

Molecular cloning and expression of a multiubiquitin chain binding subunit of the human 26S protease

Katherine Ferrell^a, Quinn Deveraux^a, Steven van Nocker^b, Martin Rechsteiner^{a,*}

^aDepartment of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84132, USA

^bDepartment of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA

Received 7 December 1995; revised version received 21 January 1996

Abstract S5a is a subunit of the 26S protease that binds and presumably selects multiubiquitinated proteins for destruction. We recently identified an *Arabidopsis* protein, MBP1, that is physically, immunologically and biochemically similar to S5a from the human erythrocyte 26S protease. Based upon the MBP1 cDNA sequence we have now isolated a HeLa cell cDNA coding for human S5a. The HeLa cDNA sequence is highly similar to MBP1 and it encodes peptides obtained directly from human erythrocyte S5a. Moreover, expression of the isolated cDNA in *E. coli* results in a recombinant protein with an apparent molecular mass and multiubiquitin binding properties that match those of human S5a obtained from the purified 26S enzyme.

Key words: Human 26S protease; Multiubiquitin chain binding subunit; Antisecretory factor

1. Introduction

Ubiquitin mediated proteolysis is a selective and well controlled pathway involved in regulating many cellular processes such as cell cycle progression [1,2], transcriptional activation [3] and plant photoreceptor response [4]. The pathway is characterized by covalent addition of multiubiquitin chains to substrate proteins [5,6], a reaction that targets the modified protein for destruction by the 26S ATP-dependent protease [7]. This large multisubunit enzyme is assembled from two sub-complexes; the well characterized multicatalytic protease (MCP) or proteasome [8,9] and a regulatory complex [10–13]. MCP confers proteolytic activity to the 26S enzyme, and the regulatory complex confers ATPase activity as well as the ability to degrade specific protein substrates [7,14].

The 26S protease regulatory complex is composed of at least 15 distinct subunits with molecular masses ranging from 25 to 110 kDa. They can be divided into a family of ATPases, comprising subunits S4, S6, S7 and S8 [15,16] and possibly two additional subunits. Eight other components (S1 or SEN3, S2 or TRAP2, S3, S5a, S5b, S10, S12, S13/14 or NIN1) are likely to be involved in substrate selection or in maintaining the structural integrity of the regulatory complex (refs. [15–21]). In 1994, we identified a 50 kDa component of the regulatory complex that binds multiubiquitinated lysozyme and free multiubiquitin chains [22]. The protein was called subunit 5 (S5) based upon its migration on SDS polyacrylamide gels. Subsequent studies demonstrated that two proteins are present in the band designated S5 and that the more acidic member of the pair (S5a) binds multiubiquitin chains [19]. Recently we have identified a multiubiquitin chain

binding protein (MBP1) from *Arabidopsis* that by numerous criteria is homologous to human S5a [23]. Based on our knowledge of the *Arabidopsis* MBP1 sequence, we designed PCR primers for isolating human S5a. Here we present the sequence of a HeLa cDNA that encodes peptides obtained directly from S5a of the human 26S protease. Furthermore, we show that expression of the cDNA in *E. coli* results in a recombinant protein with biochemical properties matching those of the human 26S protease subunit.

2. Materials and methods

2.1. Isolation of a human cDNA encoding S5a

The DNA sequence of *Arabidopsis* MBP1 [23] was used to search the Genbank and EMBL databases using the tblastn program. This search produced matches to several human expressed sequence tagged clones (H83202, H24730, HUML13720, HSB34E112, T97122, T30391, R58821, T97007, T30383, T25023, H24731, H82970, HUMGS01820, R95474). Because a conceptual translation product of clone HUML13720 (EMBL D31409) corresponded almost exactly to the N-terminus of *Arabidopsis* MBP1, a 54 nucleotide oligomer matching the HUML13720 DNA sequence was designed. Using standard techniques the synthetic oligonucleotide probe was used to screen a lambda ZAPII HeLa cDNA library (Stratagene) [24]. Two positive clones were identified and the isolated cDNA inserts were sequenced on both strands (Sequenase protocol, U.S. Biochemical Corp.). Sequence alignments and protein analysis was made using the DNASTAR program. The *Arabidopsis* MBP1 sequence and sequences of peptides obtained directly from purified human erythrocyte 26S protease S5a have been described in refs. [23] and [19], respectively.

2.2. In vivo expression of recombinant S5a and multiubiquitin chain binding assays

PCR was used to introduce a *NdeI* restriction site at the methionine initiating the open reading frame of the isolated cDNA. This construct was subcloned into pAMP1 using the CloneAMP system (Life Technologies Inc.). The cDNA was then excised from the pAMP vector at the *NdeI* and *HindIII* sites and ligated into an expression vector, pET16b for a decahistidine N-terminal fusion with S5a or pAED4 for a non-histidine 'tagged' version [19]. This expression construct was sequenced to verify PCR fidelity. The vector alone or expression construct was used to transform *E. coli* BL21(DE3) cells. Protein expression at 30°C, resolution of bacterial lysate on 10% SDS-PAGE and transfer of proteins to nitrocellulose membranes was performed using published procedures [19]. Multiubiquitin chains were obtained from C. Pickart and radioiodinated using iodogen [22]. Purification of 26S protease regulatory complexes (RC), SDS-PAGE, binding and elution of [¹²⁵I]multiubiquitin chains to nitrocellulose bound proteins was performed as described previously [10,22].

