Cell Death and Differentiation (1999) 6, 916–930

http://www.stockton-press.co.uk/cdd

Transglutaminase crosslinking and structural studies of the human small proline rich 3 protein

Peter M Steinert[1], Eleonora Candi[1,2], Edit Tarcsa[1], Lyuben N Marekov[1], Marco Sette[3], Maurizio Paci[3], Barbara Ciani[2], Pietro Guerrieri[2] and Gerry Melino*[,2]

[1] Laboratory of Skin Biology, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland, MA 20892-2752, USA
[2] Laboratory of Biochemistry, Istituto Dermopatico dell'Immacolata, Department of Experimental Medicine, University of Tor Vergata, Rome, Italy
[3] Department of Chemical Science and Technology, University of Tor Vergata, Rome, Italy
* corresponding author: G Melino, IDI-IRCCS, c/o Laboratory of Biochemistry, Department of Experimental Medicine and Biochemical Sciences, Room D26/F153, University Tor Vergata, Via di Tor Vergata 135, 00133 Rome, Italy. tel: +39-06-20427290; fax: +39-06-20427299; e-mail: gerry.melino@uniroma2.it

Received 22.4.99; revised 6.7.99; accepted 19.7.99
Edited by P Davies

Abstract

The cell envelope (CE) is a vital structure for barrier function in terminally differentiated dead stratified squamous epithelia. It is assembled by transglutaminase (TGase) cross-linking of several proteins, including SPR3 in certain specialized epithelia normally subjected to mechanical trauma. We have expressed recombinant human SPR3 in order to study its cross-linking properties. It serves as a complete substrate for, and is cross-linked at similar efficiencies by, the three enzymes (TGases 1, 2 and 3) that are widely expressed in many epithelia. Multiple adjacent glutamines (4, 5, 16, 17, 18, 19 and 167) and lysines (6, 21, 164, 166 and 168) of only head and tail domain sequences are used for cross-linking. However, each enzyme preferentially uses certain residues on the head domain. Moreover, our *in vitro* data suggest a defined temporal order of cross-linking of SPR3 *in vivo*: It is first cross-linked by TGase 3 into short intra- and inter-chain oligomers which are later further cross-linked to the CE by TGase 1. To investigate the absence of cross-linking in the central domain (e.g. lysine in position 2 of each of the 16 repeats) we performed structural studies on recombinant SPR3 and on a synthetic peptide containing three repeats of the central domain. 2D H-1 NMR spectroscopy, TOCSY and ROESY, shows strong and medium intensity NOEs connectivities along the amino acid sequence with one weak long range NOE contact between Thr and Cys of subsequent repeats. Distance geometry computation on the basis of intensities of NOEs found generated 50 compatible structures grouped in three main families differing by the number of H-bonds. These measurements were repeated at different concentrations of trifluoroethanol (TFE)-water mixture, an α-helical promoting solvent, in order to check the stability of the conformations determined; no changes were observed up to 50% TFE in solution. Also temperature changes did not produce any variation in the ROESY spectrum in the same condition as above. The NMR and circular dichroism data strongly indicate the presence of an ordered (not α-helix nor β-sheet) highly flexible structure in the eight amino acids repetitive units of SPR3, confirming the prediction of one possible β-turn per each repeating unit. Thus, biochemical and biophysical data, strongly support SPR3 to function as a flexible cross-bridging protein to provide tensile strength or rigidity to the CE of the stratified squamous epithelia in which it is expressed.

Keywords: barrier function; cell envelope; circular dichroism; epithelia; nuclear magnetic resonance; small proline-rich proteins; skin; transglutaminase; cross-linking

Abbreviations: CD, circular dichroism; CE, cell envelope; GdHCl, guanidinium chloride; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; RMSD, root mean square deviation; ROE, rotating-frame nuclear Overhauser effect; ROESY, rotating-frame nuclear Overhauser effect spectroscopy; SPR, small proline rich (protein); TFE, 2,2,2-trifluoroethanol; TGase, transglutaminase; TOCSY, total correlation spectroscopy

Introduction

Stratified squamous epithelia undergo a complex terminal differentiation program which has many features typical of programmed cell death, including loss of the nucleus, dissolution of most 'house-keeping' cytoplasmic constituents, and cell death.[1–3] However, unlike most such programs, the dead epithelial cells are eventually lost from the epithelial surface by desquamation. Such layers of dead epithelial cells function largely to prevent water loss and as a physical barrier against the environment. A major component of this physical barrier is a specialized structure termed the cell envelope (CE) which is assembled in the living cell during terminal differentiation and is deposited as a 10–15 nm thick layer of highly insoluble protein just beneath the plasma membrane.[4–13] The CE is assembled by cross-linking several defined structural proteins by both disulfide and N^{ε}-(γ-glutamyl)lysine isopeptide bonds formed by the action of transglutaminases (TGases).[4–6] Several TGase enzymes are likely to be involved, including the membrane-bound TGase-1 enzyme and its various highly active isoforms, and the cytoplasmic TGase 2, TGase 3 and TGase X enzymes.[9,14–18] The structural proteins include: various cell peripheral and desmosomal proteins such as desmoplakin, envoplakin and periplakin; calcium binding proteins such as

annexin I; involucrin; cystatin α; elafin; various select members of the small proline rich (SPR) family; as yet uncharacterized cysteine-rich proteins in the case of the hair cuticle; trichohyalin; repetin; pancornulins; and loricrin in the epidermis (see [7–10] for references). To date, only anecdotal information is available on the structures of these protein components.

Of these, the SPR group is the most complex, consisting in human and mouse of 11–12 members divided into three families, SPR1 (two members), SPR2 (8–11 members), and SPR3 (one member) (see [19–22] for reviews). All members are built according to a common plan of amino (head) and carboxy (tail) domains containing several adjacent glutamine and lysine residues, which span a central domain composed of a series of peptide repeats of 8–9 residues which are highly enriched in prolines. The precise sequence of these repeats permits distinction into the three families. The numbers of repeats varies widely both between members of the family and between species: for example, human SPR1 proteins contain six repeats while mouse contain 13 or 14; all human SPR2 members contain three repeats, while mouse proteins contain 3.5–9 repeats; and the number of repeats in SPR3 varies from about 16–30 in different species. In addition, various individual members of SPR proteins are expressed in widely differing amounts in different epithelia, and their expression is generally upregulated in response to epithelial injury or disease.

We have deduced from sequencing analyses of cross-linked peptides obtained from CEs isolated from human epidermis and mouse forestomach epithelia that the SPRs serve pivotal roles in the CE structure. We have found that glutamine and lysine residues on only head and tail domains are used in cross-linking *in vivo* and *in vitro*.[8–10,16,17,21] The data imply that the SPRs serve as cross-bridging proteins by adjoining themselves or other proteins such as loricrin, involucrin, desmosomal proteins, etc., by use of multiple adjacent residues of the end domains.[21] Furthermore, we have found a correlation between the amount of SPR proteins used in CEs and the presumed physical requirements for mechanical strength and toughness of the epithelium.[10] From this, we have concluded that the SPRs serve as biomechanical modifiers of the physical properties of the CE structures in order to fulfill the particular requirements of different epithelia to withstand physical trauma.[10]

Recent biochemical experiments have shown that for the SPR1[16] and SPR2[17] proteins, each TGase enzyme preferentially cross-links certain glutamine and lysine residues with high specificity, from which we could conclude that multiple enzymes are required to cross-link them *in vivo*. Moreover, in the case of SPR1 proteins, the data imply an obligatory temporal order to this process: first cross-linking by the cytosolic TGase 3 into short oligomers, which in turn are later cross-linked by the membrane-associated TGase 1 enzyme into large polymers onto the CE barrier.[16]

To date, however, there are no equivalent biochemical data for the SPR3 protein. In this study, we have extended the *in vitro* experiments to bacterially expressed human SPR3. We show here that, unlike SPR1 and SPR2, it is approximately equally cross-linked by three TGases commonly expressed in such epithelia, using multiple different residues only on the head and tail domains but with limited sequence specificity. Our data suggest that the SPR3 protein is more aggressively cross-linked than the other classes of SPR proteins, which seems consistent with the hypothesis that it is an even more important effector of the biomechanical properties of the epithelia in which it is expressed. In addition, we have used CD and NMR to examine the conformational features of the uncrosslinked central repeats of this protein. These studies have revealed a strong tendency to remain unstructured, which indicates a marked degree of flexibility.

Results and Discussion

The SPR3 protein is widely expressed in a variety of terminally differentiated, dead stratified squamous epithelia that are normally subjected to significant mechanical trauma, such as the esophagus, filiform ridges of the tongue, inner root sheath of the hair follicle, and rodent forestomach.[10,22–25] The purpose of this study is to provide basic biochemical information on the cross-linking properties of SPR3 in an attempt to correlate its presumed functional role in these tissues.

Expression and purification of the recombinant human SPR3 protein

Following expression in bacteria using the pET11a system, the recombinant human SPR3 protein was purified to homogeneity following removal of bacterial proteins by precipitation in 25 mM sodium acetate buffer and chromotography on a Mono-S FPLC column.[16–17] It runs on SDS–PAGE with a substantially higher apparent size than its true molecular weight of 19 kDa (Figure 1, lanes 1 and 2 of insert).

Isolation of native mouse SPR3 protein

The mouse SPR3 protein was readily purified from tissue homogenates because of its high solubility in citrate buffer. This protein contains about 24.5 repeats in its central domain, compared to 16 for human, so it has a molecular size of 26 kDa and runs with an apparent molecular weight of about 36 kDa (Figure 1, lanes 3 and 4 of insert). Notably also, there were minor amounts of apparent oligomers that were soluble in the citrate buffer. The yield was about 0.2 mg/g (wet weight of tissue). Assuming SPR3 constitutes about 2% (molar basis) of the CE,[10] we estimate the forestomach tissue should contain about 0.5 mg/g wet weight. Thus, a considerable amount of SPR3 remains in a soluble form, some of which is already oligomerized, which thereby affords an important clue on its temporal incorporation into the CE structure.

