Human stratum corneum proteomics reveals cross-linking of a broad spectrum of proteins in cornified envelopes

Short title: Corneocyte envelope proteomics

Noreen Karim[1,2], Brett S. Phinney[3], Michelle Salemi[3], Pei-Wen Wu[2,4], Muhammad Naeem[1], Robert H. Rice[2,4]

[1]Medical Genetics Research Laboratory, Department of Biotechnology, Quaid-i-Azam University, Islamabad, Pakistan

[2]Department of Environmental Toxicology, University of California, Davis, CA

[3]Proteomics Core Facility, University of California, Davis, CA

[4]Forensic Science Program, University of California, Davis, CA

Correspondence: Dr. Robert H. Rice, Department of Environmental Toxicology, University of California, Davis, CA 95616-8588; Tel 530-752-5176, Fax 530-752-3394, Email rhrice@ucdavis.edu

Key Words: Keratin, Keratinocyte, Loricrin, Proteomics, TGM1

Concise Communication

Abstract

Defects in keratinocyte transglutaminase (TGM1), resulting in an improper protein scaffold for deposition of the lipid barrier, comprise a major source of autosomal recessive congenital ichthyosis. For that reason, the composition and formation of the cornified (cross-linked) protein envelope of the epidermis have been of considerable interest. Since the isopeptide cross-linked protein components are not individually isolable once incorporated, purified envelopes were analyzed by mass spectrometry after trypsin digestion. Quantitative estimates of the identified components revealed some 170 proteins, each comprising at least 0.001% of the total, of which keratins were major constituents accounting for ≈74% of the total. Some prevalent non-keratin constituents such as keratinocyte proline rich protein, loricrin and late envelope protein-7 were preferentially incorporated into envelopes. The results suggest a model where, as previously observed in hair shaft and nail plate, a diversity of cellular proteins are incorporated. They also help rationalize the minimal effect on epidermis of ablating genes for specific single envelope structural components. The quantitative profile of constituent proteins provides a foundation for future exploration of envelope perturbations that may occur in pathological conditions.

Background

The cornified outer layer of epidermis (*stratum corneum*) is our major extra-pulmonary barrier to the environment. This layer results from a well-orchestrated program of terminal differentiation producing corneocytes filled primarily with keratin intermediate filaments tightly connected by disulfide bonding.[1] This cytoskeleton is interconnected to an isopeptide cross-linked protein envelope at the cell periphery stabilized by transglutaminase cross-linking.[2] The envelope serves as a scaffold for deposition of the lipid barrier containing ω-OH-ceramides secreted from inside the maturing keratinocytes.[3] That defects in the transglutaminase gene *TGM1* are a major cause of autosomal recessive congenital ichthyosis[4,5] testifies to the importance of proper envelope formation for barrier function.

Questions addressed

The importance of transglutaminase cross-linking to barrier function suggested that transglutaminase substrate proteins were also critical to barrier function. If the envelope consisted of a small number of major protein constituents, then their loss would also result in barrier defects and hence skin disease. While ablation of *TGM1* produced a lethal barrier defect in mice,[6] ablation of genes encoding proposed major envelope proteins such as loricrin and involucrin had minimal effect.[7,8] Barrier defects from ablation of multiple genes encoding envelope components[9] support the suggestion of backup or redundant substrate proteins.[10] This suggestion has been borne out and augmented by proteomic analysis of envelopes from newborn mouse skin.[11] Present work extends the analysis to adult human epidermis. Developing a quantitative accounting of these components is anticipated to permit analyzing changes in the proteomic profile of envelopes that may occur during pathological processes.

Experimental design

Samples of epidermal squames were collected with CuDerm D110 D-Squame Pro 2.2 cm adhesive circles from the forearms of 3 male and 3 female healthy individuals with informed consent (University of California Davis IRB#217868-14) and eluted from the tapes as previously described.[12] Some samples were processed with reduction and alkylation for mass spectrometry without fractionation, while others were separated into SDS-dithioerythritol

insoluble (envelope) and solubilized fractions.[11,13] As previously found for SDS-dithioerythritol insoluble fractions of hair shaft and nail plate, >90% of the protein was solubilized by the digestion.[14,15] An envelope is shown by electron microscopy (Supplementary Figure S1) after purification as described.

