Complete amino acid sequence and homologies of human erythrocyte membrane protein band 4.2

(factor 13/transglutaminase/cDNA sequence)

Catherine Korsgren*, Jack Lawler[†], Stephen Lambert^{*}, David Speicher[‡], and Carl M. Cohen^{*§}

*Department of Biomedical Research, St. Elizabeth's Hospital, Boston, MA and Tufts University School of Medicine, Boston, MA; [†]Vascular Research Division, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; and [‡]The Wistar Institute, Philadelphia, PA 19104

Communicated by L. Lorand, October 30, 1989 (received for review September 19, 1989)

ABSTRACT The complete amino acid sequence for human erythrocyte band 4.2 has been derived from the nucleotide sequence of a full-length 2.35-kilobase (kb) cDNA. The 2.35-kb cDNA was isolated from a human reticulocyte cDNA library made in the expression vector $\lambda gt11$. Of the 2348 base pairs (bp), 2073 bp encode 691 amino acids representing 76.9 kDa (the SDS/PAGE molecular mass is 72 kDa). RNA blot analysis of human reticulocyte total RNA gives a message size for band 4.2 of 2.4 kb. The amino acid sequence of band 4.2 has homology with two closely related Ca²⁺-dependent crosslinking proteins, guinea pig liver transglutaminase (proteinglutamine γ -glutamyltransferase; protein-glutamine:amine γ glutamyltransferase, EC 2.3.2.13) (32% identity in a 446amino acid overlap) and the a subunit of human coagulation factor XIII (27% identity in a 639-amino acid overlap), a transglutaminase that forms intermolecular γ -glutamyl- ε -lysine bonds between fibrin molecules. The region of greatest identity includes a 49-amino acid stretch of band 4.2, which is 69% and 51% identical with guinea pig liver transglutaminase and the a subunit of factor XIII, respectively, within the regions that contain the active sites of these enzymes. Significantly, within the five contiguous consensus residues of the transglutaminase active site, Gly-Gln-Cys-Trp-Val, band 4.2 has an alanine substituted for cysteine (which is apparently essential for activity). Consistent with this active site substitution, erythrocyte membranes or inside-out vesicles, which contain band 4.2, show no evidence of transglutaminase activity by two types of in vitro assay.

Band 4.2 is a major protein (5% by weight) of the human erythrocyte membrane (1). This protein has an apparent mass of 72 kDa on SDS gels but forms oligomers both in solution and on the erythrocyte membrane (2, 3). Previous studies from our laboratory have shown that band 4.2 binds to the 43-kDa cytoplasmic domain of band 3, the erythrocyte anion transport protein (3, 4). While this binding is likely responsible for the association of band 4.2 with the cytoplasmic surface of the membrane, band 4.2 also has other, lessunderstood associations. We have shown that band 4.2 can bind to purified erythrocyte ankyrin in solution with a $K_d \approx$ 10^{-7} M and may also associate with band 4.1 and spectrin (4). Binding measurements and competition studies suggest that ankyrin and band 4.2 bind to distinct sites on the cytoplasmic domain of band 3 (4). While we and others have not found any effect of band 4.2 on ankyrin association with erythrocyte membranes (4-6), other studies suggest that band 4.2 may stabilize ankyrin-band 3 associations (7).

While the exact function of band 4.2 on the erythrocyte membrane is not known, individuals whose erythrocytes lack or are deficient in band 4.2 have hemolytic anemia associated with spherocytic or elliptocytic erythrocytes (7, 8). Absence of band 4.2 associated with spur or target erythrocytes has also been reported (9). These findings suggest that band 4.2 probably plays an important role in the regulation of erythrocyte shape and mechanical properties. Band 4.2 is of additional interest because analogues of this protein have been detected in a wide variety of nonerythroid cells (10).

Here we report the complete amino acid sequence of human erythrocyte band 4.2 obtained from a reticulocyte cDNA expression library.[¶] Of particular interest is that the sequence shows homology with two well-characterized enzymes—guinea pig liver transglutaminase (protein-glutamine γ -glutamyltransferase; protein-glutamine:amine γ -glutamyltransferase, EC 2.3.2.13) and the a subunit of human factor XIII. Both of these are transglutaminase-type enzymes, which catalyze the formation of covalent isopeptide bonds via γ -glutamyl- ε -lysine linkages. Our results show that band 4.2 has extensive homology with these enzymes in the vicinity of the active site. However, the substitution in band 4.2 of an alanine for a cysteine in the enzyme active site may be responsible for the lack of transglutaminase activity of band 4.2. These studies may shed light on the origin and function of both band 4.2 and intracellular transglutaminases.

MATERIALS AND METHODS

Reagents. Nucleotide sequencing was performed with Sequenase from United States Biochemical. Restriction enzymes were obtained from New England Biolabs. The vectors p-Gem3 and p-Gem4 and Erase-a-Base were from Promega. The Multiprime DNA labeling system was from Amersham. [³²P]dCTP and [¹⁴C]putrescine dihydrochloride were obtained from New England Nuclear.