Three epidermal TGases use recombinant human SPR3 as a complete substrate but in different extents

Three TGase enzymes which are expressed in stratified squamous epithelia were used to cross-link *in vitro* the recombinant human and native mouse SPR3 proteins. In

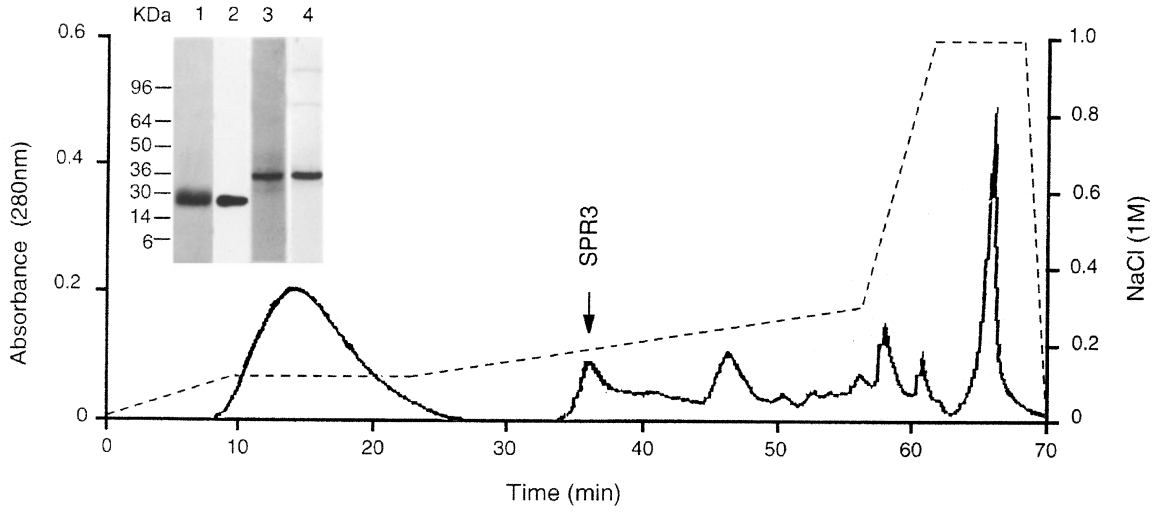


Figure 1 Isolation and purification of recombinant human and native mouse forestomach SPR3 proteins. The SPR-enriched citrate extracts were chromatographed on the Mono-S column, from which the highly purified proteins were recovered in the peak shown. Inset shows SDS–PAGE gels and Western blots of the purified proteins. Lanes 1 and 2, recombinant human SPR3; lanes 3 and 4, native mouse SPR3 protein; lanes 1 and 3, SDS gels developed with Coomassie stain; lanes 2 and 4, Western blots developed using the broadly-reacting SPR1/3 antibody.[20] kDa, molecular mass sizes of standards are shown. Note the presence of apparent dimers/trimers of mouse SPR3 in lane 4

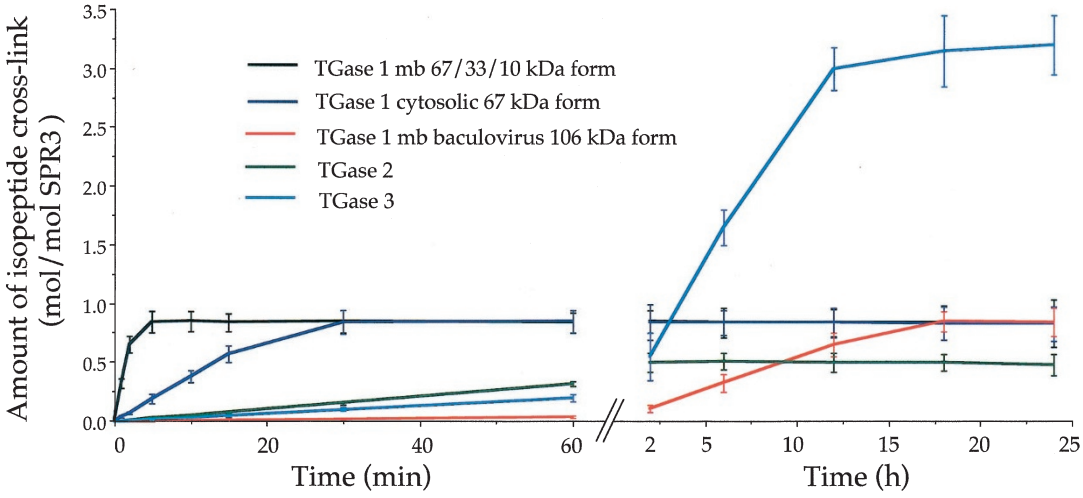


Figure 2 Isolation of isodipeptide reveals that three TGases use the recombinant SPR 3 to different degrees. Equimolar amounts (700 nM) of the five isoforms of the TGase 1 enzyme as shown, and TGases 2 and 3, were used to cross-link 1 nmol of SPR3 for varying times as indicated. Aliquots were removed and digested to completion to release the free isodipeptide, which was then quantitated by amino acid analysis. The data are the averages (±S.D.) of 2–3 separate experiments

order to make direct comparisons of their reactions, we used the same molar amount (typically 700 nM) of each of several isoforms of TGase1, the only known active form of TGase 2, and the activated form of TGase 3.

First, we determined that each enzyme inserted significant amounts of isodipeptide cross-link into SPR3, which means that the enzymes use it as a complete substrate, in the sense that this substrate provides both the donor glutamine and acceptor lysine residues. However, the extent and rates of reactions varied widely. The TGase 2 enzyme inserted about 0.5 mol/mol cross-link; the activated TGase 3 enzyme inserted the most (3.2 mol/mol); and a maximal amount of 0.8–0.9 mol/mol was inserted by each of four TGase 1 isoforms employed, but at different rates (Figure 2). The reaction was completed within 5 min with the highest specific activity 67/33/10 kDa complex form recovered from the membrane fraction of NHEK cells, within 30 min by the cytosolic 67 kDa forms from NHEK cells, but 15–18 h was required for the lowest specific activity membrane-bound form expressed in baculovirus cells or in NHEK cells (data not shown). Previous work has shown that there is about a 200-fold difference in specific activity between the intact and activated forms,[26] which is reflected in these observed

reaction rates. In general, the times required for reaction completion for each of the TGase 1 isoforms mirror accurately their specific activities.

Second, we resolved the cross-linking reactions by SDS–PAGE and then performed autoradiography or Western blotting. In order to obtain quantitative information, we cut out ^{35}S-labeled bands from the SDS gels. With each of the TGase 1 isoforms, >95% of the SPR3 protein remained as a monomer (Figure 3A for 67/33/10 kDa complex recovered from the NHEK membrane fraction, cytosolic 67 kDa form, and baculovirus intact form; data for other isoforms not shown); only traces of protein were oligomerized into apparent dimers. Similarly, <5% of the SPR3 protein was oligomerized by the TGase 2 enzyme (Figure 3B). However, Western blotting analyses with a TGase 2 specific antibody revealed certain diffuse high molecular bands were due to autocatalytic cross-linking to TGase 2 itself (data not shown), a phenomenon described previously for this[16,17,27] and the related factor XIIIa[28] TGase enzymes. On the other hand, the TGase 3 enzyme cross-linked the SPR3 protein readily so that 30% of the formed short oligomeric products (Figure 3C). Although we did not use labeled protein, very similar data were obtained for the native mouse SPR3 protein as observed by Coomassie staining (not shown). These data thus reveal rather different reaction processes between the three TGases. However, in all cases, the monomer SPR3 band appeared to have migrated at a faster rate than in the EDTA controls. Based on Coomassie staining, Western blots, and amino acid analyses (data not shown), this shift was likely due to intrachain cross-linking. A similar observation has been made previously in the *in vitro* cross-linking of loricrin.[14]

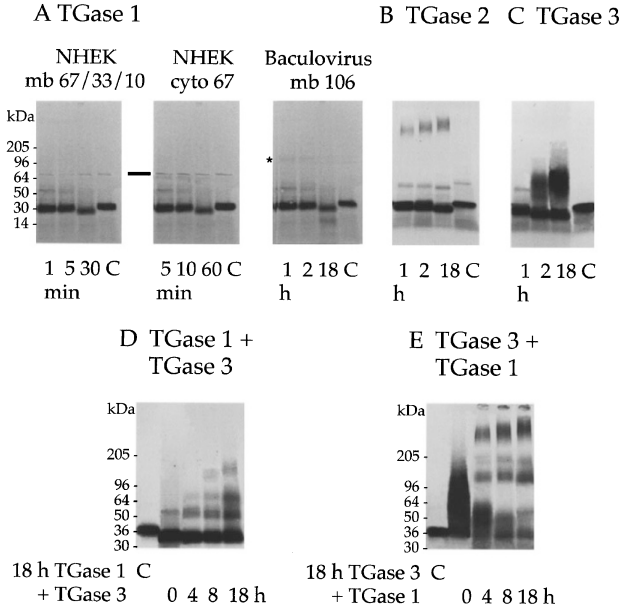


Figure 3 Cross-linking of recombinant human and native mouse SPR3 proteins by three TGases *in vitro*. (**A–C**) Recombinant ^{35}S-labeled human SPR3 protein. In (**A**) 700 nM amounts two different forms of the TGase 1 enzyme recovered from normal human epidermal keratinocytes and the baculovirus full-length form were used; in (**B, C**) 700 nM TGases 2 and 3 were used. Times for cross-linking varied with the enzymes as identified in Figure 2. The bar marks the position of the 67 kDa form of the enzymes; the asterisk marks the full-length enzyme. In (**D, E**) native mouse forestomach SPR3. Only the baculovirus TGase 1 enzyme was used, and the proteins were monitored by Western blotting with a broadly-reacting SPR1/3 antibody

The most likely explanation for this is that the enzymes inserted extensive intrachain cross-links. However, this observation makes little biological sense, as the available data shows that the SPR3 protein forms extensive interchain linkages instead[10] in its proposed cross-bridging role in CE structures. Therefore, we performed double reactions in which we first cross-linked to completion with either the TGase1 or 3 enzyme, and then reacted those products with either the TGase 3 or 1 enzyme, respectively. These experiments were monitored by Western blotting. Using native mouse SPR3 protein, the TGase 3 enzyme permitted little additional cross-linking of the initial TGase 1 cross-linking reaction, as there was little change in the mobility of bands (Figure 3D), and the amount of isodipeptide incorporated increased from 0.8 to 1 mol/mol. However, the TGase 1 enzyme performed much additional cross-linking of the initial TGase 3 products into large polymers, as evident from the observation that most of the material could no longer enter the gel (Figure 3E). Further, the total amount of isodipeptide cross-link increased from 3.2 to 5 mol/mol. Similar data were also obtained for the ^{35}S-labeled recombinant human SPR3 wherein we found that >70% of the ^{35}S-label did not enter the gel (not shown). These *in vitro* data support the view that there is a well-defined preferred temporal order for the cross-linking of the SPR3 protein *in vivo*.

Finally, these experiments also document that the three TGases use the native mouse and recombinant human SPR3 proteins in the same way. While the proteins differ in the numbers of the central domain peptide repeats, the distributions of glutamine and lysine residues on the head and tail domains are almost identical. These data therefore offer strong indirect support for the notion that the recombinant human protein had been properly refolded following expression and isolation, and is thus a valid source of the SPR3 protein for further biochemical experiments.

