala Ala Pro Leu Pro
1 5

<210> 144
<211> 5
<212> PRT
<213> Homo sapiens

<400> 144

Leu Ser Glu Glu Ala
1 5

<210> 145
<211> 13
<212> PRT
<213> Homo sapiens

<400> 145

Ala Ala Pro Leu Pro Met Met Pro Leu Ser Glu Glu Ala
1 5 10

<210> 146

WO 03/012122

34/75

PCT/US02/23763

<211> 5
<212> PRT
<213> Homo sapiens

<400> 146

Thr Ala Pro Ile Pro
1 5

<210> 147
<211> 5
<212> PRT
<213> Homo sapiens

<400> 147

Val Ser Gln Thr Gln
1 5

<210> 148
<211> 13
<212> PRT
<213> Homo sapiens

<400> 148

Thr Ala Pro Ile Pro Met Met Pro Val Ser Gln Thr Gln
1 5 10

<210> 149
<211> 11
<212> PRT
<213> Homo sapiens

<400> 149

Pro Met Pro Met Ser Met Ala Pro Ile Thr Gly
1 5 10

<210> 150
<211> 11
<212> PRT
<213> Homo sapiens

<400> 150

Leu Met Pro Met Asn Met Ala Pro Ile Thr Gly
1 5 10

<210> 151
<211> 11
<212> PRT

WO 03/012122

35/75

PCT/US02/23763

<213> Homo sapiens

<400> 151

Pro Met Pro Met Ser Met Gln Pro Leu Lys Gly
1 5 10

<210> 152

<211> 9

<212> PRT

<213> Homo sapiens

<400> 152

Pro Leu Pro Met Met Pro Val Ala Glu
1 5

<210> 153

<211> 9

<212> PRT

<213> Homo sapiens

<400> 153

Pro Leu Pro Met Met Pro Leu Ser Glu
1 5

<210> 154

<211> 9

<212> PRT

<213> Homo sapiens

<400> 154

Pro Ile Pro Met Met Pro Val Ser Gln
1 5

<210> 155

<211> 4

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (3)..(3)

<223> X stands for alanine, serine, or aspartate.

<220>

<221> MISC_FEATURE

<222> (4)..(4)

<223> X stands for glutamate or glutamine.

WO 03/012122

36/75

PCT/US02/23763

<400> 155

Ala Gly Xaa Xaa
1<210> 156
<211> 4
<212> PRT
<213> Homo sapiens

<400> 156

Ala Gly Ala Glu
1<210> 157
<211> 4
<212> PRT
<213> Homo sapiens

<400> 157

Ala Gly Ser Glu
1<210> 158
<211> 4
<212> PRT
<213> Homo sapiens

<400> 158

Ala Gly Asp Glu
1<210> 159
<211> 4
<212> PRT
<213> Homo sapiens

<400> 159

Ala Gly Asp Gln
1<210> 160
<211> 5
<212> PRT
<213> Homo sapiens

<400> 160

WO 03/012122

37/75

PCT/US02/23763

Val Met His Thr Ser
1 5

<210> 161
<211> 5
<212> PRT
<213> Homo sapiens

<400> 161

Gly Ser Gly Gln Ala
1 5

<210> 162
<211> 14
<212> PRT
<213> Homo sapiens

<400> 162

Val Met His Thr Ser Ala Gly Ala Glu Gly Ser Gly Gln Ala
1 5 10

<210> 163
<211> 5
<212> PRT
<213> Homo sapiens

<400> 163

Ile Met Phe Thr Ser
1 5

<210> 164
<211> 5
<212> PRT
<213> Homo sapiens

<400> 164

Gly Thr Gly Gln Ala
1 5

<210> 165
<211> 14
<212> PRT
<213> Homo sapiens

<400> 165

Ile Met Phe Thr Ser Ala Gly Ser Glu Gly Thr Gly Gln Ala

WO 03/012122

38/75

PCT/US02/23763

1 5 10

<210> 166
<211> 5
<212> PRT
<213> Homo sapiens
<400> 166
Met Met His Thr Ser
1 5

<210> 167
<211> 14
<212> PRT
<213> Homo sapiens
<400> 167
Met Met His Thr Ser Ala Gly Ala Glu Gly Ser Gly Gln Ala
1 5 10

<210> 168
<211> 5
<212> PRT
<213> Homo sapiens
<400> 168
Leu Met His Thr Ala
1 5

<210> 169
<211> 5
<212> PRT
<213> Homo sapiens
<400> 169
Gly Gly Gly Gln Ser
1 5

<210> 170
<211> 14
<212> PRT
<213> Homo sapiens
<400> 170
Leu Met His Thr Ala Ala Gly Asp Glu Gly Gly Gly Gln Ser
1 5 10

WO 03/012122

39/75

PCT/US02/23763

<210> 171
<211> 5
<212> PRT
<213> Homo sapiens

<400> 171

Leu Met His Thr Gly
1 5

<210> 172
<211> 5
<212> PRT
<213> Homo sapiens

<400> 172

Gly Gly Gly Gln Ala
1 5

<210> 173
<211> 14
<212> PRT
<213> Homo sapiens

<400> 173

Leu Met His Thr Gly Ala Gly Asp Gln Gly Gly Gly Gln Ala
1 5 10

<210> 174
<211> 14
<212> PRT
<213> Homo sapiens

<400> 174

Leu Met His Thr Ala Ala Gly Ala Glu Gly Gly Gly Gln Ser
1 5 10

<210> 175
<211> 12
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X stands for histidine, asparagine, glutamine, or serine.

<220>

WO 03/012122

40/75

PCT/US02/23763

<221> MISC_FEATURE
<222> (5)..(5)
<223> X stands for glycine, arginine, alanine, or is absent.

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> X stands for arginine or glutamine.

<220>
<221> MISC_FEATURE
<222> (10)..(10)
<223> X stands for asparagine or histidine.

<220>
<221> MISC_FEATURE
<222> (11)..(11)
<223> X stands for phenylalanine or tyrosine.

<220>
<221> MISC_FEATURE
<222> (12)..(12)
<223> X stands for phenylalanine or tyrosine.

<400> 175

Glu Cys Xaa Gly Xaa Xaa Gly Thr Cys Xaa Xaa Xaa
1 5 10

<210> 176
<211> 11
<212> PRT
<213> Homo sapiens

<400> 176

Glu Cys His Gly Arg Gly Thr Cys Asn Tyr Tyr
1 5 10

<210> 177
<211> 12
<212> PRT
<213> Homo sapiens

<400> 177

Glu Cys Asn Gly Gly Arg Gly Thr Cys His Tyr Tyr
1 5 10

WO 03/012122

41/75

PCT/US02/23763

<210> 178
<211> 12
<212> PRT
<213> Homo sapiens

<400> 178

Glu Cys Gln Gly Arg Gln Gly Thr Cys His Phe Phe
1 5 10

<210> 179
<211> 12
<212> PRT
<213> Homo sapiens

<400> 179

Glu Cys Ser Gly Ala Arg Gly Thr Cys His Tyr Phe
1 5 10

<210> 180
<211> 5
<212> PRT
<213> Homo sapiens

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> X stands for proline, serine, or alanine.