Screening of λ gt11 Library. A human reticulocyte cDNA library made in the expression vector λ gt11 constructed by John Conboy (11) was screened as described by Huynh *et al.* (12) with a band 4.2 antibody (4) affinity purified by the method of Olmstead (13). Positive plaques were visualized by an avidin-biotin conjugated second antibody enhancement system, Vectastain (Vector Laboratories). The positive clone designated 18 was labeled with [³²P]dCTP by primer extension using the Multiprime DNA labeling system and was used to screen the cDNA library a second time. Positive plaques were identified by autoradiography. Phage DNA was subcloned in the plasmid vectors p-Gem3 and p-Gem4.

Sequence Analysis. Cesium chloride banded plasmid DNA was sequenced by the dideoxynucleotide chain-termination method (14) using a Sequenase DNA sequencing kit. Exonuclease III deletions were made using Erase-a-Base.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

[§]To whom reprint requests should be addressed.

[¶]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M29399).

RNA and DNA Blot Analysis. Human reticulocyte total RNA was prepared by the method of Goosens and Kan (15). RNA was electrophoresed on a 1.2% formaldehyde agarose gel with Mops (4-morpholinepropanesulfonic acid) running buffer, transferred to nitrocellulose (16), and hybridized with cDNA probe 1A (see Fig. 1). Chromosomal DNA from human peripheral blood lymphocytes was prepared by the procedure of Sykes (17), digested with a series of restriction enzymes and electrophoresed on a 0.7% agarose gel. The DNA was denatured and transferred to nitrocellulose (18) and hybridized as described for RNA.

Assay for Ca²⁺-Stimulated Transamidation. Erythrocyte ghosts and inside-out vesicles (which contain normal amounts of band 4.2) and pH 11-stripped inside-out vesicles (which lack band 4.2 and other peripheral membrane proteins) were prepared as described by Korsgren and Cohen (4) and tested for transamidating (transglutaminase) activity by two procedures. (i) The method of Lorand et al. (19) was used to measure transglutaminase-catalyzed covalent crosslinking of a radioactive amine-rich protein (in this case ¹⁴C]putrescine) into membrane proteins. Erythrocyte membranes at a concentration of 1.0 mg/ml were incubated for 90 min at 37°C in the presence of 10 mM CaCl₂/20 mM dithiothreitol/90 mM Tris·HCl, pH 7.5/leupeptin (10 μ g/ml)/0.5 mM EGTA (cross-linking buffer) with 0.8 mM [¹⁴C]putrescine (10⁴ cpm/ μ mol). In some experiments, guinea pig liver transglutaminase was added as a positive control. The reaction was terminated by washing the membranes three times in 5 vol of 10 mM Tris·HCl (pH 7.5) at 4°C and the sedimented membranes were counted for ¹⁴C in a Beckman LS 5000 TD with Aquasol-2 (New England Nuclear). (ii) In the presence of Ca^{2+} , transglutaminases can induce the formation of covalently cross-linked aggregates of erythrocyte membrane proteins (20–22). To determine whether erythrocyte membranes containing or lacking band 4.2 evidenced such activity, ghosts or inside-out vesicles (which contain band 4.2) or pH 11-stripped vesicles (which lack band 4.2) were incubated in cross-linking buffer as described above. These samples were then solubilized in SDS and electrophoresed in a polyacrylamide gel in the discontinuous buffer system of Laemmli (23), stained with Coomassie blue, and scanned for the presence of protein aggregates. In some instances, the electrophoresed membrane proteins were transferred to nitrocellulose and incubated with antiserum to erythrocyte band 3, ankyrin, band 4.1, or spectrin to provide a more sensitive detection of any cross-linked complexes that might have formed. Immunoreactive protein bands were visualized with a peroxidase-conjugated second antibody and 4chloro-1-naphthol.

Sequence Analysis of Band 4.2 Peptides. Band 4.2 was extracted from high salt (1 M KCl)-stripped inside-out vesicles (see above) by alkaline extraction (pH \approx 13) and was further purified by HPLC gel filtration in 200 mM Tris·HCl/8 M urea/10 mM 2-mercaptoethanol, pH 7.0. Purified protein

was digested with endoproteinase Lys-C (Boehringer Mannheim) at an enzyme/substrate ratio of 1:200 (wt/wt) for 5 hr at 37°C in 100 mM Tris·HCl/1.5 M urea/0.06% SDS/2.5 mM EDTA/2.5 mM glycine/0.005% azide, pH 7.7. Digestion was terminated by titrating the sample to pH 2 and peptides were separated by reverse-phase HPLC on a C18 column (4.6 \times 250 mm) using a gradient of acetonitrile in 0.1% trifluoroacetic acid. Peptides were quantitated by amino acid analysis and sequences were determined by automated Edman degradation.