Peptides were subjected to LC-MS/MS analysis on a Thermo Scientific Q Exactive Plus Orbitrap mass spectrometer essentially as previously described.[12] Intensity-based absolute quantitation (iBAQ) values of samples analyzed were used to calculate the relative molar amount of each protein.[16] To this end, iBAQ values were calculated by importing MaxQuant (version 1.5.7.4) MS1 intensity values and identification results into Scaffold (version Scaffold_4.8.2, Proteome Software Inc., Portland) and having Scaffold calculate the iBAQ values. To obtain the Scaffold output, the MS/MS generated spectral data were searched with Uniprot Human Proteome database (7/13/2017), a database of common laboratory contaminants (www.thegpm.org/crap) and an equal number of reverse decoy sequences, employing X! Tandem (The GPM, thegpm.org; version X! Tandem Alanine (2017.2.1.4)). Proteins that could not be differentiated by MS/MS analysis due to presence of shared peptides were clustered. The spectral count values of exclusive peptides (peptides belonging to only one protein) were used to delete those proteins with few or no exclusive peptides from the analysis. The raw data, peak lists and Scaffold file used in the study are available in the Massive Proteomics repository (massive.ucsd.edu/#MSV000083265) and ProteomeExchange (http://www.proteomexchange.org/ #PXD012122).

Results and conclusions

Corneocytes of epidermal squames are known to be comprised primarily of keratin intermediate filaments. Quantitative estimation of these proteins in squames from forearm stratum corneum yields a lineup of >150 constituents (Supplementary Table S1), of which the top 25 are shown in Table 1. Keratin family members account for ≈93% of the total according to this method of quantitation and are known to be a critical constituent of envelopes.[17] Similarly, the protein solubilized from squames by vigorous extraction with SDS under reducing conditions has the same keratin content and essentially the same lineup of protein constituents with some minor variations (Table S2). By contrast, the cross-linked envelope fraction exhibited a lower keratin content, though still high (74%), and a distinctive lineup of constituents (Table S3).

Among the proteins in the envelope fraction are a number long known to be envelope components including several keratins, loricrin, small proline rich proteins, S100 proteins, and some reported more recently such as keratinocyte proline rich protein and late cornified envelope proteins.[18-20] The lineup strongly resembles a large group of proteins recently reported in envelopes from newborn mouse skin,[11] but numerous quantitative differences are evident. In present work, 172 proteins were detected in the solubilized fraction of the squames at levels of at least 0.001% of the total, while 194 were detected in the envelope fraction, more possibly due to less masking by a lower keratin content. Thus, the cross-linked fraction appears to be a sampling of most of the proteins detected in the corneocytes. However, the relative amounts differed. Of the total proteins detected, 109 were estimated to differ by at least 10-fold in the solubilized versus cross-linked fractions, with 54% of them being more prevalent in envelopes. Table 2 lists those proteins of levels at least 0.01%.

The estimated relative amounts of envelope components are only semi-quantitative, but are highly informative whether a given protein is highly, moderately or not prevalent in them. An

additional uncertainty in the case of envelopes is the influence of cross-linking, since an estimated 18% of envelope protein lysines participate in isopeptide cross-linking[21] and thereby are not subject to cleavage. Thus a minority of tryptic peptides containing isopeptide bonds will not be identified in database searches unless specific cross-linked peptides (largely unknown) are targeted. In principle, this phenomenon could apply to any of the proteins in the envelope fraction, and it would help rationalize the low content of involucrin (0.05%) if a large fraction of its numerous glutamines served as transglutaminase substrates. Saponification has been reported to assist involucrin detection,[22] but in our hands that treatment markedly decreased the yield of peptides from the digest without increasing involucrin yield. Involucrin was not detected among the solubilized proteins despite being expressed at a high level, probably due to its marked sensitivity to proteolytic degradation during the extensive remodeling process of terminal differentiation.[23] Loricrin differs substantially in prevalence from a previous estimate based on amino acid compositions of envelopes.[24] This discrepancy likely reflects that elegant models for envelopes previously proposed[2] consist of only a small number of constituent proteins from which their relative amounts were calculated. Reportedly subject to uncertainties,[25] such calculations previously gave high levels of keratin that were discounted in skin of mice in which the loricrin gene was ablated.[11]