<220>
<221> MISC_FEATURE
<222> (2)..(2)
<223> X stands for glutamate, aspartate, or serine.

<220>
<221> MISC_FEATURE
<222> (4)..(4)
<223> X stands for leucine or valine.

<400> 180

Xaa Xaa Thr Xaa Lys
1 5

<210> 181
<211> 5
<212> PRT
<213> Homo sapiens

<400> 181

WO 03/012122

42/75

PCT/US02/23763

Pro Ser Thr Leu Lys
1 5

<210> 182
<211> 5
<212> PRT
<213> Homo sapiens

<400> 182

Pro Ser Thr Val Lys
1 5

<210> 183
<211> 5
<212> PRT
<213> Homo sapiens

<400> 183

Ser Glu Thr Leu Lys
1 5

<210> 184
<211> 5
<212> PRT
<213> Homo sapiens

<400> 184

Ala Asp Thr Leu Lys
1 5

<210> 185
<211> 5
<212> PRT
<213> Homo sapiens

<400> 185

Pro Asp Thr Leu Lys
1 5

<210> 186
<211> 5
<212> PRT
<213> Homo sapiens

<400> 186

Phe Lys Lys Pro Thr

WO 03/012122

43/75

PCT/US02/23763

1 5

<210> 187
<211> 5
<212> PRT
<213> Homo sapiens

<400> 187

Ala Gly Glu Leu Arg
1 5

<210> 188
<211> 15
<212> PRT
<213> Homo sapiens

<400> 188

Phe Lys Lys Pro Thr Pro Ser Thr Leu Lys Ala Gly Glu Leu Arg
1 5 10 15

<210> 189
<211> 5
<212> PRT
<213> Homo sapiens

<400> 189

Phe Arg Lys Pro Ile
1 5

<210> 190
<211> 5
<212> PRT
<213> Homo sapiens

<400> 190

Ala Gly Glu Leu Glu
1 5

<210> 191
<211> 15
<212> PRT
<213> Homo sapiens

<400> 191

Phe Arg Lys Pro Ile Pro Ser Thr Val Lys Ala Gly Glu Leu Glu
1 5 10 15

WO 03/012122

44/75

PCT/US02/23763

<210> 192
<211> 5
<212> PRT
<213> Homo sapiens

<400> 192

Phe Ser Lys Pro Gln
1 5

<210> 193
<211> 5
<212> PRT
<213> Homo sapiens

<400> 193

Ala Gly Asp Leu Arg
1 5

<210> 194
<211> 15
<212> PRT
<213> Homo sapiens

<400> 194

Phe Ser Lys Pro Gln Ser Glu Thr Leu Lys Ala Gly Asp Leu Arg
1 5 10 15

<210> 195
<211> 5
<212> PRT
<213> Homo sapiens

<400> 195

Gln Gly Ser Pro Ser
1 5

<210> 196
<211> 5
<212> PRT
<213> Homo sapiens

<400> 196

Ala Gly Leu Ile Arg
1 5

<210> 197

WO 03/012122

45/75

PCT/US02/23763

<211> 15
<212> PRT
<213> Homo sapiens
<400> 197

Gln Gly Ser Pro Ser Ala Asp Thr Leu Lys Ala Gly Leu Ile Arg
1 5 10 15

<210> 198
<211> 5
<212> PRT
<213> Homo sapiens

<400> 198

Ser Ser Ala Pro Ala
1 5

<210> 199
<211> 5
<212> PRT
<213> Homo sapiens

<400> 199

Glu Ser Gln Ala Gln
1 5

<210> 200
<211> 15
<212> PRT
<213> Homo sapiens

<400> 200

Ser Ser Ala Pro Ala Pro Asp Thr Leu Lys Glu Ser Gln Ala Gln
1 5 10 15

<210> 201
<211> 5
<212> PRT
<213> Homo sapiens

<400> 201

Gly Glu Leu Pro Val
1 5

<210> 202
<211> 5
<212> PRT

WO 03/012122

46/75

PCT/US02/23763

<213> Homo sapiens

<400> 202

Ala Gly Gln Leu His
1 5

<210> 203

<211> 15

<212> PRT

<213> Homo sapiens

<400> 203

Gly Glu Leu Pro Val Ser Glu Thr Leu Lys Ala Gly Gln Leu His
1 5 10 15

<210> 204

<211> 5

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> X stands for serine, glutamine, or arginine.

<400> 204

Ala Xaa Arg Asn Asp
1 5

<210> 205

<211> 5

<212> PRT

<213> Homo sapiens

<400> 205

Ala Ser Arg Asn Asp
1 5

<210> 206

<211> 5

<212> PRT

<213> Homo sapiens

<400> 206

Ala Gln Arg Asn Asp
1 5

WO 03/012122

47/75

PCT/US02/23763

<210> 207
<211> 5
<212> PRT
<213> Homo sapiens

<400> 207

Ala Arg Arg Asn Asp
1 5

<210> 208
<211> 5
<212> PRT
<213> Homo sapiens

<400> 208

Asn Val Cys Asn Phe
1 5

<210> 209
<211> 5
<212> PRT
<213> Homo sapiens

<400> 209

Tyr Ser Tyr Trp Leu
1 5

<210> 210
<211> 15
<212> PRT
<213> Homo sapiens

<400> 210

Asn Val Cys Asn Phe Ala Ser Arg Asn Asp Tyr Ser Tyr Trp Leu
1 5 10 15

<210> 211
<211> 5
<212> PRT
<213> Homo sapiens

<400> 211

Asp Val Cys Asn Phe
1 5

<210> 212

WO 03/012122

48/75

PCT/US02/23763

<211> 15
<212> PRT
<213> Homo sapiens

<400> 212

Asp Val Cys Asn Phe Ala Ser Arg Asn Asp Tyr Ser Tyr Trp Leu
1 5 10 15

<210> 213
<211> 5
<212> PRT
<213> Homo sapiens

<400> 213

Asp Val Cys Tyr Tyr
1 5

<210> 214
<211> 5
<212> PRT
<213> Homo sapiens

<400> 214

Lys Ser Tyr Trp Leu
1 5

<210> 215
<211> 15
<212> PRT
<213> Homo sapiens

<400> 215

Asp Val Cys Tyr Tyr Ala Ser Arg Asn Asp Lys Ser Tyr Trp Leu
1 5 10 15

<210> 216
<211> 5
<212> PRT
<213> Homo sapiens

<400> 216

Gln Val Cys His Tyr
1 5

<210> 217
<211> 5
<212> PRT

WO 03/012122

49/75

PCT/US02/23763

<213> Homo sapiens

<400> 217

Arg Ser Tyr Trp Leu
1 5

<210> 218

<211> 15

<212> PRT

<213> Homo sapiens

<400> 218

Gln Val Cys His Tyr Ala Gln Arg Asn Asp Arg Ser Tyr Trp Leu
1 5 10 15

<210> 219

<211> 5

<212> PRT

<213> Homo sapiens

<400> 219

Glu Val Cys His Tyr
1 5

<210> 220

<211> 15

<212> PRT

<213> Homo sapiens

<400> 220

Glu Val Cys His Tyr Ala Arg Arg Asn Asp Lys Ser Tyr Trp Leu
1 5 10 15

<210> 221

<211> 5

<212> PRT

<213> Homo sapiens

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> X stands for tyrosine or phenylalanine.