Kinetics of cross-linking of recombinant SPR3 by TGases 1, 2 and 3

These were determined using the recombinant full-length TGase 1 enzyme expressed in baculovirus, the guinea-pig TGase 2 enzyme, and dispase-activated TGase 3 enzyme (Table 1), and saturating amounts of putrescine to suppress oligomerization of the recombinant SPR3 protein by putrescine. Note that the kinetic values listed are in fact the averages for the several glutamine residues that are utilized by the enzymes (see below). The data show that there are no significant differences in the kinetic efficiencies between the three TGases. This observation is in marked contrast to previous data for recombinant SPR1,[16] trichohyalin[15] and loricrin[14] (Table 1), where the TGase 3 enzyme was the most efficient. On the other hand, the kinetic efficiency values for the three enzymes with SPR3 are two- to eightfold higher than

Table 1 Kinetic parameters for cross-linking of the recombinant SPR3 protein by TGases 1, 2 and 3

	k_{cat} min^{-1}	K_M μM	k_{cat}/K_M $min^{-1}.\mu M$	V_{max} $pmol.min^{-1}$
Recombinant human SPR3				
TGase 1	43.6±9.6	18.7±8.3	2.3±0.8	16.3±0.9
TGase 2	30.8±7.7	18.1±8.0	1.7±0.5	1.7±0.8
TGase 3	20.7±6.8	10.5±3.3	2.0±0.7	8.8±0.2
Recombinant human loricrin				
TGase 1	5.0±0.8	16.9±3.2	0.3±0.1	39.8±7.9
TGase 2	4.5±0.9	16.0±3.2	0.3±0.1	2.0±0.4
TGase 3	5.7±1.1	5.0±1.4	1.1±0.2	45.6±9.1
Recombinant human SPR1				
TGase 1	6.2±0.2	10.4±1.1	0.6±0.2	
TGase 2	5.9±1.3	19.9±1.9	0.3±0.2	
TGase 3	6.5±1.3	5.4±0.4	1.1±0.2	
Succinylated casein				
TGase 1	1.7±0.2	57±5	0.03±0.005	
TGase 3	2.2±0.5	31±4	0.07±0.01	

The full-length baculovirus TGase 1 enzyme[51] was used for the present SPR3 experiments. The data for loricrin[14], SPR1[16] and succinylated casein[60] are shown for the full-length TGase 1 enzyme from normal human epidermal keratinocytes[26] or recombinant enzyme expressed in bacteria[60]

TGase 3 for these other substrate proteins, which together, suggests that SPR3 is far more efficiently cross-linked by all enzymes. In contrast, the generic TGase substrate, succinylated casein, is much less efficiently used.

Amino acid sequencing analyses of cross-linking reactions by the TGase 1, 2 and 3 enzymes *in vitro*

In view of the observed different amounts of cross-linking (Figure 2), we performed additional experiments to identify which glutamine and lysine residues were utilized by each of the three TGases. Samples of uncross-linked and cross-linked SPR3 protein were digested to completion with trypsin. The peptides were resolved by HPLC and sequencing was performed on reduced or shifted peaks to obtain specific quantitative information.

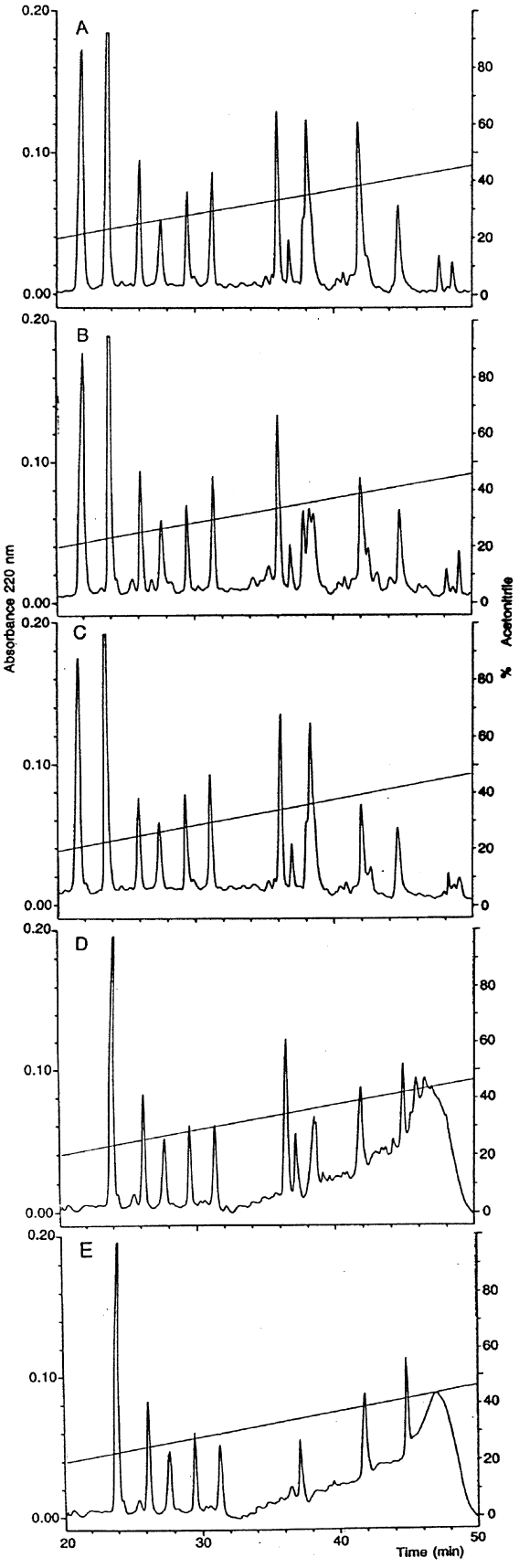
We used two forms of the TGase 1 enzyme. In the case of the baculovirus expressed full-length form (Figure 4B), most of the peptide peaks were unchanged in position, although some were reduced in amount. The peptide which eluted at 22 min and contained residues 1–6 was unchanged. The peptide eluted at 38 min containing residues 7–21 was reduced by about 50%. The peptide eluted at 42 min containing residues 22–37 was reduced by about 20% because Lys21 was partially used in cross-linking. The two short tryptic peptides corresponding to residues 165–166 and 167–168 near the carboxy terminus could not be resolved by this HPLC system. However, 30–50% of their expected molar amounts was identified in the minor peaks eluted after 46 min and were cross-linked to sequences containing residues 7–21 or 22–37. The several tryptic peptides corresponding to the central domain were not significantly reduced and were not seen cross-linked to other sequences. The total extent of involvement of Gln and Lys residues (about 0.8 mol/mol) (Figure 4F, uppermost numbers), corresponded closely to the amount of isolated isodipeptide (Figure 1). Essentially identical sequencing and quantitative data were obtained for the high specific activity 67/33/10 kDa form of TGase 1 isolated from the membrane fraction of NHEK cells (not shown). Together with the data from Figures 1 and 2, we can conclude that the various forms of the TGase 1 enzyme treat this substrate in solution assays in the same way. Notably, the TGase 1 enzymes essentially exclusively utilized only head domain B sequences (residues 14–25) for cross-linking to tail domain sequences.

With the TGase 2 enzyme, essentially only two resolvable tryptic peptides were changed (Figure 4C): the peptide eluted at 38 min (residues 7–21) was reduced by about 10%, and that eluted at 42 min (residues 22–37) was reduced by about 50%. These sequences were recovered in the minor peaks eluted after 46 min cross-linked to the two terminal tryptic peptides. Again, the total recovery of isodipeptide (Figure 1) in this reaction mirrored the extent of involvement of Gln and Lys residues (about 0.45 mol/mol) (Figure 4F). In this case, only three residues in the head B domain were used for cross-linking to tail domain residues.

In the case of the reaction with the activated TGase 3 enzyme, many tryptic peptides were changed and a broad poorly resolved region was seen late in the HPLC profile (Figure 4D). Notably, peptides containing residues 1–6 (eluted at 22 min) and 7–21 (38 min) were largely lost, but that containing residues 22–37 (eluted at 42 min) was reduced only about 50%, mostly due to the utilization of the preceding Lys21. The peptide eluted at 36 min (containing residues 148–164) was diminished by about 5% due to utilization of Lys164 with Gln19. Peptides eluted between 26–37 min and at 45 min containing central domain sequences were reduced by about 5% each indicating that a minor amount of their Lys residues was used for cross-linking. Likewise, the peptide eluted at 24 min comprising five exact repeats of the sequence VPEPGCTK (residues 60–99) was reduced 15% following cross-linking; perhaps each of the five lysine residues was utilized to a minor extent ($<5\%$) in cross-linking (Figure 4F). By sequencing 1 min time aliquots across the broad peak from 44–50 min, it was possible to account for all of the diminished peptides of the head domain cross-linked to glutamine and lysine residues on the end of the tail domain, even though at least three separate sequencing were running simultaneously. Altogether, this region could account for about 3 of 3.1 mol of isodipeptide residues (Figure 4F). As can be seen, about 70% of the cross-linking involved head B domain sequences, but a significant proportion involved the head A domain.

Next, we performed these experiments on SPR3 protein that had been first cross-linked with the TGase 1 enzyme and then followed by the TGase 3 enzyme. In this case, there was only a minor change in the HPLC profile in comparison to the TGase 1 enzymes alone (profile not shown), reflected by somewhat increased usage of the same glutamine and lysine residues (summarized in Figure 4F), although the peptide eluted 22 min (residues 1–6) was reduced by about 20%.

In contrast, when we examined the tryptic peptides of the reverse experiment, cross-linking by TGase 3 followed



F

Head Domain

	S	S	Y	Q	Q	K	Q	T	F	T[10]	P	P	P
TGase 1				0	0	0	0						
TGase 2				0	0	0	0						
TGase 3				40	40	5	5						
TGase 1+3					5	10	5						
TGase 3+1				40	40	10	5						

	Q	L	Q	Q	Q	Q	V[20]	K	Q	P	S	Q[25]
TGase 1			10	10	20	20		20				
TGase 2						10			20			30
TGase 3			30	40	50	60		60				
TGase 1+3			10	10	20	30		30				
TGase 3+1			50	90	90	90		100	10			

Central Domain — *Repeats*

P	P	P	Q	E	I	F	V[33]	*1°*
P	T	T	K	E	P	C	H[41]	*2°*
S	K	V	P	Q	P	G	N[49]	*3°*
T	K	I	P	E	P	G	C[57]	*4°*
T	K	V	P	E	P	G	C[65]	*5°*
T	K	V	P	E	P	G	C[73]	*6°*
T	K	V	P	E	P	G	C[81]	*7°*
T	K	V	P	E	P	G	C[89]	*8°*
T	K	V	P	E	P	G	C[97]	*9°*
T	K	V	P	E	P	G	Y[105]	*10°*
T	K	V	P	E	P	G	S[113]	*11°*
I	K	V	P	D	Q	G	F[121]	*12°*
I	K	F	P	E	P	G	A[129]	*13°*
I	K	V	P	E	Q	G	Y[137]	*14°*
T	K	V	P	V	P	G	Y[145]	*15°*
T	K	L	P	E	P	C	P[153]	*16°*

TGase 1	0		0	0	0	*(for each repeat)*
TGase 2	0		0	0	0	*(for each repeat)*
TGase 3	5		0	0	0	*(for each repeat)*
TGase 1+3	0		0	0	0	*(for each repeat)*
TGase 3+1	5		0	0	0	*(for each repeat)*

Tail Domain

	S	T	V	T	P	G	P[160]	
	A	Q	Q	K	T	K	Q	K[168]
TGase 1						30	20	30
TGase 2						20	10	20
TGase 3				5		90	90	90
TGase 1+3						40	30	40
TGase 3+1		30	30	40		100	100	100

Figure 4 Identification and quantitation of glutamine and lysine residues used for cross-linking of recombinant SPR3 protein by three TGases *in vitro*. Tryptic peptides before (**A**) or after cross-linking with the baculovirus TGase 1 (**B**), TGase 2 (**C**), activated TGase 3 (**D**), or TGase 3 followed by TGase 1 (**E**). The products were resolved by HPLC. In **F** is the primary sequence of human SPR3 protein to summarize the per cent utilization of the glutamine and lysine residues as estimated following sequencing of the lost/new peptide peaks. This protein features 16 repeats of eight residues of the central domain. The conserved lysine residue in the second repeat position may be slightly used for crosslinking. The sequences underlined delineate the peptide of three repeats which was synthetized for CD and NMR studies