RESULTS

Isolation and Characterization of Erythrocyte Band 4.2 cDNA. Erythrocyte band 4.2 cDNA was isolated from a human reticulocyte cDNA library constructed in the expression vector λ gt11. Three hundred thousand plaque-forming units were initially screened with an affinity-purified antibody to band 4.2. One positively reacting clone (number 18, nucleotides 650–2348; Fig. 1) was verified as an authentic cDNA to band 4.2 by matching peptide amino acid sequence with nucleotide-derived sequence. This cDNA was used as a probe to rescreen another 300,000 plaque-forming units from the same library, resulting in the identification of another 33 positively reacting clones. One full-length 2.35-kilobase (kb) cDNA (designated 1A) was sequenced. The strategy for sequencing included subcloning from the major restriction sites and making a series of exonuclease III deletions (Fig. 1).

The authenticity of this cDNA clone was confirmed by matching amino acid sequences obtained from four separate peptides, derived from purified human erythrocyte band 4.2, with the predicted amino acid sequence of cDNA 1A (Fig. 2). Clone 1A contained 2348 base pairs (bp) and had an open reading frame from nucleotide 188 to 2261, representing 691 residues and a protein molecular weight of 76,981. This is in good agreement with the molecular mass of band 4.2 observed on denaturing SDS/polyacrylamide gels of 72 kDa (1). The nucleotide sequence contains a consensus sequence for initiation of translation as described by Kozak (24) at nucleotides 185-191, ACCATGG. When cDNA 1A was used to probe an RNA blot of total human reticulocyte RNA, a single band corresponding to a transcript size of 2.4 kb was found (data not shown). Enzyme digestion of human genomic DNA and analysis on a Southern blot (18) using full-length 1A cDNA as a probe gave an approximate size of 17 kb for the band 4.2 gene (data not shown).

Analysis of Derived Amino Acid Sequence. Nucleotide and primary amino acid sequence for erythrocyte band 4.2 was analyzed with the DNASTAR computer program. The predicted isoelectric point for the 691 amino acids of band 4.2 is 7.9 and the predicted molecular weight is 76,981 (Table 1). There are 37% hydrophobic residues and 28% polar residues. Band 4.2 does not show any obvious repeating primary structure, and computer generated secondary structure anal-



FIG. 1. (A) Restriction map of the full-length cDNA clone 1A showing the major enzyme restriction sites. Shaded area represents the coding sequence. (B) Sequence strategy for overlapping cDNA clones 1A and 18.