<220>

<221> MISC_FEATURE

<222> (2)..(2)

<223> X stands for tyrosine or phenylalanine.

WO 03/012122

50/75

PCT/US02/23763

<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> X stands for alanine or serine.

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> X stands for alanine, serine, or lysine.

<400> 221

Xaa Xaa Xaa Asn Xaa
1 5

<210> 222
<211> 5
<212> PRT
<213> Homo sapiens

<400> 222

Tyr Tyr Ala Asn Ala
1 5

<210> 223
<211> 5
<212> PRT
<213> Homo sapiens

<400> 223

Tyr Tyr Ser Asn Ser
1 5

<210> 224
<211> 5
<212> PRT
<213> Homo sapiens

<400> 224

Tyr Tyr Ala Asn Ser
1 5

<210> 225
<211> 5
<212> PRT
<213> Homo sapiens

WO 03/012122

51/75

PCT/US02/23763

<400> 225
Tyr Tyr Ala Asn Lys
1 5

<210> 226
<211> 5
<212> PRT
<213> Homo sapiens

<400> 226
Phe Phe Ala Asn Lys
1 5

<210> 227
<211> 5
<212> PRT
<213> Homo sapiens

<400> 227
Tyr Phe Ala Asn Lys
1 5

<210> 228
<211> 5
<212> PRT
<213> Homo sapiens

<400> 228
Arg Gly Thr Cys Asn
1 5

<210> 229
<211> 5
<212> PRT
<213> Homo sapiens

<400> 229
Tyr Ser Phe Trp Leu
1 5

<210> 230
<211> 15
<212> PRT
<213> Homo sapiens

<400> 230

WO 03/012122

52/75

PCT/US02/23763

Arg Gly Thr Cys Asn Tyr Tyr Ala Asn Ala Tyr Ser Phe Trp Leu
1 5 10 15

<210> 231
<211> 15
<212> PRT
<213> Homo sapiens

<400> 231

Arg Gly Thr Cys Asn Tyr Tyr Ser Asn Ser Tyr Ser Phe Trp Leu
1 5 10 15

<210> 232
<211> 15
<212> PRT
<213> Homo sapiens

<400> 232

Arg Gly Thr Cys Asn Tyr Tyr Ala Asn Ser Tyr Ser Phe Trp Leu
1 5 10 15

<210> 233
<211> 5
<212> PRT
<213> Homo sapiens

<400> 233

Arg Gly Thr Cys His
1 5

<210> 234
<211> 15
<212> PRT
<213> Homo sapiens

<400> 234

Arg Gly Thr Cys His Tyr Tyr Ala Asn Lys Tyr Ser Phe Trp Leu
1 5 10 15

<210> 235
<211> 5
<212> PRT
<213> Homo sapiens

<400> 235

Gln Gly Thr Cys His
1 5

WO 03/012122

53/75

PCT/US02/23763

<210> 236
<211> 15
<212> PRT
<213> Homo sapiens

<400> 236

Gln Gly Thr Cys His Phe Phe Ala Asn Lys Tyr Ser Phe Trp Leu
1 5 10 15

<210> 237
<211> 15
<212> PRT
<213> Homo sapiens

<400> 237

Arg Gly Thr Cys His Tyr Phe Ala Asn Lys Tyr Ser Phe Trp Leu
1 5 10 15

<210> 238
<211> 11
<212> PRT
<213> Homo sapiens

<400> 238

Ile Glu Arg Ser Glu Met Phe Lys Lys Pro Thr
1 5 10

<210> 239
<211> 11
<212> PRT
<213> Homo sapiens

<400> 239

Leu Asn Pro Glu Arg Met Phe Arg Lys Pro Ile
1 5 10

<210> 240
<211> 11
<212> PRT
<213> Homo sapiens

<400> 240

Val Asp Val Ser Asp Met Phe Ser Lys Pro Gln
1 5 10

WO 03/012122

54/75

PCT/US02/23763

<210> 241
<211> 11
<212> PRT
<213> Homo sapiens

<400> 241

Ile Pro Glu Gln Ser Phe Gln Gly Ser Pro Ser
1 5 10

<210> 242
<211> 12
<212> PRT
<213> Homo sapiens

<400> 242

Val Lys Ala Asp Leu Gln Phe Ser Ser Ala Pro Ala
1 5 10

<210> 243
<211> 12
<212> PRT
<213> Homo sapiens

<400> 243

Val Glu Glu Arg Gln Gln Phe Gly Glu Leu Pro Val
1 5 10

<210> 244
<211> 5
<212> PRT
<213> Homo sapiens

<400> 244

Phe Trp Leu Ala Thr
1 5

<210> 245
<211> 21
<212> PRT
<213> Homo sapiens

<400> 245

Phe Trp Leu Ala Thr Ile Glu Arg Ser Glu Met Phe Lys Lys Pro Thr
1 5 10 15

Pro Ser Thr Leu Lys
20

WO 03/012122

55/75

PCT/US02/23763

<210> 246
<211> 5
<212> PRT
<213> Homo sapiens

<400> 246

Phe Trp Leu Ala Ser
1 5

<210> 247
<211> 21
<212> PRT
<213> Homo sapiens

<400> 247

Phe Trp Leu Ala Ser Leu Asn Pro Glu Arg Met Phe Arg Lys Pro Ile
1 5 10 15

Pro Ser Thr Val Lys
20

<210> 248
<211> 21
<212> PRT
<213> Homo sapiens

<400> 248

Phe Trp Leu Ala Thr Val Asp Val Ser Asp Met Phe Ser Lys Pro Gln
1 5 10 15

Ser Glu Thr Leu Lys
20

<210> 249
<211> 5
<212> PRT
<213> Homo sapiens

<400> 249

Phe Trp Leu Thr Thr
1 5

<210> 250
<211> 21
<212> PRT

WO 03/012122

56/75

PCT/US02/23763

<213> Homo sapiens

<400> 250

Phe Trp Leu Thr Thr Ile Pro Glu Gln Ser Phe Gln Gly Ser Pro Ser
 1 5 10 15

Ala Asp Thr Leu Lys
 20

<210> 251

<211> 22

<212> PRT

<213> Homo sapiens

<400> 251

Phe Trp Leu Thr Thr Val Lys Ala Asp Leu Gln Phe Ser Ser Ala Pro
 1 5 10 15

Ala Pro Asp Thr Leu Lys
 20

<210> 252

<211> 22

<212> PRT

<213> Homo sapiens

<400> 252

Phe Trp Leu Thr Thr Val Glu Glu Arg Gln Gln Phe Gly Glu Leu Pro
 1 5 10 15

Val Ser Glu Thr Leu Lys
 20

<210> 253

<211> 18

<212> PRT

<213> Homo sapiens