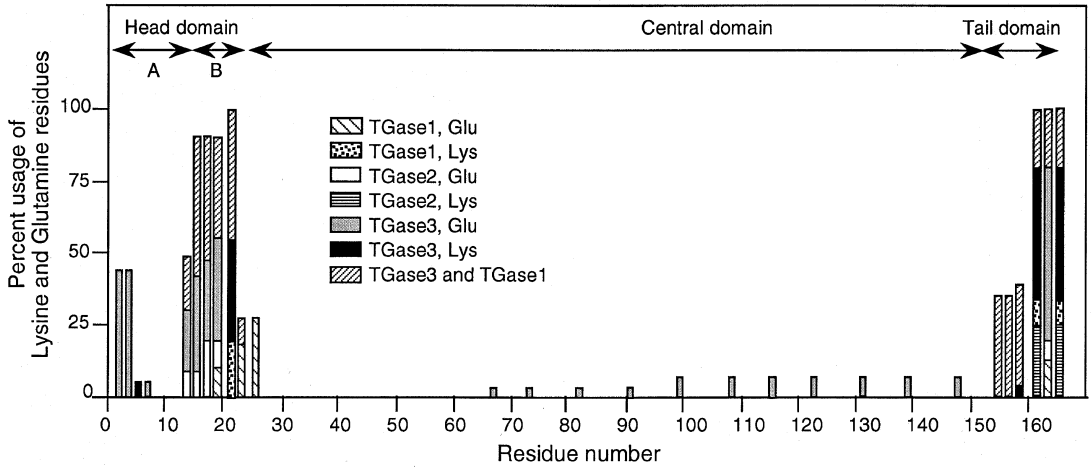


Figure 5 Summary of utilization of glutamine and lysine residues used for cross-linking of recombinant SPR3 protein by three TGases *in vitro*. The sites of the additional (2 mol/mol) of cross-link inserted by the TGase 1 enzyme after complete TGase 3 reaction are also indicated

by TGase 1, there were significant changes from the SPR3 reaction alone (Figure 4E): peptides eluted at 22, 36 and 38 min were completely lost; and a much broader less resolved profile appeared at the end of the HPLC run. The peptide eluted at 42 min, and the several peptides containing the central domain sequences, were essentially unchanged from the reaction with TGase 3 alone. Sequencing analyses revealed that most of the additional cross-linking (about 1.8 mol/mol) involved near maximal utilization of several of the head B domain glutamine and lysine residues to tail domain residues (summarized in lowest line of Figure 4F).

Secondary structure features of SPR3

The biochemical data reported in the previous paragraphs and summarized in Figure 5 show a very different behavior of lysines present in the head and tail domains *versus* the lysines present in the central domain repeats. In order to evaluate the structural properties of the rod domain, thus outlining a possible structure-function relationship, we performed both a structure prediction study and an experimental evaluation by circular dicroism (CD) and ^{1}H-nuclear magnetic resonance (NMR) on the recombinant SPR3 protein or three repeats from the central domain.

Prediction studies indicate that the dominant structural feature in the central repeating domain SPR3 protein and the synthetic peptide (see Figure 4F) is the presence of a β-turn conformation. Repetitive β-turns are thus predicted for the central repeated and rich in proline domain, the probabilities depending on the tetrapeptide sequences (see Table 2).

Table 2

Region	Consensus	Probability
EPGC	TKVPEPGC	4.10×10^{-4}
EPGY	T/IKVPEPGY/F	4.00×10^{-4}
EPGF	same	2.08×10^{-4}

The absence of predicted α-helical or β-sheet structures in the repeated sequences can be accounted for by the high proline content. A minimum of five favorable residues are required to form an α-helix and six for a β-sheet, proline destabilizing either structure. A polyproline II structure is also unlikely to form, since consecutive proline residues or proline residues are present at every third position along the polypeptide chain.[31,32]

Circular dichroism spectroscopy

CD spectra of human protein SPR3 and peptide show a very similar profile (Figure 6A). The different value of molar ellipticity at 200 nm for SPR3 peptide compared to that of the protein is probably caused by the different numbers of repeats[33,34] as found for SPR1 proteins.[16] The far-UV spectrum is dominated by transitions associated with the amide groups of the peptide backbone. The most striking feature in this region is the absence of characteristic shapes with either an α-helical or β-sheet conformation, supporting the indications of the structure prediction. Indeed, the spectrum most closely resembles that of a random coil.

However, several lines of evidence suggest the presence of some ordered structure. First, significant differences are apparent when the spectrum of human SPR3 protein and peptide in water is compared with that obtained in solution with guanidinium chloride (GdHCl) up to 5 M (Figure 6B). The spectrum in water shows a greater ellipticity value around 225 nm than the spectrum in 5 M GdHCl, consistent with the presence of ordered structures. The protein denaturant induces marked unfolding at 5 M, and is partially although not completely restored when the denaturant is removed (Figure 6B).

Second, similar data were obtained with temperature changes. On increasing the temperature, the spectrum of human SPR3 protein shows significant unfolding at 50°C, but the spectrum is not fully restored after the transition 23-50-23°C (Figure 6C). Raising the temperature from 23 to 70°C (Figure 6D) for the SPR3 peptide also decreases

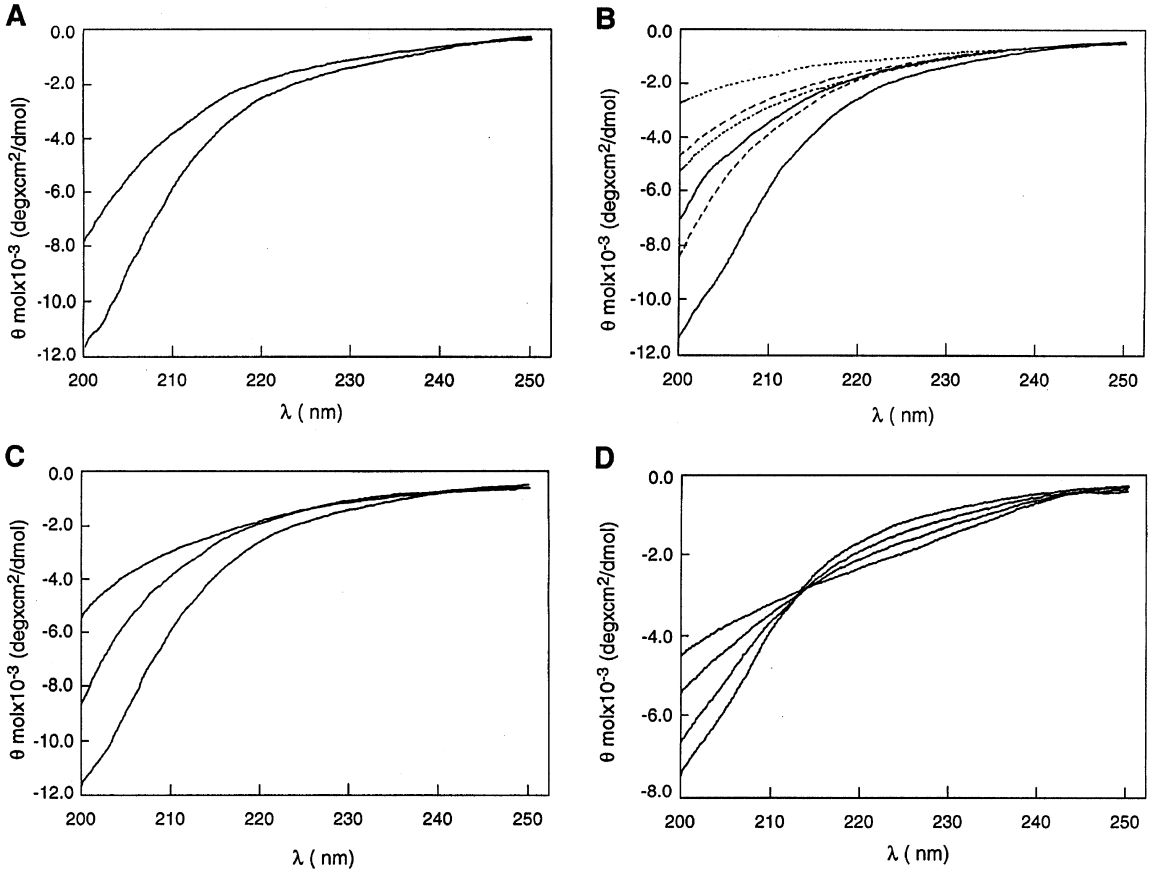


Figure 6 (**A**) CD spectra of recombinant human SPR3 protein (1.5 mg/ml; lower line) and the 24 residue SPR 3 peptide (1 mM) corresponding to three repeats (upper line). (**B**) CD spectra of human SPR3 protein (lower solid lane) and the 24 residues SPR3 peptide (upper solid line) in 5 M GdHCl (lower and upper dotted lines, respectively) or after a 0-5-0 M transition (lower and upper dashed lines, respectively). (**C**) CD spectra of SPR 3 protein at 20°C (lower line); after a 20-50-20°C transition (middle line); and at 50°C (upper line). (**D**) CD spectra of the peptide measured at 20, 40, 50 or 70°C (from lower to upper lines)

ellipticity at 200 nm, thus indicating a loss of secondary structure by disrupting the stabilizing interactions. The spectrum shows an increase in absorption around 225 nm with the increase in temperature (Figure 6D). The partial thermal reversibility and the presence of an isodichroic point at 212 nm suggest that one equilibrium with at least two conformers are observable by CD. Higher temperatures shift the equilibrium between the conformers. This can be explained as a result of the stabilization by hydrophobic interactions of the secondary structures which are favored at high temperatures.[35] These temperature effects suggest partial denaturation or may reflect changes in the conformation of the proline peptide bonds from trans to cis as previously suggested.[36] Note that neither the intact SPR3 protein nor peptide precipitated in solution at 60°C.

Experiments were also performed upon addition of TFE (2,2,2-trifluoroethanol) in order to study the conformation of the peptide (Figure 7A). TFE enhances the intramolecular interactions such as hydrogen-bonding and electrostatic salt bridges in proteins and peptides. Increasing the concentration of TFE from 10% to 70% (v/v) in water resulted in similar changes in the far- and near-UV regions as those observed on heating. There was a decrease in intensity and a shift in the peak wavelength from 200 to 203–204 nm, and an increase of the shoulder at 225 nm (Figure 7A). Also, an isodichroic point is evident at 212 nm, suggestive of the existence of a conformational equilibrium.

A far-UV difference spectrum was obtained by subtracting the spectrum in aqueous 30% TFE (v/v) from that in 70% TFE (v/v) (Figure 7B). A similar spectrum was obtained by subtracting the trace at 23°C from the 70°C in Figure 6D (not shown). The differential CD spectrum could correspond to a class-B β-turn spectrum according to the classification scheme of Woody,[37] with a minimum around 225 nm and a maximum in the region 200–205 nm. This class-B spectrum is the most common class associated with a β-turn conformation.[38] The increase in β-turn conformation content with both increasing temperature and increasing hydrophobicity of the solvent would indicate that the conformation rich in β-turn minimizes both hydrophobic interactions and hydrogen-bonding.