	С	GAA	GAA	ACA	TGT	CAG	GGT	GCT	CAC	AGG	AGT	AGT	GGG	GGG	AGG	TTT	TGC	: 49	TGG	ATG	ACG	CGG	CCT	GCC	TTG	222	CAG	GGT	TAT	GAT	GGA	TGG	CAG	ATT	CTC	GAC	1237
TAT	TTC	CAG	ATT	CTT	AAG	CCA	ACA	AAA	GTG	CCT	TCA	TAT	III	CTG	TCT	GG/	AGA	103	W	M	I	R	P	A	L	P	9	G	I.	D	G	¥	9	1	L	0	350
CAG	AAA	GCC	CAG	AAG	GAG	CCC	AGA	AGC	AAC	AGT	TTG	AGA	GAG	GCG	CTT	CIG	CGG	15/	CCA	AGT	GCT	CCI	AAI	GGA	GGT	GGA	GIC	CIG	GGG	ICC	161	GAI	CIG	GIG	CCG	GIC	1291
CCA	AGT	GGA	TAA	GAG	GAG	CGG	CCT	GCA	ACC	ATG	GGA	CAG	GCC	CIG	GGT	ATC	; AAG	i 211	P	S	A	2.0	N	6	6	6	V COO	L	6	5	500	U	L	V.	P	***	17/5
		~ • •		~ • •		~~~				M	G	Q	A	L	G	1	K	245	AGA	GCA	GIL	AAG	GAG	666	ALL	616	666	LIG	ALL	LLA	LCA	516	LCA	UAL	- CIT	111	1345
AGU	161	GAU	111	CAG	GLA	GLA	AGA	AAL	AAT	686	GAG	LAL	LAL	ALL	AAG	4		205	ССТ	A	А.Т.А.	A	500	TCA	ter	CTC.	CTC.	TCC	hac	TCC	TOT	CAC	CAT	000	ACA.	CTC.	1300
3	tre			CTC.	*	CTC.	ACC	ACC	acc.	E AC	500	110	- Acc	ATC	ATC	ÊTC	140	310		A	1	NA1		c	r 01	v	v	000	K NAG	ruc.	C	E	0.41	600		1	404
c	100		D	1	111	v	P	P	000	0	D	110	T	1	1	1	Ŷ	44	242	îte	ACT	GAC	fre	AAC	ACA.	ÅÅG	TAT	πī	ົງຄຸດ	ĂAC.	ĂAC.	ĂTC.	204	ACC.	ÅAG	Taa	1453
ŤTC	777	TĴa -	<u> 77</u>	<u> 7 7 8 -</u>	tán -	- GCA	TÎT	र्गेट	сčт	220	อรัว	AAG	AAG	ata	222	cīc	ACT	373	F	1	ĩ	D	s	N	T	K	Ý	v.	G	N	N	ĩ	S	T	ĸ	G	422
F	R	A .	P	v	R	A	F	ĩ	P	A	L	ĸ	ĸ	v	Ă	L	T	62	ĞΤG	GGC	ÅGT	GAC	ČGC	ŤGC	ĠAG	GAC	ÁTC.	Á CT	ČAG	AAC	TAC	ÅÅG	ŤAT	CCT	GAA	GGG	1507
GCA	CAA	ACT	GGA	ĠAG	CAG	CCT	İCC	ĀAG	ATC	AAC	ĀGG	ACC	CAA	ĠCC	ACA	ŤTC	ĊCA	427	v	G	S	D	R	Ċ	E	D	I	T	Q	N	Y	K	Ŷ	P	E	G	440
Ă	0	T	G	E	9	P	S	K	1	N	R	T	Q	A	T	F	P	80	TCT	CTT	CAG	GAA	AAA	GAG	GTG	CTG	GAG	AGA	GTC	GAG	AAA	GAG	AAA	ATG	GAA	CGT	1561
ATT	TCC	AGT	CTG	GGG	GAC	CGA	AAG	TGG	TGG	AGT	GCA	GTG	GTG	GAG	GAG	AGA	GAT	481	S	Ł	٩	E	ĸ	Ε	٧	L	E	R	٧	E	ĸ	E	ĸ	M	Ε	R	458
I	S	S	L	G	D	R	ĸ	Μ	W	S	A	۷	v	E	Ε	R	D	98	GAG	AAA	GAC	AAC	GGC	ATC	CGT	CCT	CCC	AGT	CTC	GAG	ACT	GCC	AGT	ССТ	CTG	TAC	1615
GCC	CAG	TCC	TGG	ACC	ATC	TCT	GTG	ACC	ACA	CCT	GCG	GAC	GCT	GTC	ATT	GGC	CAC	535	E	κ	<u>p</u>	N	G	1	R	<u>P</u>	P	<u>s</u>	L	E	T_	<u> </u>	<u>s</u>	<u>P</u>	L	<u>Y</u>	476
A	Q	S	W	T	I	S	v	T	T	P	Α	D	A	V	I	G	H	116	CTG	CTC	TTG	AAA	GCA	CCC	AGC	TCC	CTA	CCC	CTG	AGA	GGG	GAT	GCC	CAG	ATC	TCA	1669
TAC	TCG	CTT	CTG	CTG	CAG	GTC	TCA	GGC	AGG	AAG	CAA	CTC	CTC	TTG	GGT	CAG	TIC	589	L.	L	L	ĸ	A	P	S	S	L	P	L	R	G	D	A	9	I	S	494
Y.	S	L	L	L	Q	V.	S	G	R	ĸ	0	L	L	L	G	Q	F	154	GIG	ACG	CIG	GIT	AAT	CAC	AGT	GAG	CAG	GAG	AAG	GCA	GIG	CAG	CIG	GCA	ΑH.	GGG	1/25
ACA	CIG	Ç11	ÎH	AAC	CCC.	IGG	AAT	AGA	GAG	GAT	GCI	GII	inc.	CIG	AAG	AAI	GAG	043	V		L	V.	N	H	S	E	Q ATA	E .	K	A	¥	u ato	L	A CC	1	5	1777