Epidermal corneocytes share with the hair shaft and nail plate the incorporation of a wide range of proteins into isopeptide structures comprising 10-20% of the total cellular protein.[26] Incorporating a diversity of protein substrates to create a protective peripheral shield would appear advantageous to supplement the highly disulfide bonded cell interior. The transglutaminase encoded by TGM1 has long been known to be capable of cross-linking involucrin in cell culture to membrane-associated proteins that are not expressed exclusively in keratinocytes.[27] The availability of a multitude of alternative substrates could rationalize the minimally perturbed phenotype arising from ablation of single transglutaminase substrate proteins such as involucrin or loricrin.[7,8] This finding supports the previously anticipated existence of such redundancy or a backup system.[10]

Present results are reminiscent of the “dustbin hypothesis” proposed on the basis of ionophore treatment of SV40-immortalized human epidermal cells.[28] The concept that envelopes could be formed from “waste” proteins (those surviving proteolytic remodeling) at the time cross-linking is initiated during terminal differentiation was superseded by a proposed model envelope comprised of a small number of distinct proteins. In contrast, present work is consistent with the use of many or most of the proteins present, as proposed earlier, where the data indicate clearly that some proteins are incorporated to a larger extent than others. The quantitative aspect of this study provides a more realistic accounting of the envelope proteome than previously available, providing a foundation for improving modeling of envelope formation. It also offers a basis for analyzing possible perturbation of the envelope proteome by aging, environmental pollution and disease states.[29-31]

Acknowledgments

We thank Ms. Patricia Kysar for expert technical assistance in electron microscopy, and the International Research Support Initiative Program of the Higher Education Commission of Pakistan and the USDA(NIFA)/University of California Agricultural Experiment Station (CA-D-ETX-2152-H) for financial support of this work.

Conflict of Interest

156 The authors state that they have no conflicts of interest.

157 Author Contributions

158 NK collected the samples and processed them for mass spectrometry. BSP and MS
159 performed the mass spectrometry and processed the data. PWW performed the transmission
160 electron microscopy. MN and RHR performed supervisory and organizational functions. NK and
161 RHR wrote the manuscript with input and editing from the other authors.