<400> 253

Phe Ser Thr Met Pro Phe Leu Phe Cys Asn Ile Asn Asn Val Cys Asn
 1 5 10 15

Phe Ala

WO 03/012122

57/75

PCT/US02/23763

<210> 254
<211> 18
<212> PRT
<213> Homo sapiens

<400> 254

Phe Thr Thr Met Pro Phe Leu Phe Cys Asn Val Asn Asp Val Cys Asn
1 5 10 15

Phe Ala

<210> 255
<211> 18
<212> PRT
<213> Homo sapiens

<400> 255

Phe Ser Thr Met Pro Phe Met Phe Cys Asn Ile Asn Asn Val Cys Asn
1 5 10 15

Phe Ala

<210> 256
<211> 18
<212> PRT
<213> Homo sapiens

<400> 256

Phe Ser Thr Met Pro Phe Leu Tyr Cys Asn Pro Gly Asp Val Cys Tyr
1 5 10 15

Tyr Ala

<210> 257
<211> 18
<212> PRT
<213> Homo sapiens

<400> 257

Phe Ser Thr Leu Pro Phe Ala Tyr Cys Asn Ile His Gln Val Cys His
1 5 10 15

Tyr Ala

WO 03/012122

58/75

PCT/US02/23763

<210> 258
<211> 18
<212> PRT
<213> Homo sapiens

<400> 258

Phe Ser Thr Met Pro Phe Ile Tyr Cys Asn Ile Asn Glu Val Cys His
1 5 10 15

Tyr Ala

<210> 259
<211> 18
<212> PRT
<213> Homo sapiens

<400> 259

Pro Phe Leu Phe Cys Asn Ile Asn Asn Val Cys Asn Phe Ala Ser Arg
1 5 10 15

Asn Asp

<210> 260
<211> 18
<212> PRT
<213> Homo sapiens

<400> 260

Pro Phe Leu Phe Cys Asn Val Asn Asp Val Cys Asn Phe Ala Ser Arg
1 5 10 15

Asn Asp

<210> 261
<211> 18
<212> PRT
<213> Homo sapiens

<400> 261

Pro Phe Met Phe Cys Asn Ile Asn Asn Val Cys Asn Phe Ala Ser Arg
1 5 10 15

WO 03/012122

59/75

PCT/US02/23763

Asn Asp

<210> 262
<211> 18
<212> PRT
<213> Homo sapiens

<400> 262

Pro Phe Leu Tyr Cys Asn Pro Gly Asp Val Cys Tyr Tyr Ala Ser Arg
1 5 10 15

Asn Asp

<210> 263
<211> 18
<212> PRT
<213> Homo sapiens

<400> 263

Pro Phe Ala Tyr Cys Asn Ile His Gln Val Cys His Tyr Ala Gln Arg
1 5 10 15

Asn Asp

<210> 264
<211> 18
<212> PRT
<213> Homo sapiens

<400> 264

Pro Phe Ile Tyr Cys Asn Ile Asn Glu Val Cys His Tyr Ala Arg Arg
1 5 10 15

Asn Asp

<210> 265
<211> 22
<212> PRT
<213> Homo sapiens

<400> 265

WO 03/012122

60/75

PCT/US02/23763

Phe Ser Thr Met Pro Phe Leu Phe Cys Asn Ile Asn Asn Val Cys Asn
1 5 10 15

Phe Ala Ser Arg Asn Asp
20

<210> 266
<211> 22
<212> PRT
<213> Homo sapiens

<400> 266

Phe Thr Thr Met Pro Phe Leu Phe Cys Asn Val Asn Asp Val Cys Asn
1 5 10 15

Phe Ala Ser Arg Asn Asp
20

<210> 267
<211> 22
<212> PRT
<213> Homo sapiens

<400> 267

Phe Ser Thr Met Pro Phe Met Phe Cys Asn Ile Asn Asn Val Cys Asn
1 5 10 15

Phe Ala Ser Arg Asn Asp
20

<210> 268
<211> 22
<212> PRT
<213> Homo sapiens

<400> 268

Phe Ser Thr Met Pro Phe Leu Tyr Cys Asn Pro Gly Asp Val Cys Tyr
1 5 10 15

Tyr Ala Ser Arg Asn Asp
20

<210> 269
<211> 22
<212> PRT

WO 03/012122

61/75

PCT/US02/23763

<213> Homo sapiens

<400> 269

Phe Ser Thr Leu Pro Phe Ala Tyr Cys Asn Ile His Gln Val Cys His
 1 5 10 15

Tyr Ala Gln Arg Asn Asp
 20

<210> 270

<211> 22

<212> PRT

<213> Homo sapiens

<400> 270

Phe Ser Thr Met Pro Phe Ile Tyr Cys Asn Ile Asn Glu Val Cys His
 1 5 10 15

Tyr Ala Arg Arg Asn Asp
 20

<210> 271

<211> 14

<212> PRT

<213> Homo sapiens

<400> 271

Pro Phe Ile Glu Cys His Gly Arg Gly Thr Cys Asn Tyr Tyr
 1 5 10

<210> 272

<211> 14

<212> PRT

<213> Homo sapiens

<400> 272

Pro Phe Leu Glu Cys His Gly Arg Gly Thr Cys Asn Tyr Tyr
 1 5 10

<210> 273

<211> 15

<212> PRT

<213> Homo sapiens

<400> 273

Pro Phe Ile Glu Cys Asn Gly Gly Arg Gly Thr Cys His Tyr Tyr

WO 03/012122

62/75

PCT/US02/23763

1 5 10 15

<210> 274
<211> 15
<212> PRT
<213> Homo sapiens

<400> 274

Pro Phe Leu Glu Cys Gln Gly Arg Gln Gly Thr Cys His Phe Phe
1 5 10 15

<210> 275
<211> 15
<212> PRT
<213> Homo sapiens

<400> 275

Pro Phe Ile Glu Cys Ser Gly Ala Arg Gly Thr Cys His Tyr Phe
1 5 10 15

<210> 276
<211> 18
<212> PRT
<213> Homo sapiens

<400> 276

Ile Glu Arg Ser Glu Met Phe Lys Lys Pro Thr Pro Ser Thr Leu Lys
1 5 10 15

Ala Gly

<210> 277
<211> 18
<212> PRT
<213> Homo sapiens

<400> 277

Leu Asn Pro Glu Arg Met Phe Arg Lys Pro Ile Pro Ser Thr Val Lys
1 5 10 15

Ala Gly

<210> 278
<211> 18

WO 03/012122

63/75

PCT/US02/23763

<212> PRT
<213> Homo sapiens

<400> 278

Val Asp Val Ser Asp Met Phe Ser Lys Pro Gln Ser Glu Thr Leu Lys
1 5 10 15

Ala Gly

<210> 279
<211> 18
<212> PRT
<213> Homo sapiens

<400> 279

Ile Pro Glu Gln Ser Phe Cln Gly Ser Pro Ser Ala Asp Thr Leu Lys
1 5 10 15

Ala Gly