	L	L	P ATC	N	7.0	WTC	NTTC	K	E.C.	0	A	V.T.C	ATC	L.C.	K	N	E.	102	616	LAG		UIA.	LAL	JAL .	AAL					4	AAG	LIL .	166	AUU	AAG	AAG	530
4	CAG	LUL	AIG	646	VAL	110	110	AAL	CAG	AA I	001	LIC.	ALC.	VAL	1	661	T	170	ČTC.	č.	ALC:	ACC	CTC	ACT	and a	AAC	CTC		AAC .	A.T.A.	Ата	Acc	ATC	CCC	CTC.	TTC.	1931
CCT.	GAC	TCC	ATC	CAC.	cc.	CAC	fre	TCC	GAC.	TTT	000	CAC	ite	cac.		CAT	άτς.	751	1	unc.	1	T	1	6	A	N		E .	r		1 .	T	1	600		F	548
4	D D	ruc	1	0	A	F	5	0	n	È.	600	0	F	F	600	D. 1	v	188	717	ücc.	ÅAT.	τ́ττ	o a o	ADD	ÂAC .	A 11	ČCC .	GAG .	AAC .	ACC .	Ťτc	Ċττ .	ÅGA -	čτc	ACC.	228	1885
ÂTT	GAC	čīc	ÅGC	DTD	222	חדד	οτ <u>ο</u>	AGC	AAG	DAD	AAG	CAG	GTA	DA	AAG	201	ÅGC	805	F	ŝ	N	Ê.	E	R	N	P	P	Ē	N	T	F	Ľ Í	R	Ľ	Ť	Ă	566
î	D	ĩ.	S	L	R	i.	L.	S	ĸ	D	ĸ	0	v	Ē	ĸ	ý.	S	206	ÂTG	ĞCA	ÄCA	ĊAC	TCT	GAA	ŤCC	AAC	СТТ	AGC	TGC	İΤΤ	GCT (CAG I	GAA I	GAC	ATT	GCC	1939
ĊAG	CCG	G TG	ČAC	ĞΤG	GCC	ČGT	G TG	ŤTG	GGT	GCC	TTG	CTG	CAT	ĪTT	CTC.	ÄAG	ĞAG	859	M	A	T	H	S	E	S	N		S I	Ċ	F.	A I	Q	E I	D	I	Ă	584
0	P	V.	H	Ŷ.	A	R	V.	Ĺ	G	Â	Ĺ	Ĺ	H	F	Ĺ	K	E	224	ATT	TGT	AGA	CCA	CAC	CTT	GCC .	ATC /	AAG	ATG	CCA	GAG .	AAA	GCA	GAG	CAG	TAT	CAA	1993
CAG	AGG	GTC	CTG	CCC	ACC	CCG	CAG	ACC	CAG	GCC	ACC	CAG	GAA	GGG	GCC	TTG	CTG	913	I	С	R	Ρ	H	L.	A	II	(I	4 1	P I	ΕI	ĸ	A I	Εı	Q	Y .	Q	602
Q	R	٧	L	Ρ	T	P	Q	T	٩	A	T	Q	E	G	A	L	L	242	CCC	CTC	ACA	GCC	TCA	GTC .	AGC	CTC	CAG	AAC 1	TCC	CTA	GAT	GCC (CCC /	ATG	GAG	GAC	2047
AAC	AAG	CGC	CGG	GGC	AGC	GTG	CCC	ATC	CTG	CGG	CAG	TGG	CTC	ACC	GGC	CGA	GGC	967	Р	L	T	A	S	V	S	L (2 1	N :	S	LI	D	A I	P	M	E	D	620
N	ĸ	R	R	G	s	V	P	1	L	R	Q	W	L	Ţ	G	R	G	260	TGT	GTG	ATC .	TCC	ATC	CTG	GGA	AGG	GGG	CTC /	ATT	CAC	AGA	GAG /	AGG /	AGC	TAC .	AGA	2101
CGA	CCT	GTG	TAT	GAT	GGC	CAG	GCC	TGG	GTG	TTG	GCT	GCT	GTT	GCT	IGC	ACA	GTG	1021	C	V AAT	1	S	1		G	R	5			H	RI	E I	?	S	I an	R	038
R	P	V.	Y	D	G	9	A	W.	V.	L	A .	A	¥	A	C	1	¥	2/8	inc.		TCA	GIG	166		GAA I	AAC /		IG.	GI		AAG		AG		ACG	CCA	2122
CIG	CGA	ICC	CIG	GGA	AIC.	CC1	GCC	CGC	616	GIG	ACC	ACG	Į.	GCC	ICA	GCA	CAG	10/5	F.C.A.	K CAT	5	V CCC	W CTC			N CTC							ine i	TC -		CAC .	2200
L	K	COT	L	GOT		P	<u>.</u>	K CAT	V CAA	V.	1		F.	A	S	A .	Q.	1120	ALA		616	666			AGA I			, 16 (GAA (BAL			AIG		LAG .	47/
000	T	001	000				1	0.41	GAA C	V			5	5	C ADD		0	316	AAC .	CTA	ÅCC .	AAC .	TAT		ion i	ато и		тс и	TA I		έcτ λ		TA 1		CCT -	TAA '	2263
ÅÅC.	400	CAA	222	ČAG	A CA	cor	ÅCA	ATC	100	ATC	tte l		ACT	Trr	ACA.	ČAC.	701	1183	N		T	NUC .	v~ 1	K (1 1		/ \						5	A 1		692
N	C	E	C	0	D	600	D D	1	000	1	5	0	T	, LL C	T	E		332	Аст	Trr	Å CC	їст .	ite l	in i	Íct i	70 0	TC	CA I	ice i	••••	TT C	TA	-	ře í		. Ann	2317
	9	-		•	n	9	n .	•	-	•				5	•	-	U	336	AAC .	ATC .	TCC	tec 1											ma			con i	