The increase in order with increase in temperature would suggest the SPR3 repeats have the same properties of elastic biopolymers like elastin, in which the repeats assemble in β-spirals at high temperatures.[39]

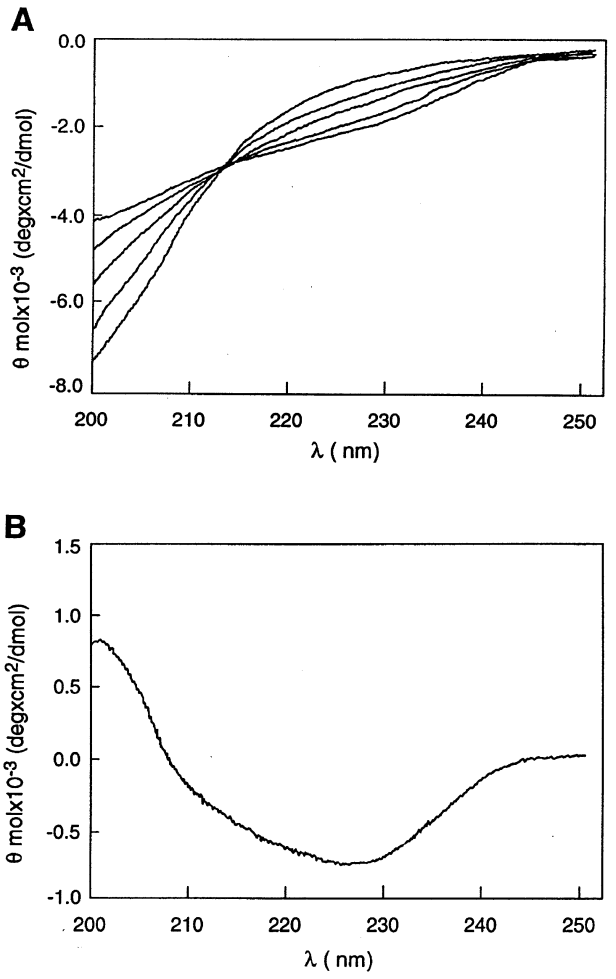


Figure 7 (**A**) Effect of TFE on the CD spectra of the SPR3 peptide using TFE concentration of 0, 10, 30, 50 and 70% (from lower to upper line). (**B**) CD spectrum of SPR3 peptide obtained subtracting the one in 30% TFE from the 70% TFE

NMR spectroscopy

The total correlation spectroscopy (TOCSY) and rotating-frame Overhauser effect spectroscopy (ROESY) were used to obtain the sequential assignment of the protons belonging to the constituent amino acids; ROESY data were also used to estimate inter-proton distances. Two dimensional NMR studies on the SPR3 peptide show a substantially extended conformation for the molecule. The internal dynamics of the molecule are such that the effective correlation times for most of the pair-wise interproton interactions are too fast to permit analyses of the nuclear Overhauser effect cross peaks in the laboratory-frame by NOESY experiments. However, several inter-residue interactions have been identified in the ROESY spectra.

Figure 8A shows the fingerprint region (Hα-NH) of the ROESY spectrum of the repeats of the SPR3 peptide. The strong sequential Hα(i)-NH(i+1) and the absence of NH(i)-NH(i+1) NOEs indicate an extended conformation. Further, the absence of most of the inter-proton interactions in the NOESY indicates that the extended structure is substantially flexible. Due to the repetitive nature of the peptide sequences, some of the resonances in the proton NMR spectra were partly overlapped. Assignments were achieved by acquiring TOCSY and ROESY at different temperatures and at different concentrations of deuterated TFE.

For the spin system assignment, identification of the amino acid spin system was achieved by reference to published random-coil chemical shift tables.[40]

Other spin systems were connected through interresidue Hαi-NHi+1 ROEs. This method breaks down at proline residues because of the absence of the backbone amide proton, a severe limitation in the assignment of these proline-rich peptides. However, the sequential Hα(i)-ProHδ(i+1) ROE connectivities were useful to identify the sequential connections. Proton assignments of the SPR3 peptide in water at pH 5.3 at 298°C were obtained and are presented in Table 3A. Figure 8B summarizes the sequential ROEs data for SPR3 peptide.

The only ROEs seen were intraresidue, sequential and short-range. Spin system of residue number 1 was not recognized in the spectra probably due to its fast conformational interconversion, but a NOE cross peak was found connecting its CHβ to the NH of residue 2. The residue number 2 was recognized as for the starting position for the sequential assignment because it is usually the one with the lowest field NH.[41] In this way, assignment of the first repeat was achieved. Central repeated sequences were assigned on the basis of the over imposition of some signals, such as Thr10/18 and Val12/20 (that is, they sample the same environment). It was also noticed that Hα of prolines preceding glycines were more downfield than those of prolines preceding glutamic acid. This afforded the only real discrimination between prolines, e.g., only those inside each repeated unit which showed significantly different chemical shift values.

All Glu22 cross peaks appeared very weak in intensity because of the higher mobility of the C-terminal tail. Glu6 showed a weak Hα-NH ROE but strong CHβ,β'-NH; Glu14 showed strong internal ROEs plus the presence of close weaker cross peaks probably due to the contribution of another conformer. Since the Hα of the residue Xaa (in the sequence Yaa-Pro-Xaa-Pro) is closer than the amide proton to the peptide bond Xaa-Pro by one covalent bond, it is likely to experience greater changes in chemical shift between the conformers, resulting from that proline's isomerization than the amide proton for residue Xaa. Similarly, the amide proton of Xaa would show a greater difference in chemical shift than the Hα if a proline preceding it were isomerizing.[42] For this reason the amide bond Val12Pro13 (in the sequence Val12Pro13Glu14Pro15) (Xaa-Pro) could be involved in a trans/cis isomerization. Unfortunately, proline signals were too overimposed to discriminate between the conformers. Inversion of Hα/Hβ chemical shifts values for all threonines was also observed.

The temperature dependences of the amide proton chemical shifts were measured to provide information about hydrogen bonding of the NH groups.[43,44] Table 3B

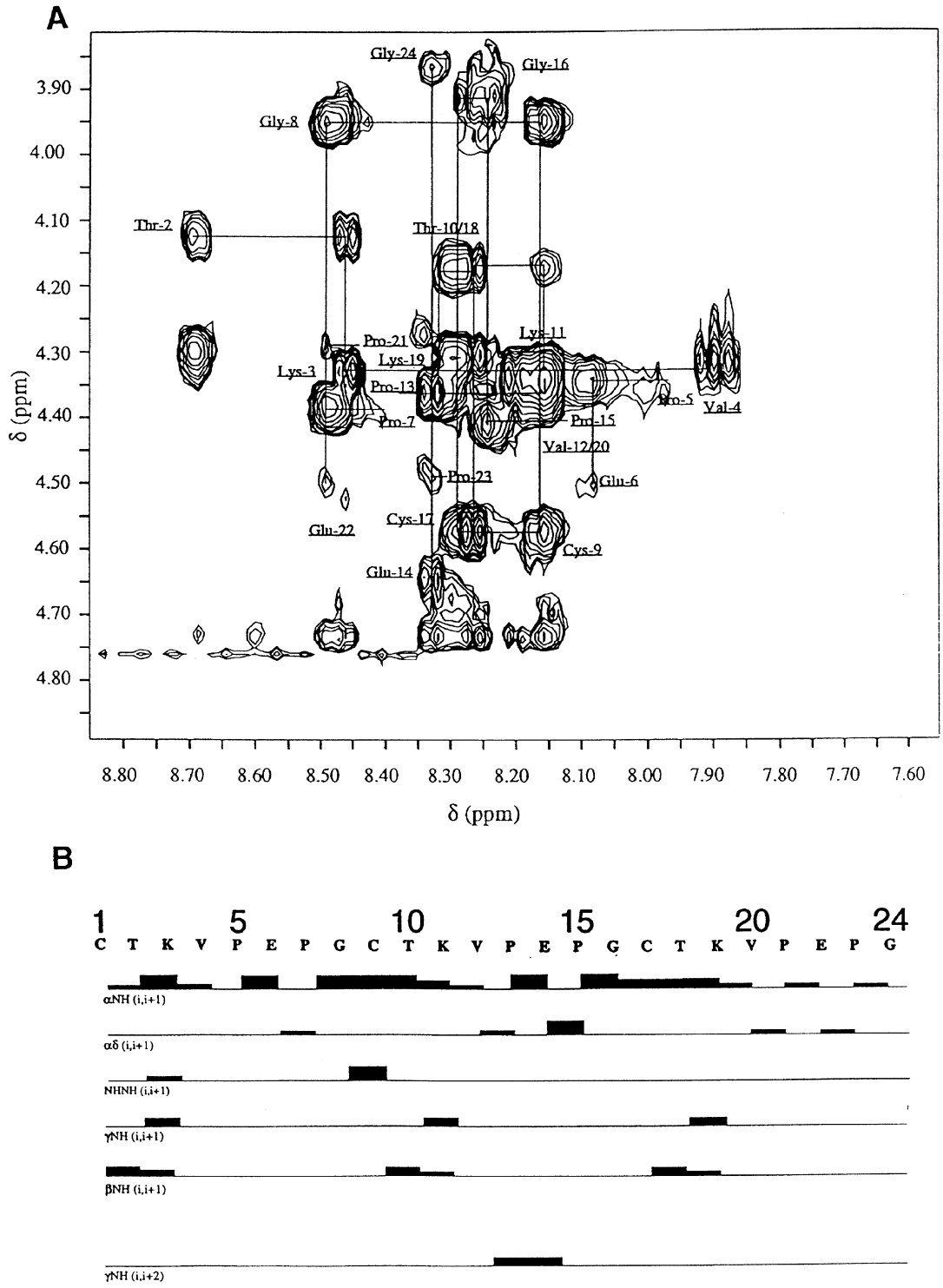


Figure 8 (**A**) ROESY spectrum (at 300 ms, 298°C) of 3 mM SPR3 peptide in phosphate buffer (pH 5.3) containing 15 mM dithiothreitol, fingerprint region. Hα-NH sequential connectivities are indicated by lines. (**B**) NOEs contacts in ROESY spectra of SPR3 peptide in TFE-aqueous solution

lists the temperature shifts of amide protons in water which show that some residues are likely to be protected from the solvent exchange.

Increasing the temperature ^{3}JNHa couplings of Gly16 appear slightly affected as it could be recognized by a little difference between Ha and Ha′ protons (data not shown). Some ROEs seemed to resemble the J coupling of the connecting NH(i+1) residue peak.

Addition of deuterated TFE made two changes on ROESY spectra: most cross peaks disappeared (ex-

changed with solvent), and the Hα of Gly8 split into Hα and Hα′ resonances (data not shown). Additional ROEs were not found either upon increasing temperature or by further TFE addition.