<210> 280
<211> 19
<212> PRT
<213> Homo sapiens

<400> 280

Val Lys Ala Asp Leu Gln Phe Ser Ser Ala Pro Ala Pro Asp Thr Leu
1 5 10 15

Lys Glu Ser

<210> 281
<211> 19
<212> PRT
<213> Homo sapiens

<400> 281

Val Glu Glu Arg Gln Gln Phe Gly Glu Leu Pro Val Ser Glu Thr Leu
1 5 10 15

Lys Ala Gly

WO 03/012122

64/75

PCT/US02/23763

<210> 282
<211> 10
<212> PRT
<213> Homo sapiens

<400> 282

Gly Ser Cys Leu Arg Lys Phe Ser Thr Met
1 5 10

<210> 283
<211> 10
<212> PRT
<213> Homo sapiens

<400> 283

Gly Ser Cys Leu Gln Arg Phe Thr Thr Met
1 5 10

<210> 284
<211> 10
<212> PRT
<213> Homo sapiens

<400> 284

Gly Ser Cys Leu Arg Arg Phe Ser Thr Met
1 5 10

<210> 285
<211> 10
<212> PRT
<213> Homo sapiens

<400> 285

Gly Ser Cys Leu Ala Arg Phe Ser Thr Met
1 5 10

<210> 286
<211> 10
<212> PRT
<213> Homo sapiens

<400> 286

Gly Ser Cys Leu Pro Val Phe Ser Thr Leu
1 5 10

<210> 287

WO 03/012122

65/75

PCT/US02/23763

<211> 10
<212> PRT
<213> Homo sapiens

<400> 287

Gly Ser Cys Leu Pro Arg Phe Ser Thr Met
1 5 10

<210> 288
<211> 20
<212> PRT
<213> Homo sapiens

<400> 288

Leu Arg Lys Phe Ser Thr Met Pro Phe Leu Phe Cys Asn Ile Asn Asn
1 5 10 15

Val Cys Asn Phe
20

<210> 289
<211> 20
<212> PRT
<213> Homo sapiens

<400> 289

Leu Gln Arg Phe Thr Thr Met Pro Phe Leu Phe Cys Asn Val Asn Asp
1 5 10 15

Val Cys Asn Phe
20

<210> 290
<211> 20
<212> PRT
<213> Homo sapiens

<400> 290

Leu Arg Arg Phe Ser Thr Met Pro Phe Met Phe Cys Asn Ile Asn Asn
1 5 10 15

Val Cys Asn Phe
20

<210> 291
<211> 20

WO 03/012122

66/75

PCT/US02/23763

<212> PRT
<213> Homo sapiens

<400> 291

Leu Ala Arg Phe Ser Thr Met Pro Phe Leu Tyr Cys Asn Pro Gly Asp
1 5 10 15

Val Cys Tyr Tyr
20

<210> 292
<211> 20
<212> PRT
<213> Homo sapiens

<400> 292

Leu Pro Val Phe Ser Thr Leu Pro Phe Ala Tyr Cys Asn Ile His Gln
1 5 10 15

Val Cys His Tyr
20

<210> 293
<211> 20
<212> PRT
<213> Homo sapiens

<400> 293

Leu Pro Arg Phe Ser Thr Met Pro Phe Ile Tyr Cys Asn Ile Asn Glu
1 5 10 15

Val Cys His Tyr
20

<210> 294
<211> 4
<212> PRT
<213> Homo sapiens

<400> 294

Leu Arg Lys Phe
1

<210> 295
<211> 4
<212> PRT

WO 03/012122

67/75

PCT/US02/23763

<213> Homo sapiens

<400> 295

Leu Ala Arg Phe
1

<210> 296

<211> 2

<212> PRT

<213> Homo sapiens

<400> 296

Gln Asp
1

<210> 297

<211> 3

<212> PRT

<213> Homo sapiens

<400> 297

Asn Tyr Tyr
1

<210> 298

<211> 10

<212> PRT

<213> Homo sapiens

<400> 298

Phe Cys Asn Val Asn Asp Val Cys Asn Phe
1 5 10

<210> 299

<211> 10

<212> PRT

<213> Homo sapiens

<400> 299

Tyr Cys Asn Pro Gly Asp Val Cys Tyr Tyr
1 5 10

<210> 300

<211> 10

<212> PRT

<213> Homo sapiens

WO 03/012122

68/75

PCT/US02/23763

<400> 300

Tyr Cys Asn Ile His Gln Val Cys His Tyr
 1 5 10

<210> 301

<211> 10

<212> PRT

<213> Homo sapiens

<400> 301

Tyr Cys Asn Ile Asn Glu Val Cys His Tyr
 1 5 10

<210> 302

<211> 229

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> alpha 1 chain

<400> 302

Ser Val Asp His Gly Phe Leu Val Thr Arg His Ser Gln Thr Ile Asp
 1 5 10 15

Asp Pro Gln Cys Pro Ser Gly Thr Lys Ile Leu Tyr His Gly Tyr Ser
 20 25 30

Leu Leu Tyr Val Gln Gly Asn Glu Arg Ala His Gly Gln Asp Leu Gly
 35 40 45

Thr Ala Gly Ser Cys Leu Arg Lys Phe Ser Thr Met Pro Phe Leu Phe
 50 55 60

Cys Asn Ile Asn Asn Val Cys Asn Phe Ala Ser Arg Asn Asp Tyr Ser
 65 70 75 80

Tyr Trp Leu Ser Thr Pro Glu Pro Met Pro Met Ser Met Ala Pro Ile
 85 90 95

Thr Gly Glu Asn Ile Arg Pro Phe Ile Ser Arg Cys Ala Val Cys Glu
 100 105 110

WO 03/012122

69/75

PCT/US02/23763

Ala Pro Ala Met Val Met Ala Val His Ser Gln Thr Ile Gln Ile Pro
 115 120 125

Pro Cys Pro Ser Gly Trp Ser Ser Leu Trp Ile Gly Tyr Ser Phe Val
 130 135 140

Met His Thr Ser Ala Gly Ala Glu Gly Ser Gly Gln Ala Leu Ala Ser
 145 150 155 160

Pro Gly Ser Cys Leu Glu Glu Phe Arg Ser Ala Pro Phe Ile Glu Cys
 165 170 175

His Gly Arg Gly Thr Cys Asn Tyr Tyr Ala Asn Ala Tyr Ser Phe Trp
 180 185 190

Leu Ala Thr Ile Glu Arg Ser Glu Met Phe Lys Lys Pro Thr Pro Ser
 195 200 205

Thr Leu Lys Ala Gly Glu Leu Arg Thr His Val Ser Arg Cys Gln Val
 210 215 220

Cys Met Arg Arg Thr
 225

<210> 303
 <211> 227
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> alpha 2 chain