FIG. 2. Nucleotide and derived amino acid sequence for erythrocyte band 4.2 full-length cDNA. Nucleotides and amino acids in (single-letter code) are numbered on the right. Amino acid residues matched by peptide analysis are underlined.

ysis of the primary amino acid sequence suggests characteristics of a globular protein. There are no extended stretches of β -sheet or α -helix; rather, the protein is characterized by

Table 1. Amino acid composition and properties of band 4.2 and related proteins

		Guinea pig	
		liver	Factor
	Band	trans-	XIII a
Amino acid	4.2	glutaminase	subunit
Leucine	79	69	48
Arginine	43	40	45
Glutamic acid	45	50	48
Glutamine	41	27	27
Valine	49	60	70
Threonine	43	31	45
Alanine	56	45	37
Serine	45	43	46
Lysine	32	29	38
Isoleucine	31	34	39
Phenylalanine	24	23	32
Asparagine	30	37	40
Proline	34	32	33
Glycine	47	57	50
Aspartic acid	26	41	47
Tyrosine	16	24	29
Tryptophan	14	10	15
Histidine	12	12	14
Cysteine	15	17	9
Methionine	9	9	19
Total amino acids	691	690*	731†
pI [‡]	7.96	4.92	5.88
Calculated molecular			
mass, kDa	76.9	76.6*	83.1 [†]
SDS molecular mass, kDa	72	77*	75†

Amino acid composition pI values and calculated molecular masses were computed from the deduced amino acid sequences obtained from the cDNA in each case. *Data from Ikura *et al.* (25).

^tData from Ikura *et al.* (25).

Data from Ichinose et al. (26).

[‡]Predicted pI values calculated from the amino acid composition.

short segments, interrupted with turn structures consistent with electron microscopic studies of band 4.2 showing a globular protein (3). A hydropathy plot of band 4.2 showed short alternating regions of hydrophobic and hydrophilic character; however, the region of the protein between amino acids 265 and 475 was characterized by two sets of alternating, prominent hydrophobic and hydrophilic domains (not shown).

The Swiss-Prot protein sequence data base (release 10) was searched for protein sequences homologous to that of band 4.2 by the method of Pearson and Lipman (27) using the FASTA program on Bionet (IntelleGenetics). By this method, band 4.2 was found to have homology with two related Ca²⁺-stimulated transamidating (transglutaminase type) enzymes, guinea pig liver transglutaminase (31.8% identity in a 446-amino acid overlap) and the a subunit of human coagulation factor XIII (26.9% identity in a 639-amino acid overlap). Fig. 3 shows the region of greatest homology among the proteins, comprising a 204-amino acid segment of band 4.2. The segment of band 4.2 between amino acids 247 and 295 is 69% identical with transglutaminase and 51% identical to factor XIII. This segment contains the active sites for both guinea pig liver transglutaminase and factor XIII (25, 26, 30; shaded region in Fig. 3). Of the five consensus residues contained within the active site of the enzymes, Gly-Gln-Cys-Trp-Val, band 4.2 is missing the cysteine residue, for which an alanine has been substituted. The substituted alanine is located exactly at the transition point between a short hydrophilic segment and the major hydrophobic domain of band 4.2 (not shown). Hydropathy plots of factor XIII and guinea pig liver transglutaminase (not shown) show that in both of these proteins the active cysteines are also placed at identical transition points at the start of a major hydrophobic domain. Fig. 3 shows that there is also a homology between band 4.2 and a region of the factor XIII a subunit which likely contains the site for Ca^{2+} binding that is required for factor XIII activation (30). At this time, it is not known whether band 4.2 binds Ca²⁺

The N-terminal amino acid sequence of band 4.2 predicted from the cDNA is Met-Gly-Gln-Ala-Leu-Gly. In the case of factor XIII a subunit and guinea pig liver transglutaminase, the terminal methionine is known to be cleaved (25, 26, 30),



FIG. 3. Amino acid sequence (single-letter code) homologies between erythrocyte band 4.2 (4.2), guinea pig liver transglutaminase (TGL), and the a subunit of human factor XIII (F 13). Sequences were aligned and gapped using the ALP 3 program (Bionet version 1.0, IntelleGenetics). Amino acid identities are boxed, and the shaded area represents the active site regions of guinea pig liver transglutaminase (TGL) and factor XIII. The underlined region of factor XIII sequence denotes the Ca^{2+} binding domain of this protein (29). The sequence numbering for transglutaminase was taken from Ikura *et al.* (31) and that for factor XIII (the a subunit) is from Takahashi *et al.* (29).

but we have no direct evidence that this is the case for band 4.2. Several attempts to determine the authentic N-terminal amino acid sequence from purified band 4.2 were unsuccessful, suggesting that the N terminus is blocked. The penultimate glycine and adjacent amino acids are consistent with the consensus sequence recognized by N-myristoyltransferase (28, 31), suggesting that band 4.2 may be cotranslationally myristoylated at this glycine residue. This feature may account for some of the unusual solubility and membrane binding properties of the protein (3, 29).