Local structure of the repeated sequence

The NOE pattern of particular interest involved the four residues for which the secondary structure predictions indicated the presence of a β-turn, that is, Glu-Pro-Gly-Cys. The strong sequential Hα-HN and HN(i+2)-HN(i+3) NOEs observed are indicative of a type II β-turn centered around Pro7 (i+1) and Gly8(i+2).[45] In contrast, a type I β-turn at this position should be expected to produce weak sequential Hα-HN and Hα(i)-HN(i+2) NOEs.[46] Using a set of 59 protein crystal structures, Wilmot and Thornton[47] found a sequence preference for type II β-turns that included a proline residue at position i+1 followed by a glycine in the i+2 position. A medium range ROE involving Hγ of Val12 and the amide proton of Glu14 was found. This could be indicative of a tight turn centered around Pro-Glu in the Val-Pro-Glu-Pro sequence as the temperature dependence for the amide protons of Val12 and Glu14 suggests. Although few ROEs were found, some of them involved sidechain protons. All threonines showed NOEs between CHγ and the amide proton of i+1 lysines. These data indicate that it is likely that side chains are folded towards the peptide chain.

Due to the small number of observed NOEs, we cannot rule out the presence of a less populated conformer and a possible conformational exchange in the intermediate NMR time scale. The distance constraints were thus used in the calculation not as rigid constraints but only as upper limits. Structure comparison by RMSD for backbone atom coordinates showed that, in any case, only a limited convergence was obtained. Thus the three-dimensional structure of the peptide could not be achieved due to the lack of long-range NOE connectivities.

Nevertheless, examinations of calculated structures confirm that the molecule is flexible and there are only partially local rigid regions. Local structures have been recognized as open β-turn structures in the residue sequence Glu-Pro-Gly-Cys as predicted (see above). The first turn is recognized in residues 6−9 and suggests the presence of a type II β-turn. The conformation of this region can be described by two clusters families (Figure 9A). Moreover, in all structures the proline has a trans conformation, and the distance between the Cα of Glu6 and Cys9 is <7 Å, a characteristic distance for a turn as defined by Chou and Fasman.[48] The resulting structures did not exhibit significant differences if analyzed in terms of energy (Table 4). A second region of local rigidity was

Table 3 ^{1}H-NMR values for SPR3 peptide

A Chemical shifts in phosphate buffer, pH 5.3 (ROESY, 300 ms, 298°C)

Residue	NH	Hα	Hβ	Others
Cys-1	–	–	3.09	
Thr-2	8.70	4.12	4.30	Hγ=1.20
Lys-3	8.47	4.33	1.76(β′)	Hγ=1.36(γ′) Hδ=1.65(δ′) Hε=– Hζ=7.50
Val-4	7.91	4.32	1.96	Hγ=0.87(γ′)
Pro-5	*	4.34	2.24(γ)	Hδ=3.87 Hδ′=3.66 Hβ′=1.86(γ′)
Glu-6	8.10	4.50	1.81	Hβ′=2.05 Hγ=2.44(γ′)
Pro-7	*	4.39	2.28(γ)	Hδ=3.78 Hδ′=3.70 Hβ′=1.94 Hγ′-2.04
Gly-8	8.50	3.96	*	
Cys-9	8.18	4.58	2.94	
Thr-10	8.28	4.18	4.31	Hγ=1.18
Lys-11	8.17	4.39	1.76(β′)	Hγ=1.36(γ′) Hδ=1.65(δ′) Hε=2.97(ε′) Hζ=7.50
Val-12	8.20	4.39	2.06	Hγ=0.91(γ′)
Pro-13	*	4.36	2.25(γ)	Hδ=3.87 Hδ′=3.66 Hβ′=1.86(γ′)
Glu-14	8.34	4.64	1.88	Hβ′=2.09 Hγ=2.50(γ′)
Pro-15	*	4.41	2.28(γ)	Hδ=3.78 Hδ′=3.70 Hβ′=1.97 Hγ′=2.02
Gly-16	8.25	3.91	*	
Cys-17	8.30	4.58	2.94	Hβ′=–
Thr-18	8.33	4.18	4.31	Hγ=1.18
Lys-19	8.31	4.36	1.76(β′)	Hγ=1.36(γ′) Hδ=1.65(δ′) Hε=2.97(ε′) Hζ=7.50
Val-20	8.20	4.39	2.06	Hγ-0.91(γ′)
Pro-21	*	4.30	2.25	Hδ=3.87 Hδ′=3.66 Hβ′=1.86(γ′)
Glu-22	8.50	4.50	1.94	Hβ′=2.09 Hγ=–
Pro-23	*	4.49	–	Hδ=3.78 Hδ′=3.70 Hβ′=1.97 Hγ′=2.02
Gly-24	8.34	3.97	*	

B Amide temperature shifts

Residue (ppb/K)	$\Delta\delta_{NH}/\Delta T$
2	5.62
3	2.49
4	3.13
6	3.08
8	4.38
9	6.92
10	3.08
11	2.38
12	1.89
14	0.65
16	5.08
17	4.38
18	3.08
19	3.73
20	1.89
22	3.78
24	5.71

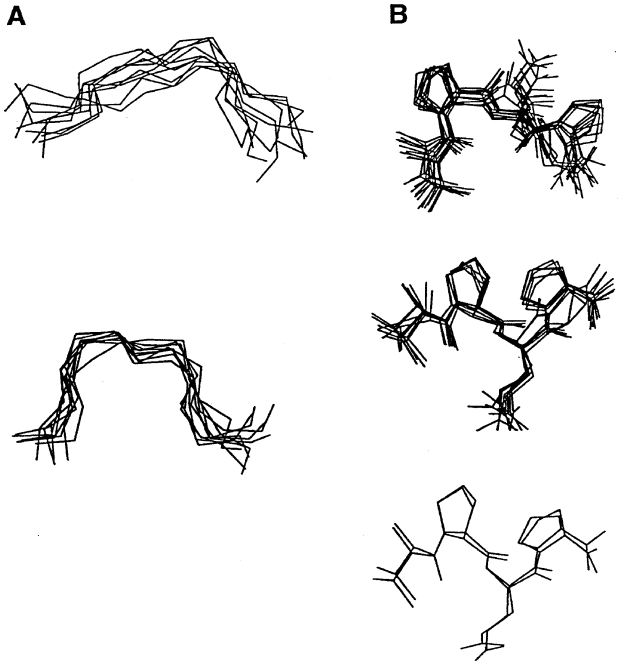


Figure 9 Intensities of NOEs were introduced in a distance geometry program which generated 50 compatible structures grouped in two main families differing by the number of H-bonds. (**A**) Two cluster families for the proposed β-turn conformation of the tetrapeptide region $Glu_6Pro_7Gly_8Cys_9$. Root mean square deviation (r.m.s.d.) for the upper conformer is 2.52, and for the lower conformer is 2.22. (**B**) Three cluster families for the conformation of the sequence $Val_{12}Pro_{13}Glu_{14}Pro_{15}$. Root mean square deviation for the upper conformer is 0.90, for the middle conformer is 0.57, and for the lower conformer is 0.56

centered around the sequence Val12-Pro13-Glu14-Pro15 where a medium range NOE contact was present between the sidechain of Val12 and the Glu14 amide proton. Cluster analyses allowed to distinguish three families for the possible conformation of this region (Figure 9B), overimposing the structures on the heavy atoms. All the conformations allow the hydrogen bond between the amide group of Glu14 and the carbonyl oxygen of Val12 as the low temperature coefficient of this residue requires (Table 3B). This sequence aligns with the epitope of the tumor associated antigen human Muc-1, a tandemly repeated protein that shares similar CD behavior and β-turn propensity with the SPR3 peptide.[49] The antigenic epitope Ala-Pro-Asp-Thr-Arg has a β-turn centered around Pro-Asp. The homologous SPR3 sequence (that is, aliphatic, proline, acidic residues) is likely to form a β-turn centered on Pro-Glu.

In vivo consequences of the present *in vitro* data

The foregoing sequencing data on the utilization of the Gln and Lys residues of SPR3 are summarized in Figure 5.

First, very few central domain residues were used in the *in vitro* reactions: $>95\%$ involved head A/B and tail domain residues. These data are consistent with our earlier observations for the cross-linking of the SPR1[16] and SPR2[17] proteins. Furthermore, they lend further strong support for our view that the SPRs in general serve as cross-bridging proteins in the CE structures of epithelia.

Table 4 Total energies for the best 27 structure generated for SPR3 peptide

Structure number	Energy (Kcal/mol)	Structure number	Energy (Kcal/mol)
1	55.6	15	55.0
2	55.6	16	54.1
3	54.2	17	52.0
4	54.8	18	53.9
5	52.5	19	54.8
6	51.0	20	55.7
7	53.1	21	54.5
8	54.3	22	50.7
9	56.0	23	55.2
10	54.0	24	55.7
11	52.2	25	52.6
12	55.4	26	55.4
13	59.3	27	54.0
14	49.6		

The 24 amino acid peptide analyzed by 2D ^{1}H-NMR generated a family of structures by distance geometry program differing in the energy involved

Second, we have identified some evidence of residue specificity by the TGase enzymes. TGase 1 used only head domain B sequences. This observation is identical to that made for the SPR1 proteins.[16] The TGase 2 enzyme also used primarily the head B domain residues. The TGase 3 enzyme, however, used both head A and B domain sequences in SPR3. In contrast, much more specificity for the head A domain was seen for SPR1,[16] and only the TGase 3 enzyme significantly cross-linked the SPR2 proteins which possess only head A domain-like sequences.[17]

Third, to date, we have very few data on the cross-linking of SPR3 proteins *in vivo*: we obtained ten cross-links involving SPR3 proteins, three of which utilized head domain and seven involved tail domain sequences.[10] Although this small data set precludes any definitive conclusions, it is nevertheless consistent with the *in vitro* data and lends robust support for the general pattern seen so far for other SPR proteins that essentially only head and tail domain sequences participate in cross-linking reactions.

Fourth, the data of Figures 2, 4 and Table 1 together allow the conclusion that at least three TGase enzymes may be involved in the cross-linking of the SPR3 protein *in vivo*. In contrast, essentially only the TGase 3 enzyme cross-links the SPR2, and TGases 1 and 3 coordinately cross-link the SPR1 proteins. In addition, we note that many of the epithelia which express SPR3 also co-express the novel TGase X enzyme. Although its properties have not yet been defined, it is possible that this enzyme might also be involved. From this we conclude that the SPR3 protein may be the most efficiently used member of the SPR family by TGases, a view supported by the kinetic data of Table 1. There may be important *in vivo* reasons for this unique property of the SPR3 protein. We note that tissues such as the forestomach, inner root sheath, tongue filiform ridges are especially hardened epithelia and have high turnover rates. Thus the rapid and promiscuous cross-

linking of the SPR3 protein by all available TGase enzymes may be required for hasty yet effective barrier formation.