162 References

163 [1] T.T. Sun, H. Green, *J. Biol. Chem.* 1978, *253,* 2053
164 [2] E. Candi, R. Schmidt, G. Melino, *Nature Rev. Molec. Cell Biol* 2005, *6,* 328
165 [3] P.M. Elias, R. Gruber, D. Crumrine, G. Menon, M.L. Williams, J.S. Wakefield, W.M. Holleran,
166 Y. Uchida, *Biochim. Biophys. Acta* 2014, *1841,* 314
167 [4] M. Huber, I. Rettler, K. Bernasconi, E. Frenk, S.P.M. Lavrijsen, M. Ponec, A. Bon, S.
168 Lautenschlager, D.F. Schorderet, D. Hohl, *Science* 1995, *267,* 525
169 [5] L.J. Russell, J.J. DiGiovanna, G.R. Rogers, P.M. Steinert, N. Hashem, J.G. Compton, S.J.
170 Bale, *Nat. Genet.* 1995, *9,* 279
171 [6] M. Matsuki, F. Yamashita, A. Ishida-Yamamoto, K. Yamada, C. Kinoshita, S. Fushiki, E.
172 Ueda, Y. Morishima, K. Tabata, H. Yasuno, M. Hashida, H. Iizuka, M. Ikawa, M. Okabe, G.
173 Kondoh, T. Kinoshita, J. Takeda, K. Yamanishi, *Proc. Natl. Acad. Sci. USA* 1998, *95,* 1044
174 [7] P.J. Koch, P.A. de Viragh, E. Scharer, D. Bundman, M.A. Longley, J. Bickenbach, Y.
175 Kawachi, Y. Suga, Z. Zhou, M. Huber, D. Hohl, T. Kartasova, M. Jarnik, A.C. Steven, D.R.
176 Roop, *J. Cell Biol.* 2000, *151,* 389
177 [8] P. Djian, K. Easley, H. Green, *J. Cell Biol.* 2000, *151,* 381
178 [9] L.M. Sevilla, R. Nachat, K.R. Groot, J.F. Klement, J. Uitto, P. Djian, A. Määttä, F.M. Watt, *J.*
179 *Cell Biol.* 2007, *179,* 1599
180 [10] M. Jarnik, P.A. de Viragh, E. Scharer, D. Bundman, M.N. Simon, D.R. Roop, A.C. Steven,
181 *J. Invest. Dermatol.* 2002, *118,* 102
182 [11] R.H. Rice, B.P. Durbin-Johnson, Y.I. Ishitsuka, M. Salemi, B.S. Phinney, D.M. Rocke, D.R.
183 Roop, *J. Proteome Res.* 2016, *15,* 2560
184 [12] R.H. Rice, B.P. Durbin-Johnson, M. Salemi, M.E. Schwartz, D.M. Rocke, B.S. Phinney, *J.*
185 *Proteomics* 2017, *165,* 132
186 [13] R.H. Rice, *Int. J. Cosmetic Sci.* 2011, *62,* 229
187 [14] Y.J. Lee, R.H. Rice, Y.M. Lee, *Molec. Cell Proteom.* 2006, *5,* 789
188 [15] R.H. Rice, Y. Xia, R.J. Alvarado, B.S. Phinney, *J. Proteome Res.* 2010, *9,* 6752
189 [16] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M.
190 Selbach, *Nature* 2011, *473,* 337
191 [17] V. Kumar, J.-E. Bouameur, J. Bär, R.H. Rice, H.-T. Hornig-Do, D.R. Roop, N. Schwarz, S.
192 Brodesser, S. Thiering, R.E. Leube, R.J. Wiesner, P. Vijayaraj, C.B. Brazel, S. Heller, H. Binder,
193 H. Löffler-Wirth, P. Seibel, T.M. Magin, *J. Cell Biol.* 2015, *211,* 1057
194 [18] A.E. Kalinin, A.V. Kajava, P.M. Steinert, *BioEssays* 2002, *24,* 789
195 [19] W.H. Lee, S. Jang, J.S. Lee, Y. Lee, E.Y. Seo, K.H. You, S.C. Lee, K.I. Nam, J.M. Kim,
196 S.H. Kee, J.M. Yang, Y.J. Seo, J.K. Park, C.D. Kim, J.H. Lee, *J. Invest. Dermatol.* 2005, *125,*
197 995
198 [20] D. Marshall, M.J. Hardman, K.M. Nield, C. Byrne, *Proc. Natl. Acad. Sci. USA* 2001, *98,*
199 13031
200 [21] R.H. Rice, H. Green, *Cell* 1977, *11,* 417
201 [22] P.M. Steinert, L.N. Marekov, *J. Biol. Chem.* 1997, *272,* 2021
202 [23] R.M. Lavker, A.G. Matoltsy, *J. Cell Biol.* 1970, *44,* 501
203 [24] Z. Nemes, P.M. Steinert, *Exper. Molec. Med.* 1999, *31,* 5