The cysteine residue in the active site is required for Ca²⁺-stimulated transamidating activity by the two transglutaminases (32, 33), and its absence in band 4.2 suggests that the protein would lack this enzymatic activity. To determine whether this was the case, we tested band 4.2 both in solution and on the erythrocyte membrane for Ca²⁺-stimulated crosslinking activity. We tested erythrocyte ghosts and inside-out vesicles (both of which contain band 4.2) and pH 11-stripped inside-out vesicles (which lack band 4.2) for transglutaminase activity using [¹⁴C]putrescine incorporation. In this assay, transglutaminase activity is detected as the incorporation of ¹⁴Clputrescine into ervthrocyte membrane proteins, principally band 3 (19, 34). Neither the membrane preparations that contained endogenous band 4.2 (ghosts and inside-out vesicles) nor those lacking band 4.2 (pH 11-stripped inside-out vesicles) showed any evidence of [¹⁴C]putrescine incorporation. As a positive control, addition of purified guinea pig liver transglutaminase to the membranes led to significant [¹⁴C]putrescine incorporation into membrane proteins (3.5 μ mol per mg of membrane protein).

An additional test was done to determine whether band 4.2 might catalyze Ca^{2+} -dependent cross-link formation between erythrocyte membrane proteins, as has been observed for endogenous erythrocyte transglutaminase (20–22). Analysis of ghosts, inside-out vesicles, or pH 11-stripped vesicles

revealed no cross-link formation in the presence or absence of Ca^{2+} . As a positive control, purified guinea pig liver transglutaminase was added to ghosts at a molar concentration equal to that of the endogenous band 4.2. In the presence of this enzyme, >90% of the band 3 molecules formed high molecular weight aggregates.

Finally, soluble band 4.2 purified by a nondenaturing method (29) was tested for the presence of transglutaminase activity in solution by the method of Lorand *et al.* (35). Once again, no Ca^{2+} -stimulated cross-linking activity could be detected (E. Dotimas, C.K., and C.M.C., unpublished data).

DISCUSSION

Erythrocyte band 4.2 is tightly associated with the inner surface of the erythrocyte membrane. Because of its high copy number-200,000 per cell (1)-and the fact that its absence or deficiency has been associated with some cases of hemolytic anemia (7-9), it is likely that band 4.2 has a significant role in erythrocyte viability. The major membrane binding site for band 4.2 is the cytoplasmic domain of the erythrocyte anion transporter, band 3 (3, 4), and we have previously shown that band 4.2 also interacts with erythrocyte ankyrin in solution. Though ankyrin is also associated with the membrane through attachment to the cytoplasmic domain of band 3, ankyrin and band 4.2 appear to have distinct binding sites on this domain (4). This paper describes the complete amino acid sequence for erythrocyte band 4.2, which will facilitate assignment of structural and functional domains within the protein.

Our results show that band 4.2 has significant homology with two transamidating enzymes—human factor XIII (the a subunit) and guinea pig liver transglutaminase. The greatest degree of identity to both enzymes includes the active site for Ca^{2+} -stimulated cross-linking activity. In the five residues of

the consensus sequence for the active site, band 4.2 is missing the active cysteine, for which an alanine has been substituted. This is likely responsible for the lack of observable Ca^{2+} -stimulated transamidating activity of band 4.2. The homology between the proteins is significant enough that affinity-purified antibodies to ervthrocyte band 4.2 crossreact with purified guinea pig liver transglutaminase (data not shown: we do not know vet whether the reverse is true). This observation is of particular significance for studies of the presence of band 4.2 analogues in nonerythroid cells. Transglutaminases are widely distributed in many different nonerythroid cell types and are found both in the cytosol and in association with membranes (36). Thus, it is essential to document that antibodies used to screen for nonerythroid forms of band 4.2 do not also cross-react with tissue transglutaminase enzymes. In spite of the strong sequence homology between band 4.2 and the transglutaminases, the relationship between the proteins is unclear. Although band 4.2 lacks transglutaminase activity, the similarities in secondary structure around the active site suggest that the homologies between these proteins may reflect common structural features such as binding sites for other proteins. For example, recently, transglutaminase has been reported to form a complex with human plasma fibronectin (37). Using a similar approach, we have found in preliminary studies evidence for an interaction of band 4.2 with some as yet unidentified components in human plasma. Furthermore, some transglutaminase enzymes have been reported to be membrane associated (36), suggesting that they may share with band 4.2 a common recognition sequence for a band 3-like (or even ankyrin-like) membrane binding site. Confirmation of this idea will need to await the identification of the band 3 and ankyrin binding domains within band 4.2.

Note Added in Proof. A sequence for human erythrocyte band 4.2 has also been obtained by Sung *et al.* (38) and is substantially the same as the sequence presented here.