<400> 303

Val Ser Ile Gly Tyr Leu Leu Val Lys His Ser Gln Thr Asp Gln Glu
 1 5 10 15

Pro Met Cys Pro Val Gly Met Asn Lys Leu Trp Ser Gly Tyr Ser Leu
 20 25 30

Leu Tyr Phe Glu Gly Gln Glu Lys Ala His Asn Gln Asp Leu Gly Leu
 35 40 45

Ala Gly Ser Cys Leu Ala Arg Phe Ser Thr Met Pro Phe Leu Tyr Cys

WO 03/012122

70/75

PCT/US02/23763

50 55 60
 Asn Pro Gly Asp Val Cys Tyr Tyr Ala Ser Arg Asn Asp Lys Ser Tyr
 65 70 75 80
 Trp Leu Ser Thr Thr Ala Pro Leu Pro Met Met Pro Val Ala Glu Asp
 85 90 95
 Glu Ile Lys Pro Tyr Ile Ser Arg Cys Ser Val Cys Glu Ala Pro Ala
 100 105 110
 Ile Ala Ile Ala Val His Ser Gln Asp Val Ser Ile Pro His Cys Pro
 115 120 125
 Ala Gly Trp Arg Ser Leu Trp Ile Gly Tyr Ser Phe Leu Met His Thr
 130 135 140
 Ala Ala Gly Asp Glu Gly Gly Gly Gln Ser Leu Val Ser Pro Gly Ser
 145 150 155 160
 Cys Leu Glu Asp Phe Arg Ala Thr Pro Phe Ile Glu Cys Asn Gly Gly
 165 170 175
 Arg Gly Thr Cys His Tyr Tyr Ala Asn Lys Tyr Ser Phe Trp Leu Thr
 180 185 190
 Thr Ile Pro Glu Gln Ser Phe Gln Gly Ser Pro Ser Ala Asp Thr Leu
 195 200 205
 Lys Ala Gly Leu Ile Arg Thr His Ile Ser Arg Cys Gln Val Cys Met
 210 215 220
 Lys Asn Leu
 225
 <210> 304
 <211> 232
 <212> FRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> alpha 3 chain

WO 03/012122

71/75

PCT/US02/23763

<400> 304

Ala Thr Trp Thr Thr Arg Gly Phe Val Phe Thr Arg His Ser Gln Thr
 1 5 10 15

Thr Ala Ile Pro Ser Cys Pro Glu Gly Thr Val Pro Leu Tyr Ser Gly
 20 25 30

Phe Ser Phe Leu Phe Val Gln Gly Asn Gln Arg Ala His Gly Gln Asp
 35 40 45

Leu Gly Thr Leu Gly Ser Cys Leu Gln Arg Phe Thr Thr Met Pro Phe
 50 55 60

Leu Phe Cys Asn Val Asn Asp Val Cys Asn Phe Ala Ser Arg Asn Asp
 65 70 75 80

Tyr Ser Tyr Trp Leu Ser Thr Pro Ala Leu Met Pro Met Asn Met Ala
 85 90 95

Pro Ile Thr Gly Arg Ala Leu Glu Pro Tyr Ile Ser Arg Cys Thr Val
 100 105 110

Cys Glu Gly Pro Ala Ile Ala Ile Ala Val His Ser Gln Thr Thr Asp
 115 120 125

Ile Pro Pro Cys Pro His Gly Trp Ile Ser Leu Trp Lys Gly Phe Ser
 130 135 140

Phe Ile Met Phe Thr Ser Ala Gly Ser Glu Gly Ala Gly Gln Ala Leu
 145 150 155 160

Ala Ser Pro Gly Ser Cys Leu Glu Glu Phe Arg Ala Ser Pro Phe Leu
 165 170 175

Glu Cys His Gly Arg Gly Thr Cys Asn Tyr Tyr Ser Asn Ser Tyr Ser
 180 185 190

Phe Trp Leu Ala Ser Leu Asn Pro Glu Arg Met Phe Arg Lys Pro Ile
 195 200 205

Pro Ser Thr Val Lys Ala Gly Glu Leu Glu Lys Ile Ile Ser Arg Cys
 210 215 220

WO 03/012122

72/75

PCT/US02/23763

Gln Val Cys Met Lys Lys Arg His
 225 230

<210> 305
 <211> 231
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> alpha 4 chain

<400> 305

Pro Gly Tyr Leu Gly Gly Phe Leu Leu Val Leu His Ser Gln Thr Asp
 1 5 10 15

Gln Glu Pro Thr Cys Pro Leu Gly Met Pro Arg Leu Trp Thr Gly Tyr
 20 25 30

Ser Leu Leu Tyr Leu Glu Gly Gln Glu Lys Ala His Asn Gln Asp Leu
 35 40 45

Gly Leu Ala Gly Ser Cys Leu Pro Val Phe Ser Thr Leu Pro Phe Ala
 50 55 60

Tyr Cys Asn Ile His Gln Val Cys His Tyr Ala Gln Arg Asn Asp Arg
 65 70 75 80

Ser Tyr Trp Leu Ala Ser Ala Ala Pro Leu Pro Met Met Pro Leu Ser
 85 90 95

Glu Glu Ala Ile Arg Pro Tyr Val Ser Arg Cys Ala Val Cys Glu Ala
 100 105 110

Pro Ala Gln Ala Val Ala Val His Ser Gln Asp Gln Ser Ile Pro Pro
 115 120 125

Cys Pro Gln Thr Trp Arg Ser Leu Trp Ile Gly Tyr Ser Phe Leu Met
 130 135 140

His Thr Gly Ala Gly Asp Gln Gly Gly Gly Gln Ala Leu Met Ser Pro
 145 150 155 160

WO 03/012122

73/75

PCT/US02/23763

Gly Ser Cys Leu Glu Asp Phe Arg Ala Ala Pro Phe Leu Glu Cys Gln
 165 170 175

Gly Arg Gln Gly Thr Cys His Phe Phe Ala Asn Lys Tyr Ser Phe Trp
 180 185 190

Leu Thr Thr Val Lys Ala Asp Leu Gln Phe Ser Ser Ala Pro Ala Pro
 195 200 205

Asp Thr Leu Lys Glu Ser Gln Ala Gln Arg Gln Lys Ile Ser Arg Cys
 210 215 220

Gln Val Cys Val Lys Tyr Ser
 225 230

<210> 306
 <211> 229
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> alpha 5 chain

<400> 306

Ser Val Ala His Gly Phe Leu Ile Thr Arg His Ser Gln Thr Thr Asp
 1 5 10 15

Ala Pro Gln Cys Pro Gln Gly Thr Leu Gln Val Tyr Glu Gly Phe Ser
 20 25 30

Leu Leu Tyr Val Gln Gly Asn Lys Arg Ala His Gly Gln Asp Leu Gly
 35 40 45

Thr Ala Gly Ser Cys Leu Arg Arg Phe Ser Thr Met Pro Phe Met Phe
 50 55 60

Cys Asn Ile Asn Asn Val Cys Asn Phe Ala Ser Arg Asn Asp Tyr Ser
 65 70 75 80

Tyr Trp Leu Ser Thr Pro Glu Pro Met Pro Met Ser Met Gln Pro Leu
 85 90 95

Lys Gly Gln Ser Ile Gln Pro Phe Ile Ser Arg Cys Ala Val Cys Glu

WO 03/012122

74/75

PCT/US02/23763

100 105 110
 Ala Pro Ala Val Val Ile Ala Val His Ser Gln Thr Ile Gln Ile Pro
 115 120 125
 His Cys Pro Gln Gly Trp Asp Ser Leu Trp Ile Gly Tyr Ser Phe Met
 130 135 140
 Met His Thr Ser Ala Gly Ala Glu Gly Ser Gly Gln Ala Leu Ala Ser
 145 150 155 160
 Pro Gly Ser Cys Leu Glu Glu Phe Arg Ser Ala Pro Phe Ile Glu Cys
 165 170 175
 His Gly Arg Gly Thr Cys Asn Tyr Tyr Ala Asn Ser Tyr Ser Phe Trp
 180 185 190
 Leu Ala Thr Val Asp Val Ser Asp Met Phe Ser Lys Pro Gln Ser Glu
 195 200 205
 Thr Leu Lys Ala Gly Asp Leu Arg Thr Arg Ile Ser Arg Cys Gln Val
 210 215 220
 Cys Met Lys Arg Thr
 225
 <210> 307
 <211> 228
 <212> FRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> alpha 6 chain
 <400> 307
 Met Arg Val Gly Tyr Thr Leu Val Lys His Ser Gln Ser Glu Gln Val
 1 5 10 15
 Pro Pro Cys Pro Ile Gly Met Ser Gln Leu Trp Val Gly Tyr Ser Leu
 20 25 30
 Leu Phe Val Glu Gly Gln Glu Lys Ala His Asn Gln Asp Leu Gly Phe
 35 40 45