204 [25] N.A. Robinson, S. Lapic, J.F. Welter, R.L. Eckert, *J. Biol. Chem.* 1997, *272,* 12035
205 [26] R.H. Rice, B.P. Durbin-Johnson, S.M. Mann, M. Salemi, S. Urayama, D.M. Rocke, B.S.
206 Phinney, J.P. Sundberg, *Exp. Dermatol.* 2018, *27,* 931
207 [27] M. Simon, H. Green, *Cell* 1984, *36,* 827
208 [28] S. Michel, R. Schmidt, S.M. Robinson, B. Shroot, U. Reichert, *J. Invest. Dermatol.* 1987, *88,*
209 301
210 [29] M. Rinnerthaler, J. Duschl, P. Steinbacher, M. Salzmann, J. Bischof, M. Schuller, H.
211 Wimmer, T. Peer, J.W. Bauer, K. Richter, *Exp. Dermatol.* 2013, *22,* 329
212 [30] K.E. Kim, D. Cho, H.J. Park, *Life Sci.* 2016, *152,* 126
213 [31] J. Krutmann, W. Liu, L. Li, X. Pan, M. Crawford, G. Sore, S. Seite, *J. Dermatol. Sci.* 2014,
214 *76,* 163
215 [32] R.H. Rice, D.M. Rocke, H.-S. Tsai, Y.J. Lee, K.A. Silva, J.P. Sundberg, *J. Invest. Dermatol.*
216 2009, *129,* 2120

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239 **Table 1. Quantitative estimates of protein components.** Intensity-based absolute
240 quantitation (iBAQ) values of samples analyzed with the Q-Exactive Plus mass spectrometer
241 were used to calculate the relative amount (%) of each protein.[16] iBAQ values were calculated
242 by importing MaxQuant (version 1.5.7.4) MS1 intensity values and identification results into
243 Scaffold and having Scaffold calculate the iBAQ values. To this end, average calculated values
244 for each protein and subject were normalized to a total value of 100 for each category - Total:
245 unfractionated protein; Sol: solubilized by SDS and dithioerythritol; Env: insoluble after
246 extraction. Deleted from clusters based on exclusive spectral counts were keratins 7, 24, 27, 40,
247 79 and 84. Shown are the 25 most prevalent proteins in unfractionated squames (Total) or in
248 the separate fractions of solubilized (Sol) or envelope (Env) protein. To minimize ambiguity,
249 proteins are identified by gene names.

Gene name	Total	Gene name	Sol	Gene name	Env
KRT1	31.88	KRT1	30.48	KRT1	24.40
KRT10,13,15	20.66	KRT10,13,15	19.38	KRT10,13,15	20.81
KRT2,3,5,6A,6B,75	9.99	KRT2,3,5,6A,6B,75	9.04	KRT2,3,5,6A,6B,75	9.52
KRT16	6.66	KRT16	7.66	KPRP	6.85
KRT77	5.88	KRT77	4.89	KRT16	4.91
KRT14,17,19	3.90	KRT14,17,19	4.43	LOR	3.28
KRT32	2.61	KRT9	2.58	KRT77	3.09
KRT35	2.43	KRT37	2.45	c1orf68/LEP7	2.87
KRT9	2.13	KRT32	2.36	KRT14,17,19	2.60
KRT31,33A,33B,34	1.62	KRT25,27	2.27	FLG2	1.71
KRT78	1.38	KRT35	2.22	KRT9	1.56
KPRP	1.33	KRT78	1.74	S100A16	1.20
KRT37	1.07	KRT31,33A,33B,34	1.38	KRT32	1.16
KRT72	0.84	KRT72	0.68	KRT35	1.08
c1orf68/LEP7	0.65	TXN	0.67	KRT25,27	1.00
KRT81,83,86	0.57	BLMH	0.57	KRT78	0.86
KRT4	0.49	DSG1	0.51	DCD	0.82
KRT85	0.42	KRT81,83,86	0.51	DSG1	0.73
DSG1	0.42	KRT85	0.38	KRT31,33A,33B,34	0.71
TXN	0.34	FLG2	0.35	KRT72	0.65
FLG2	0.34	KPRP	0.34	SPRR2D,E,G	0.55
KRTAP2-3	0.32	KRT80	0.31	ANXA2	0.49
DCD	0.27	DSC1	0.30	S100A10	0.48
DSC1	0.25	NCCRP1	0.26	LCE6A	0.44
S100A14	0.25	GAPDH	0.25	BLMH	0.42

250
251
252
253

254 **Table 2. Proteins enriched or deficient in envelopes judging by their iBAQ values.**
255 Proteins enriched >10 fold in the envelope fraction (Env) are in bold, while those enriched >10
256 fold in the solubilized fraction (Sol) compared to envelopes are italicized. iBAQ values are listed
257 down to 0.1% in order of prevalence in the unfractionated samples. To minimize ambiguity,
258 proteins are identified by gene names.