Fifth, the double cross-linking experiments of Figure 1 offer the view that there is a well-defined temporal order for the cross-linking of the SPR3 protein during its incorporation onto the CE structure. Our data suggest that it may be cross-linked first by the cytosolic TGases 2 and/or 3 into short oligomers using head A and B domain as well as tail domain sequences. The data of Figure 1A for the native mouse SPR3 protein reveal that some of these intermediate short oligomers remain soluble, at least for some time. We propose that later these are attached permanently to the CE mostly by the membrane-associated TGase1 enzyme using predominantly the head B and tail domain sequences. However, the TGase 1 deficiency models indicate there must be some redundancy in this second step since neither transgenic null mice nor human lamellar ichthyosis patients have significant involvement of internal epithelia in which the SPR3 protein is expressed.[29,30]

Sixth, structural data (Figures 6, 7 and 9) suggest that the SPR3 central repeats are highly flexible and mobile thus the TGases might not be able to recognize the residues localized on the repeats as adequate substrate.

Conclusions

Taken together, the present SPR3 data as well as previous data for the SPR1 and 2 proteins reveal that only head and tail domain glutamines/lysines are used for cross-linking *in vitro* and *in vivo*. Yet the bulk of these proteins resides in their central domains which consist of variable numbers of proline-rich peptide repeats. Our new CD and NMR data suggest that these repeats are very flexible with only a low order of organized secondary structure. We thus conclude that these may serve as flexible cross-bridging 'spacers' or 'fillers' with variable volume. In this way the central peptide repeating domain permits extensive and promiscuous cross-linking between the functional ends of the SPR proteins and other protein constituents of the CE barrier structure in the network of CE.

Materials and Methods

Expression and purification of recombinant human SPR3

A full-length cDNA clone encoding human SPR3[19] was a gift of Dr. Claude Backendorf. Following addition of linkers, it was inserted into the pET11a bacterial expression vector (Novagen, Madison, WI, USA), and transformed into the *E. coli* B host strain BL/DE23 (Novagen). Protein expression was induced in the presence or absence of ^{35}S-cysteine (0.5 mCi/ml).[14] Bacterial pellets were lysed and dialyzed against 25 mM sodium citrate (pH 3.6), 1 mM dithiolthreitol, 1 mM EDTA and a cocktail of protease inhibitors, under which conditions the SPR3 protein remained soluble but most bacterial proteins precipitated.[14–17] Final purification of the recombinant SPR3 protein was accomplished by fast liquid chromatography on a 0.5×5 cm Mono-S column (Pharmacia Biotech, Piscataway, NJ, USA) equilibrated in the citrate buffer, and eluted by 0.2 M NaCl. The protein was analyzed on 4–25% SDS polyacrylamide gradient gels and examined by either autoradiography, or enhanced chemiluminescence using the Super Signal CL-HRP Substrate System (Pierce, Rockford, IL, USA) and a primary rabbit antibody broadly reactive against mouse and human SPR1/3 proteins.[20]

Isolation of native mouse SPR3

Forestomach tissue was obtained from adult BALB/c mice by dissection and chilled on ice. The tissue was homogenized using a Polytron probe in the above ice-cold sodium citrate buffer (10 ml/g wet weight tissue), and centrifuged at $25\,000 \times g$ for 30 min to pellet cellular debris and insoluble proteins (mostly keratins). As the SPR3 is very soluble in this buffer, in this way by SDS–PAGE, it was recovered $>75\%$ pure. The bulk of the remaining proteins soluble in this buffer were lesser amounts of histones. By chromatography on the same Mono-S column of above, native SPR3 was recovered $>95\%$ pure, as estimated by amino acid analysis of its distinctive cysteine, phenylalanine, lysine, proline, and threonine contents.[10]

TGase cross-linking of recombinant SPR3 *in vitro*

Analytical- and preparative-scale cross-linking reactions were accomplished exactly as described previously for the recombinant SPR1[16] and SPR2[17] proteins, using equivalent amounts of activity of the baculovirus-expressed TGase 1,[50] guinea-pig TGase 2 (Sigma), and activated guinea-pig TGase 3[52] enzymes. In addition, we also isolated several TGase 1 isoforms from terminally differentiating cultures of normal human epidermal keratinocytes.[26] These included: the full-length (106 kDa) membrane-bound form; the cytosolic full-length form; the membrane-bound highly active 67/33/10 kDa form; and the cytosolic 67/33 kDa and 67 kDa forms. In reactions with these forms, we used a constant molar amount (700 nM) of each enzyme form, based on the amount of ^{35}S-methionine incorporated under standard conditions during cell culture.[16] We also performed double cross-linking reactions, in which the products of an initial complete TGase 1 or TGase 3 reaction were subjected to cross-linking again by either the TGase 3 or TGase 1 enzyme.[16]

Kinetic constants of the three TGases were determined for the recombinant human SPR3 protein exactly as described previously.[16,17]

Protein chemical methods

Amounts of the recombinant SPR3 protein were determined by amino acid analyses. Uncross-linked or cross-linked proteins were digested to completion with trypsin (1 : 30 by weight, Sigma, sequencing grade). The peptides were resolved by HPLC,[16,17] collected, and sequenced. For the TGase 3 enzyme, as the peptides were poorly resolved due to extensive cross-linking, 1 min aliquots were collected across a broad peak.[17] The amount of isodipeptide formed in the *in vitro* TGase cross-linking reactions was determined by total enzymic digestion as before.[8,23]

A peptide of sequence $(TKVPEPGC)_3$, based on three consecutive repeats of human SPR3 protein[10] (see Figure 5F) was synthesized and purified to homogeneity by HPLC.

Secondary structure predictions

Secondary structure prediction and β-turn analyses were performed using the Chou-Fasman methods.[48]

Circular dichroism

Circular dichroism spectra were recorded on a Jasco 600 spectropolarimeter. Spectra were recorded from 200–250 nm with readings every 0.1 nm using a 0.1 cm pathlength curvet. Sample temperatures were maintained at 23, 30, 40, 60 or 70°C in a thermostated sample holder using external circulation for at least 1 h before measurement. The recombinant SPR3 protein (1.5 mg/ml) was equilibrated by dialysis into 10 mM phosphate buffer (pH 5.3). The synthetic peptide was purified by HPLC and dissolved into 10 mM phosphate buffer (pH 7.0) and used at a concentration of 1 mM. Solvent spectra were subtracted from those of the samples. A total of three or four scans were accumulated and averaged for each sample.

NMR experiments and sequential assignments

All spectra were recorded in a Bruker 400 MHz AM instrument at 298°C with 3 mM peptide concentration in degassed 10 mM phosphate buffer (pH 5.3), 15 mM dithiothreitol. All 2D data were acquired in the phase-sensitive mode with the saturation of the water signal during the relaxation delay. TOCSY data were collected with the data matrix (t1=512, t2=2K); relaxation delay=1.41 s; number of transients=64; isotropic mixing=0.068 s, 0.098 s; MLEV-17 pulse sequence was used for the spin-lock. NOESY[54] data were collected with similar acquisition parameters and for 0.280 and 0.300 s of mixing times. ROESY[55,56] data were collected in the phase-sensitive mode. Data matrix was (t1=512; t2=2K); mixing times applied 0.120 s and 0.24 s, and 0.30 s for measurements in TFE; relaxation delay=1.5 s; number of transients=64. The sequential assignment of SPR3 peptide was obtained by combining TOCSY and ROESY data at 298°C, and ROESY at 305°C and 278°C. ROESY experiments were also performed in different TFE concentrations with the same acquisition parameters (10, 15, 32 and 70% v/v). One TOCSY experiment was performed in 10% TFE with an isotropic mixing of 0.080 s.

Structural study from structural constraints obtained from the NMR data

An energetically stable structure was obtained for the molecular modeling of the SPR3 peptide. To analyze the ROE intensities in the 2D NMR spectrum (ROESY 0.300 s), a total of 151 ROE intensities were listed and classified into three categories. <3.0 (strong), <4.0 (medium), and <5.0 Å (weak) corresponding to distance constraints.

Three-dimensional structures were generated with these distance constraints using the XPLOR 3.1 program[57] using the default parameters, except for some minor modifications to increase the durations of the dynamics. A template structure with randomized coordinates and ideal geometry was generated. Then a substructure distance geometry protocol was applied introducing the NOE-derived distance constraints to generate 50 unregularized structures with a partial set of coordinates (protons without stereospecific assignment were treated as pseudoatoms). The final coordinates were regularized by simulated annealing[58,59] with a protocol consisting of 9 ps of heat procedure up to 2000°C. The contribution of the van der Waals terms was reduced during this stage to allow the maximal exploration of the conformational space. The system was cooled to 100°C by steps of 50°C in 12 ps with a concomitant increase of the van der Waals interactions. The structures were refined by the application of simulated annealing and energy minimization with a duration of 30 ps. Temperature was decreased from 1000°C to 100°C by steps of 25°C. Each structure was then submitted to 1000 cycles of energy minimization. The 50 structures were then checked for the NOE violations and bond and angle deviations from ideality. The lowest energy 27 structures were compared on the basis of pairwise RMSDs for the backbone atom coordinates (N, Cα, Cα′). Structure superimposition by direct examination on a graphic display was accomplished using graphical programs purchased from Molecular Simulations Inc. running on a Silicon Graphic workstation O_2.

Acknowledgments

We thank William Idler for assistance with the expression and purification of recombinant SPR3. Fabio Bertocchi is gratefully acknowledged for the execution of the NMR experiments. The work was partially carried out thanks to grants from MURST-Structural Biology and CNR-PF Biotechnology to MP, grant Telethon 417/bi to EC, and grants Min. San., MURST-40% 1997–1998, Telethon E872, and AIRC 1998 to GM.

References

1. Holbrook KA and Wolff K (1993) The structure and development of the skin. In: Dermatology in General Medicine. Fitzpatrick TB, Eisen AZ, Wolff K, Freedberg IM and Austen KF, (eds). McGraw-Hill, Inc., New York, pp 97–144
2. Polakowska RR and Haake AR (1994) The skin from a new perspective. Cell Death Differ. 1: 11–18
3. Melino G, De Laurenzi V, Catani MV, Terrinoni A, Ciani B, Candi E, Marekov LM and Steinert PM (1998) The cornified envelope: a model of cell death in the skin. In: Apoptosis: Biology, Mechanisms and Role in Disease. Kumar S (ed), Results Probl. Cell Differ. 24: 175–212
4. Reichert U, Michel S and Schmidt R (1993) The cornified envelope: A key structure of terminally differentiating keratinocytes. In Molecular Biology of the Skin. Darmon M and Blumenberg M (eds), Academic Press, Inc., New York, pp 107–150
5. Hohl D (1990) Cornified Cell envelope. Dermatologica 180: 201–211
6. Simon M (1994) The epidermal cornified envelope and its precursors. In: The Keratinocyte Handbook. Leigh IM, Lane EB and Watt FM (eds). Cambridge University Press, Cambridge, pp 275–292
7. Steinert PM (1995) A model for the hierarchical structure of the human epidermal cornified cell envelope. Cell Death Differ. 2: 33–40
8. Steinert PM and Marekov LN (1995) The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodipeptide cross-linked components of the human epidermal cornified cell envelope. J. Biol. Chem. 270: 17702–17711
9. Steinert PM and Marekov LN (1997) Involucrin is an important early component in the assembly of the epidermal cornified cell envelope. J. Biol. Chem. 272: 2021–2030
10. Steinert PM, Kartasova T and Marekov LN (1998) Biochemical evidence that small proline-rich proteins and trichohyalin function in epithelia by modulation of the biomechanical properties of their cornified cell envelopes. J. Biol. Chem. 273: 11758–11769
11. Robinson NA, Lapic S, Welter JF and Eckert RL (1997) S100A11, S100A10, annexin I, desmosomal proteins, small proline-rich proteins, plasminogen activator inhibitor-2, and involucrin are components of the cornified envelope of cultured human epidermal keratinocytes. J. Biol. Chem. 272: 12035–12046
12. Jarnik M, Kartasova T, Steinert PM, Lichti U and Steven AC (1996) Differential expression and cell envelope incorporation of small proline-rich protein 1 in different cornified epithelia. J. Cell Sci. 109: 1381–1391
13. Jarnik M, Simon MN and Steven AC (1998) Cornified cell envelope assembly: a model based on electron microscopic determinations of thickness and projected density. J. Cell Sci. 111: 1051–1060
14. Candi E, Melino G, Mei G, Tarcsa E, Chung SI, Marekov LN, Steinert PM (1995) Biochemical, structural, and transglutaminase substrate properties of human loricrin, the major epidermal cornified cell envelope protein. J. Biol. Chem.270: 26382–26390
15. Nagy L, Thomazy VA, Saydak MM, Stein JP and Davis PJA (1997) The promoter of the mouse tissue transglutaminase gene directs tissue-specific, retinoid regulated and apoptosis-linked expression. Cell Death Differ. 4: 534–547