WO 03/012122

75/75

PCT/US02/23763

Ala Gly Ser Cys Leu Pro Arg Phe Ser Thr Met Pro Phe Ile Tyr Cys
50 55 60

Asn Ile Asn Glu Val Cys His Tyr Ala Arg Arg Asn Asp Lys Ser Tyr
65 70 75 80

Trp Leu Ser Thr Thr Ala Pro Ile Pro Met Met Pro Val Ser Gln Thr
85 90 95

Gln Ile Pro Gln Tyr Ile Ser Arg Cys Ser Val Cys Glu Ala Pro Ser
100 105 110

Gln Ala Ile Ala Val His Ser Gln Asp Ile Thr Ile Pro Gln Cys Pro
115 120 125

Leu Gly Trp Arg Ser Leu Trp Ile Gly Tyr Ser Phe Leu Met His Thr
130 135 140

Ala Ala Gly Ala Glu Gly Gly Gln Ser Leu Val Ser Pro Gly Ser
145 150 155 160

Cys Leu Glu Asp Phe Arg Ala Thr Pro Phe Ile Glu Cys Ser Gly Ala
165 170 175

Arg Gly Thr Cys His Tyr Phe Ala Asn Lys Tyr Ser Phe Trp Leu Thr
180 185 190

Thr Val Glu Glu Arg Gln Gln Phe Gly Glu Leu Pro Val Ser Glu Thr
195 200 205

Leu Lys Ala Gly Gln Leu His Thr Arg Val Ser Arg Cys Gln Val Cys
210 215 220

Met Lys Ser Leu
225

【国際公開パンフレット(コレクション)】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 13 February 2003 (13.02.2003)

PCT

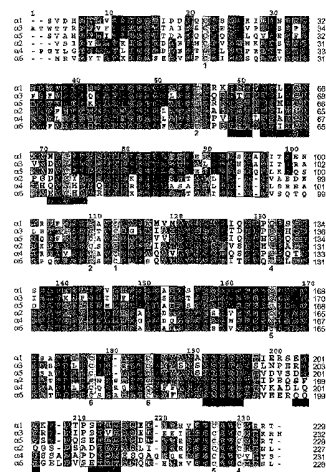
(10) International Publication Number WO 2003/012122 A2

- (51) International Patent Classification: C12Q 3901 Rainbow Blvd., Kansas City, KS 66160 NC 27703 (US).
- (21) International Application Number: PCT/US2002/023763
- (22) International Filing Date: 26 July 2002 (26.07.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
 - 60/508,523 27 July 2001 (27.07.2001) US
 - 60/551,289 29 October 2001 (29.10.2001) US
 - 60/366,854 22 March 2002 (22.03.2002) US
 - 60/385,362 3 June 2002 (03.06.2002) US
- (71) Applicant (for all designated States except US): UNIVERSITY OF KANSAS MEDICAL CENTER [US/US];
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): SUNDARAMOORTHY, Manirathnam [US/US]; S-3223 Medical Center North, 1161 Twenty-first Avenue South, Nashville, TN 37232-2372 (US). HUDSON, Billy [US/US]; Dept. of Biochemistry and Director, S-3223, Medical Center North, 1161 Twenty-first Avenue South, Nashville, TN 37232-2372, Kansas 66219 (US).
- (74) Agent: HARPER, David, S., McDonnell Boehnen Hulbert & Berghoff, 300 South Wacker Drive, Suite # 3200, Chicago, IL 60606 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,

[Continued on next page]

(54) Title: CRYSTALLIZED STRUCTURE OF TYPE IV COLLAGEN NC1 DOMAIN HEXAMER

WO 2003/012122 A2



(57) Abstract: The present invention provides a crystallized NC1 domain hexamer of Type IV collagen, and methods for making the crystal, wherein the NC1 domain hexamer is crystallized such that the three dimensional structure of the crystallized NC1 domain hexamer can be determined to a resolution of at least 3 A or better. The present invention also provides a method for designing compounds to inhibit angiogenesis, tumor growth, tumor metastasis, endothelial cell adhesion and/or proliferation, and/or basal lamina assembly, comprising analyzing the three dimensional structure of a crystallized Type IV collagen NC1 domain hexamer produced by the methods of the invention, and identifying and synthesizing compounds that target regions of the NC1 domain that have been identified by the analysis as being important for type IV collagen heterotrimer and hexamer assembly. The present invention also provides novel polypeptides designed by the rational drug design methods of the present invention, based on an analysis of the type IV collagen NC1 hexamer structure disclosed herein.

WO 2003/012122 A2



GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

Published:

— without international search report and to be republished upon receipt of that report

(48) Date of publication of this corrected version:

25 March 2004

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(15) Information about Correction:

see PCT Gazette No. 13/2004 of 25 March 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

フロントページの続き

(51) Int.Cl. ⁷	F I	テーマコード(参考)
A 6 1 P 3/10	A 6 1 P 7/06	
A 6 1 P 7/06	A 6 1 P 9/00	
A 6 1 P 9/00	A 6 1 P 9/10	
A 6 1 P 9/10	A 6 1 P 17/00	
A 6 1 P 17/00	A 6 1 P 17/02	
A 6 1 P 17/02	A 6 1 P 17/06	
A 6 1 P 17/06	A 6 1 P 19/02	
A 6 1 P 19/02	A 6 1 P 19/08	
A 6 1 P 19/08	A 6 1 P 27/02	
A 6 1 P 27/02	A 6 1 P 27/06	
A 6 1 P 27/06	A 6 1 P 29/00	
A 6 1 P 29/00	A 6 1 P 29/00	1 0 1
A 6 1 P 31/04	A 6 1 P 31/04	
A 6 1 P 31/10	A 6 1 P 31/10	
A 6 1 P 31/18	A 6 1 P 31/18	
A 6 1 P 31/22	A 6 1 P 31/22	
A 6 1 P 33/00	A 6 1 P 33/00	
A 6 1 P 33/02	A 6 1 P 33/02	
A 6 1 P 35/00	A 6 1 P 35/00	
A 6 1 P 35/02	A 6 1 P 35/02	
A 6 1 P 35/04	A 6 1 P 35/04	
A 6 1 P 37/04	A 6 1 P 37/04	
A 6 1 P 37/06	A 6 1 P 37/06	
A 6 1 P 37/08	A 6 1 P 37/08	
A 6 1 P 43/00	A 6 1 P 43/00	1 0 5
C 0 7 K 5/083	A 6 1 P 43/00	1 1 1
C 0 7 K 5/093	C 0 7 K 5/083	
C 0 7 K 5/097	C 0 7 K 5/093	
C 0 7 K 5/103	C 0 7 K 5/097	
C 0 7 K 5/107	C 0 7 K 5/103	
C 0 7 K 5/113	C 0 7 K 5/107	
C 0 7 K 7/06	C 0 7 K 5/113	
C 0 7 K 7/08	C 0 7 K 7/06	
C 0 7 K 19/00	C 0 7 K 7/08	Z T D
G 0 1 N 33/15	C 0 7 K 19/00	
G 0 1 N 33/50	G 0 1 N 33/15	Z
G 0 1 N 33/53	G 0 1 N 33/50	Z
G 0 1 N 33/566	G 0 1 N 33/53	D
	G 0 1 N 33/53	M
	G 0 1 N 33/566	
	A 6 1 K 37/02	