Gene name	Sol	Env
KPRP	0.34	6.85
c1orf68/LEP7	0.01	2.87
DSC1	0.3	0.02
LOR	0.02	3.28
CDSN	0.15	0.01
SPRRs	0.02	0.77
KRT71	0	0.13
S100A16	0	1.2
KRTAP1-5	0.05	0
ALOXE3	0.02	0.37
DSC3	0.04	0
CPNE	0.01	0.14
FLG	0.01	0.1
SPRR5	0	0.06
CRCT1	0	0.08
ASAH1	0.02	0
SERPINB7	0.02	0
KLK5	0.02	0
GNG12	0	0.05
KRTAP10-3	0	0.04
HIST1H4A	0	0.02
S100A11	0	0.11

259
260
261
262
263
264
265
266
267
268
269

270 **Supplementary Files**

271 **Table S1. Quantitative estimates of relative protein amounts in unfractionated**
272 **corneocytes.** iBAQ values calculated for each sample were averaged, and the total was
273 normalized to 100.

274 **Table S2. Quantitative estimates of relative protein amounts in solubilized fraction.** iBAQ
275 values calculated for each sample were averaged, and the total was normalized to 100.

276 **Table S3. Quantitative estimates of relative protein amounts in cross-linked envelope**
277 **fraction.** iBAQ values calculated for each sample were averaged, and the total was normalized
278 to 100.

279 **Figure S1. Envelope ultrastructure.** Squames eluted from Cu-Derm tape circles were heated
280 in 2% SDS – 25 mM dithioerythritol – 0.1 M sodium phosphate (pH 7.8) and then submitted for
281 electron microscopy. For this purpose, samples were immersed in Karnovsky's fixative,
282 postfixed in osmium tetroxide, embedded in a combination of Araldite and EMBed, sectioned,
283 stained with uranyl acetate and lead citrate and examined by transmission electron
284 microscopy.[32]

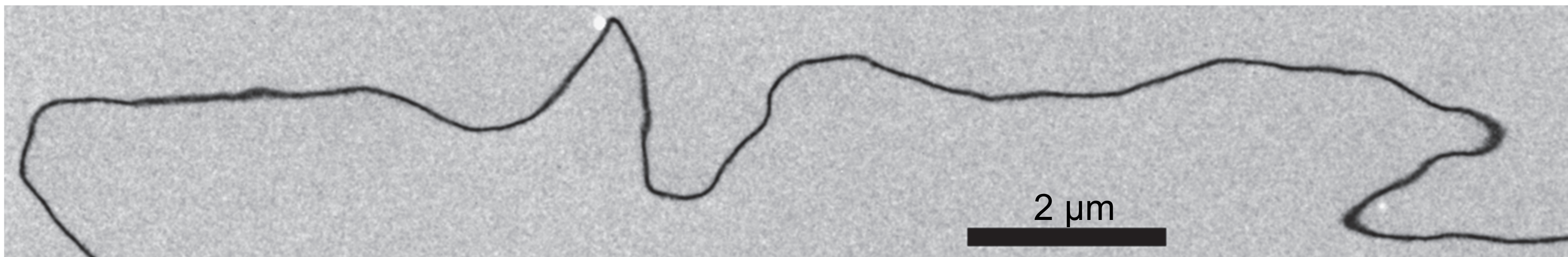

Figure S1. Envelope ultrastructure. Squames eluted from Cu-Derm tape circles were heated in 2% SDS – 25 mM dithioerythritol – 0.1 M sodium phosphate (pH 7.8) and then submitted for electron microscopy. For this purpose, samples were immersed in Karnovsky's fixative, postfixed in osmium tetroxide, embedded in a combination of Araldite and EMBed, sectioned, stained with uranyl acetate and lead citrate and examined by transmission electron microscopy.[32]