16. Candi E, Tarcsa E, Idler WW, Kartasova T, Marekov LN, Steinert PM (1999) Transglutaminase cross-linking properties of the small proline-rich 1 family of cornified cell envelope proteins: integration with loricrin. J. Biol. Chem. 274: 7226 – 7237
17. Tarcsa E, Candi E, Kartasova T, Idler WW, Marekov LN and Steinert PM (1998) Structural and transglutaminase substrate properties of the small proline-rich 2 family of cornified cell envelope proteins. J. Biol. Chem. 273: 23297 – 23303
18. Aeschlimann D, Koeller MK, Allen-Hoffmann BL and Mosher DF (1998) Isolation of a cDNA encoding a novel member of the transglutaminase gene family from human keratinocytes. Detection and identification of transglutaminase gene products based on reverse transcription-polymerase chain reaction with degenerate primers. J. Biol. Chem. 273: 3452 – 3460
19. Gibbs S, Fijneman R, Wiegant J, van Kessel AG, van de Putte P and Backendorf C (1993) Molecular characterization and evolution of the SPRR family of keratinocyte differentiation markers encoding small proline-rich proteins. Genomics 16: 630 – 637
20. Kartasova T, Darwiche N, Kohno Y, Koizumi H, Osada S, Huh N, Lichti U, Steinert PM and Kuroki T (1996) Sequence and expression patterns of mouse SPR1: Correlation of expression with epithelial function. J. Invest. Dermatol. 106: 294 – 304
21. Steinert PM, Candi E, Kartasova T and Marekov LN (1998) Small proline-rich proteins are cross-bridging proteins in the cornified cell envelopes of stratified squamous epithelia. J. Struct. Biol. 122: 76 – 85
22. Song H-J, Poy G, Darwiche N, Lichti U, Kuroki T, Kartasova T and Steinert PM (1999) Mouse Sprr2 genes: a clustered family of genes showing differential expression in epithelial tissues. Genomics 55: 28 – 42
23. Hohl D, de Viragh PA, Armiguet-Barras F, Gibbs S, Backendorf C and Huber M (1995) The small proline-rich proteins constitute a multigene family of differentially regulated cornified cell envelope precursor proteins. J. Invest. Dermatol. 106: 902 – 909
24. Austin SJ, Fujimoto W, Marvin KW, Vollberg TM, Lorand L and Jetten AM (1996) Cloning and regulation of cornifin beta, a new member of the cornifin/spr family. Suppression by retinoic acid receptor-selective retinoids. J. Biol. Chem. 271: 3737 – 3742
25. Fujimoto W, Nakanishi G, Arata J and Jetten AM (1997) Differential expression of human cornifin alpha and beta in squamous differentiating epithelial tissues and several skin lesions. J. Invest. Dermatol. 108: 200 – 204
26. Steinert PM, Chung S-I and Kim S-Y (1996) Inactive zymogen and highly active proteolytically processed membrane-bound forms of the transglutaminase 1 enzyme in human epidermal keratinocytes. Biochem. Biophys. Res. Commun. 221: 101 – 106
27. Birchbirchler PJ, Orr GR, Carter HA and Patterson MK (1977) Biochem. Res. Commun. 78: 1 – 7
28. Barry EL and Mosher DF (1990) Binding and degradation of blood coagulation factor XIII by cultured fibroblasts. J. Biol. Chem. 265: 9302 – 9307
29. Traupe H (1989) The Ichthyoses: A Guide to Clinical Diagnosis, Genetic Counseling and Therapy. Springer-Verlag, Heidelberg.
30. Matsuki M, Yamashita F, Ishida-Yamamoto A, Yamada K, Kinoshita C, Fushiki S, Ueda E, Morishima Y, Tabata K, Yasuno H, Hashida M, Iizuka H, Ikawa M, Okabe M, Kondoh G, Kinoshita T, Takeda J and Yamanishi K (1998) Defective stratum corneum and early neonatal death in mice lacking the gene for transglutaminase 1 (keratinocyte transglutaminase). Proc. Natl. Acad. Sci. USA 95: 1044 – 1049
31. Isemura T, Asakura J, Shibata S, Iasemura S, Saitoh E and Sanada K (1983) Conformational study of the salivary proline-rich polypeptides. Int. J. Peptide Protein Res. 21: 281 – 287
32. Shibata S, Asakura J, Isemura T, Isemura S, Saitoh E and Sanada K (1984) Conformational study of the basic proline-rich polypeptides from human parotid saliva. Int. J. Peptide Protein Res. 23: 158 – 165
33. Raj PA, Edgerton M and Levine MJ. (1990) Salivary Histatin 5: dependence of sequence, chain length and helical conformation for candidacidal activity. J. Biol. Chem. 265: 3898 – 3905
34. Raj PA, Johnsson M, Levine MJ and Nancollas GH (1992) Salivary staherin: dependence of sequence, charge, hydrogen bonding potency, and helical conformation for adsorption to hydroxyapatite and inhibition of mineralization. J. Biol. Chem. 267: 5968 – 5976
35. Tanford C (1970) Theoretical models for the mechanism of denaturation. Adv. Prot. Chem. 24: 1 – 95
36. Darrell Fontenot J, Tiandra N, Bu D, Ho C, Montelaro RC and Finn OJ (1993) Biophysical characterization of one-, two-, and three-tandem repeats of human mucin (muc-1) protein core. Cancer Research 53: 5386 – 5394
37. Woody RW (1976) Studies of theoretical circular dichroism of polypeptides: contributions of beta turns. In: Blout ER, *et al.* (ed.). Peptides, polypeptides and proteins. New York, Wiley, pp. 338 – 350
38. Tatham AS, Drake AF and Shewry PR (1985) A conformational study of a glutamine- and proline-rich cereal seed protein *C hordein*. Biochem. J. 226: 557 – 562
39. Groß KH and Kalbitzer HR (1998) Distribution of chemical shifts in 1H nuclear magnetic resonance spectra of proteins. J. Magn. Reson. 76: 87 – 89
40. Urry DW (1997) Physical chemistry of biological free energy transduction as demonstrated by elastic protein-based polymers. J. Phys. Chem. 101: 11007 – 11028
41. Jimenez MA, Bruix M, Gonzalez C, Blanco FJ, Nieto JL, Herranz J and Rico M (1993) CD and 1H-NMR studies on the conformational properties of peptide fragments from the C-terminal domain of thermolysin. Eur J Biochem 211(3): 569 – 581
42. O'Neal KD, Chari MV, McDonald CH, Cook RG, Yuu-Lee L-Y, Morrissett JD and Shearer WT (1986) Multiple cis-trans conformers of the prolactin receptor proline-rich motif (PRM) peptide detected by reverse-phase HPLC, CD and NMR spectroscopy. Biochem. J. 315: 833 – 844
43. Dyson HJ, Rance M, Houghten RA and Wright PE (1998) Folding of immunogenic peptide fragments in water solution. J. Mol. Biol. 201: 161 – 200
44. Williamson MP (1993) Peptide structure determination by NMR. Methods Mol. Biol. 17: 69 – 85
45. Rose DR, Gierasch LM and Smith JA (1985) Turns in peptides and proteins. Adv. Prot. Chem. 37: 1 – 109
46. Wutrich K (1986) NMR of proteins and nucleic acids. John Wiley & Sons, New York.
47. Wilmot CM and Thornton JM (1988) Analysis and prediction of the different types of beta-turn in proteins. J. Mol. Biol. 205: 221 – 232
48. Chou PY and Fasman GD (1977) β-turns in proteins. J. Mol. Biol. 115: 135 – 175
49. Fontenot JD, Mariappan SV, Catasti P, Domenech N, Finn OJ and Gupta G (1995) Structure of a tumor associated antigen containing a tandemly repeated immunodominant epitope. J. Biomol. Struct. Dyn. 13: 245 – 260
50. Leszczynski JF and Rose GD (1986) Loops in globular proteins: a novel category of secondary structure. Science 234: 849 – 855
51. Candi E, Melino G, Lahm A, Ceci R, Rossi A, Kim IG, Ciani B and Steinert PM (1998) Transglutaminase 1 mutations in lamellar ichthyosis. Loss of activity due to failure of activation by proteolytic processing. J. Biol. Chem. 273: 13693 – 13702
52. Kim H-C, Lewis MS, Gorman JJ, Park S-C, Girard JE, Folk JE and Chung S-I (1990) Protransglutaminase E from guinea pig skin. Isolation and partial characterization. J. Biol. Chem. 265: 21971 – 21978
53. Hohl D, Lichti U, Turner ML, Roop DR and Steinert PM (1991) Characterization of human loricrin. Structure and function of a new class of epidermal cell envelope proteins. J. Biol. Chem. 266: 6626 – 6636
54. Jeener J, Meier BH, Bachmann P and Ernst RR (1979) Investigation of exchange processes by two-dimensional NMR spectroscopy. J. Chem. Phys. 71: 4546 – 4553
55. Bax A and Davis DG (1985) Practical aspects of two-dimensional transverse NOE spectroscopy. J. Magn. Res. 63: 207 – 213
56. Griesinger C and Ernst RR (1987) Frequency offset effects and their elimination in NMR Rotating-Frame Cross-Relaxation spectroscopy. J. Magn. Reson. 75: 261 – 271
57. Brunger AT (1992) XPLOR, a system for crystallography and NMR. Yale University Press, New Haven, CT.
58. Nilges M, Clore GM and Gronenborn AM (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding. FEBS Lett. 229: 317 – 324
59. Nilges M, Gronenborn AM, Brunger AT and Clore GM (1998) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. Protein Eng. 2: 27 – 38
60. Kim S-Y, Kim I-G, Chung S-I and Steinert PM (1994) The structure of the transglutaminase 1 enzyme: deletion cloning reveals domains which regulate its specific activity and substrate specificity. J. Biol. Chem. 269: 27979 – 27986