The authors would like to thank Ms. Lucille Paul for preparation of the manuscript and Ms. Joan Joos for preparing the figures. We also thank Dr. John Conboy for the use of his λ gt11 library. This work was supported by National Institutes of Health Grants HL 36090 and HL 24382 (C.M.C.) and HL 38794 (D.S.).

- 1. Steck, T. L. (1974) J. Cell Biol. 62, 1-19.
- 2. Steck, T. L. (1972) J. Mol. Biol. 66, 295-305.
- Korsgren, C. & Cohen, C. M. (1986) J. Biol. Chem. 261, 5536–5543.
- Korsgren, C. & Cohen, C. M. (1988) J. Biol. Chem. 263, 10212–10218.
- Hargreaves, W. R., Giedd, K. N., Verkleij, A. & Branton, D. (1980) J. Biol. Chem. 255, 11965-11972.
- Bennett, V. & Stenbuck, P. (1980) J. Biol. Chem. 255, 6424– 6432.

- Rybicki, A. C., Heath, R., Wolf, J. L., Lubin, B. & Schwartz, R. S. (1988) J. Clin. Invest. 81, 893-901.
- Hayashi, S., Koomoto, R., Yano, A., Ishigami, S., Tsujino, G., Saeki, S. & Tanaka, T. (1974) Biochem. Biophys. Res. Commun. 57, 1038-1044.
- 9. Iida, H., Hasegawa, I. & Nozawa, Y. (1976) Biochim. Biophys. Acta 443, 394-401.
- Friedrichs, B., Koob, R., Kraemer, D. & Drenckhahn, D. (1989) Eur. J. Cell Biol. 48, 121–127.
- Conboy, J., Kan, Y. W., Shohet, S. B. & Mohandas, N. (1986) Proc. Natl. Acad. Sci. USA 83, 9512–9516.
- Huynh, T. Y., Young, R. A. & Davis, R. W. (1985) in "DNA Cloning: A Practical Approach", ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49-78.
- 13. Olmstead, J. B. (1981) J. Biol. Chem. 256, 11955-11957.
- 14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 15. Goosens, M. & Kan, Y. W. (1981) Methods Enzymol. 76, 805-817.
- 16. Thomas, P. S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
- 17. Sykes, B. G. (1983) Lancet II, 787-788.
- 18. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
- Lorand, L., Shishido, R., Parameswaran, K. N. & Steck, T. L. (1975) Biochem. Biophys. Res. Commun. 67, 1158-1166.
- 20. King, L. E. & Morrison, M. (1977) Biochim. Biophys. Acta 471, 162–168.
- Lorand, L., Weissmann, L. B., Epel, D. L. & Bruner-Lorand, J. (1976) Proc. Natl. Acad. Sci. USA 73, 4479–4481.
- 22. Lorand, L., Siefring, G. E. & Lowe-Krente, L. (1978) J. Supramol. Struct. 9, 427-440.
- 23. Laemmli, U. K. (1970) Nature (London) 227, 680-685.
- 24. Kozak, M. (1986) Cell 44, 283-292.
- Ikura, K., Nasu, T., Yokota, H., Tsuchiya, Y., Sasaki, R. & Chiba, H. (1988) *Biochemistry* 27, 2898–2905.
- Ichinose, A., Hendrickson, L. E., Fujikawa, K. & Davie, E. W. (1986) *Biochemistry* 25, 6900-6906.
- 27. Pearson, W. R. & Lipman, D. J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
- 28. Towler, D. A. & Gordon, J. I. (1988) Annu. Rev. Biochem. 57, 69-99.
- Dotimas, E., Korsgren, C., Anderson, L. & Cohen, C. M. (1989) J. Cell Biol. 109, 2652.
- Takahashi, N., Takahashi, Y. & Putnam, F. W. (1986) Proc. Natl. Acad. Sci. USA 83, 8019-8023.
- Towler, D. A., Eubanks, S. R., Towery, D. S., Adams, S. P. & Glaser, L. (1987) J. Biol. Chem. 262, 1030-1036.
- 32. Lorand, L. & Conrad, S. M. (1984) Mol. and Cell. Biochem. 58, 9–35.
- 33. Folk, J. E. (1980) Annu. Rev. Biochem. 49, 517-531.
- Dutton, A. & Singer, S. J. (1975) Proc. Natl. Acad. Sci. USA 72, 2568–2571.
- Lorand, L., Campbell-Wilkes, L. K. & Cooperstein, L. (1972) Anal. Biochem. 50, 623-631.
- 36. Fesus, L. & Thomazy, V. (1988) Adv. Exp. Med. Biol. 231, 119-134.
- Lorand, L., Dailey, J. E. & Turner, P. M. (1988) Proc. Natl. Acad. Sci. USA 85, 1057–1059.
- Sung, L. A., Chien, S., Chang, L. S., Lambert, K., Bliss, S., Bouhassira, E., Nagle, R., Schwartz, R. & Rybicki, A. (1990) *Proc. Natl. Acad. Sci. USA* 87, 955–959.