(31)優先権主張番号 60/385,362

(32)優先日 平成14年6月3日(2002.6.3)

(33)優先権主張国 米国(US)

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,

BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, N O, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW

(72)発明者 ハドソン、 ビリー

アメリカ合衆国 カンザス州 66219 ティエヌ 37232 - 2372 ナッシュビル ト
ウエンティ - ファースト アベニュー サウス 1161 メディカル センター ノース エス
- 3223 デパートメント オブ バイオケミストリー アンド ディレクター

Fターム(参考) 2G045 AA35 AA40 BA11 BB50 DA13 DA36

4C084 AA02 AA03 AA07 AA17 BA01 BA08 BA14 BA15 BA16 BA17
BA18 BA23 CA59 NA14 ZA332 ZA362 ZA452 ZA552 ZA662 ZA682
ZA892 ZA962 ZB082 ZB092 ZB112 ZB132 ZB152 ZB212 ZB262 ZB272
ZB332 ZB352 ZB382 ZC202 ZC212 ZC352 ZC552

4H045 AA10 AA30 BA12 BA13 BA14 BA15 BA16 BA17 BA18 BA41
CA40 EA20 FA10

专利名称(译)	IV型胶原NC1结构域六聚体的结晶结构		
公开(公告)号	JP2005504037A	公开(公告)日	2005-02-10
申请号	JP2003517296	申请日	2002-07-26
[标]申请(专利权)人(译)	堪萨斯大学医学中心的盐湖研究所		
申请(专利权)人(译)	堪萨斯大学医学中心的盐湖研究所		
[标]发明人	サンドラムーアシームイラシナム ハドソンビリー		
发明人	サンドラムーアシー、ムイラシナム ハドソン、ビリー		
IPC分类号	G01N33/50 A61K38/00 A61K45/00 A61P1/04 A61P3/02 A61P3/10 A61P7/06 A61P9/00 A61P9/10 A61P17/00 A61P17/02 A61P17/06 A61P19/02 A61P19/08 A61P27/02 A61P27/06 A61P29/00 A61P31 /04 A61P31/10 A61P31/18 A61P31/22 A61P33/00 A61P33/02 A61P35/00 A61P35/02 A61P35/04 A61P37/04 A61P37/06 A61P37/08 A61P43/00 C07K5/083 C07K5/093 C07K5/097 C07K5/103 C07K5 /107 C07K5/113 C07K7/06 C07K7/08 C07K14/78 C07K19/00 G01N33/15 G01N33/53 G01N33/566		
CPC分类号	A61P1/04 A61P3/02 A61P17/00 A61P17/02 A61P17/06 A61P19/02 A61P19/08 A61P27/02 A61P27/06 A61P29/00 A61P31/04 A61P31/10 C07K14/78 C07K2299/00		
FI分类号	C07K14/78.ZNA A61K45/00 A61P1/04 A61P3/02.101 A61P3/10 A61P7/06 A61P9/00 A61P9/10 A61P17/00 A61P17/02 A61P17/06 A61P19/02 A61P19/08 A61P27/02 A61P27/06 A61P29/00 A61P29 /00.101 A61P31/04 A61P31/10 A61P31/18 A61P31/22 A61P33/00 A61P33/02 A61P35/00 A61P35/02 A61P35/04 A61P37/04 A61P37/06 A61P37/08 A61P43/00.105 A61P43/00.111 C07K5/083 C07K5/093 C07K5/097 C07K5/103 C07K5/107 C07K5/113 C07K7/06 C07K7/08.ZTD C07K19/00 G01N33/15.Z G01N33/50.Z G01N33/53.D G01N33/53.M G01N33/566 A61K37/02		
F-TERM分类号	2G045/AA35 2G045/AA40 2G045/BA11 2G045/BB50 2G045/DA13 2G045/DA36 4C084/AA02 4C084 /AA03 4C084/AA07 4C084/AA17 4C084/BA01 4C084/BA08 4C084/BA14 4C084/BA15 4C084/BA16 4C084/BA17 4C084/BA18 4C084/BA23 4C084/CA59 4C084/NA14 4C084/ZA332 4C084/ZA362 4C084 /ZA452 4C084/ZA552 4C084/ZA662 4C084/ZA682 4C084/ZA892 4C084/ZA962 4C084/ZB082 4C084 /ZB092 4C084/ZB112 4C084/ZB132 4C084/ZB152 4C084/ZB212 4C084/ZB262 4C084/ZB272 4C084 /ZB332 4C084/ZB352 4C084/ZB382 4C084/ZC202 4C084/ZC212 4C084/ZC352 4C084/ZC552 4H045 /AA10 4H045/AA30 4H045/BA12 4H045/BA13 4H045/BA14 4H045/BA15 4H045/BA16 4H045/BA17 4H045/BA18 4H045/BA41 4H045/CA40 4H045/EA20 4H045/FA10		
代理人(译)	三好秀		
优先权	60/308523 2001-07-27 US 60/351289 2001-10-29 US 60/366854 2002-03-22 US 60/385362 2002-06-03 US		
外部链接	Espacenet		

摘要(译)

本发明提供了IV型胶原的结晶的NC1结构域六聚体，以及制备晶体的方法，其中NC1结构域六聚体被结晶使得结晶的NC1结构域六聚体的三维结构可以确定为至少3的分辨率。Å或更好。本发明还提供了一种用于设计化合物以抑制血管生成，肿瘤生长，肿瘤转移，内皮细胞粘附和/或增殖和/或基底层组装的方法，该方法包括分析结晶的IV型胶原NC1结构域六聚体的三维结构。通过本发明的方法制备并鉴定和合成靶向NC1结构域的区域化合物，所述NC1结构域的区域已经通过分析鉴定为对于IV型胶原异源三聚体和六聚体组装重要。本发明还基于对本文公开的TV胶原蛋白NC1六聚体结构的分析，提供了通过本发明的合理药物设计方法设计的新多肽。