3. Results

In an effort to identify a cDNA for S5a from humans, we used an *Arabidopsis* cDNA sequence (MBP1), encoding a protein homologous to human S5a [23], to search DNA databases. Although a number of human sequences with high

*Corresponding author. Fax: (1) (801) 581-7959.

AAT	TCC	CAA	ATG	ACC	TTT	TAT	TTC	ATA	CAG	AGA	TAC	AAA	GGC	AAC	TAT	GTG	CAG	CAA	CAA	60
TCT	GAT	GGG	CAG	TCC	AAA	CTC	TTG	GGA	GGA	AGT	AAA	TTC	ATG	GTA	AAT	GTC	ATG	ATG	GCG	120
GTC	GGG	AGG	GAG	GAA	GGT	GGC	AAG	ATG	GTG	TTG	GAA	AGC	ACT	ATG	GTG	TGT	GTG	GAC	AAC	180
								M	V	L	E	S	T	M	V	C	V	D	N	60
AGT	GAG	TAT	ATG	CGG	AAT	GGA	GAC	TTC	TTA	CCC	ACC	AGG	CTG	CAG	GCC	CAG	CAG	GAT	GCT	240
S	E	Y	M	R	N	G	D	F	L	P	T	R	L	Q	A	Q	Q	D	A	80
GTC	AAC	ATA	GTT	TGT	CAT	TCA	AAG	ACC	CGC	AGC	AAC	CCT	GAG	AAC	AAC	GTG	GGC	CTT	ATC	300
V	N	I	V	C	H	S	K	T	R	S	N	P	E	N	N	V	G	L	I	100
ACA	CTG	GCT	AAT	GAC	TGT	GAA	GTG	CTG	ACC	ACA	CTC	ACC	CCA	GAC	ACT	GGC	CGT	ATC	CTG	360
T	L	A	N	D	C	E	V	L	T	T	L	T	P	D	T	G	R	I	L	120
TCC	AAG	CTA	CAT	ACT	GTC	CAA	CCC	AAG	GGC	AAG	ATC	ACC	TTC	TGC	ACG	GGC	ATC	CGC	GTG	420
S	K	L	H	T	V	Q	P	K	G	K	I	T	F	C	T	G	I	R	V	140
GCC	CAT	CTG	GCT	CTG	AAG	CAC	CGA	CAA	GGC	AAG	AAT	CAC	AAG	ATG	CGC	ATC	ATT	GCC	TTT	480
A	H	L	A	L	K	H	R	Q	G	K	N	H	K	M	R	I	I	A	F	160
GTG	GGA	AGC	CCA	GTG	GAG	GAC	AAT	GAG	AAG	GAT	CTG	GTG	AAA	CTG	GCT	AAA	CGC	CTC	AAG	540
V	G	S	P	V	E	D	N	E	K	D	L	V	K	L	A	K	R	L	K	180
AAG	GAG	AAA	GTA	AAT	GTT	GAC	ATT	ATC	AAT	TTT	GGG	GAA	GAG	GAG	GTG	AAC	ACA	GAA	AAG	600
K	E	K	V	N	V	D	I	I	N	F	G	E	E	E	V	N	T	E	K	200
CTG	ACA	GCC	TTT	GTA	AAC	ACG	TTG	AAT	GGC	AAA	GAT	GGA	ACC	GGT	TCT	CAT	CTG	GTG	ACA	660
L	T	A	F	V	N	T	L	N	G	K	D	G	T	G	S	H	L	V	T	220
GTG	CCT	CCT	GGG	CCC	AGT	TTG	GCT	GAT	GCT	CTC	ATC	AGT	TCT	CCG	ATT	TTG	GCT	GGT	GAA	720
V	P	P	G	P	S	L	A	D	A	L	I	S	S	P	I	L	A	G	E	240
GGT	GGT	GCC	ATG	CTG	GGT	CTT	GGT	GCC	AGT	GAC	TTT	GAA	TTT	GGA	GTA	GAT	CCC	AGT	GCT	780
G	G	A	M	L	G	L	G	A	S	D	F	E	F	G	V	D	P	S	A	260
GAT	CCT	GAG	CTG	GCC	TTG	GCC	CTT	CGT	GTA	TCT	ATG	GAA	GAG	CAG	CGG	CAG	CGG	CAG	GAG	840
D	P	E	L	A	L	A	L	R	V	S	M	E	E	Q	R	Q	R	Q	E	280
GAG	GAG	CGC	CGG	CGG	GCA	GCT	GCA	GCT	TCT	GCT	GCT	GAG	GCC	GGG	ATT	GCT	ACG	ACT	GGG	900
E	E	R	R	R	A	A	A	A	S	A	A	E	A	G	I	A	T	T	G	300
ACT	GAA	GAC	TCA	GAC	GAT	GCC	CTG	CTG	AAG	ATG	ACC	ATC	AGC	CAG	CAA	GAG	TTT	GGC	CGC	960
T	E	D	S	D	D	A	L	L	K	M	T	I	S	Q	Q	E	F	G	R	320
ACT	GGG	CTT	CCT	GAC	CTA	AGC	AGT	ATG	ACT	GAG	GAA	GAG	CAG	ATT	GCT	TAT	GCC	ATG	CAG	1020
T	G	L	P	D	L	S	S	M	T	E	E	E	Q	I	A	Y	A	M	Q	340
ATG	TCC	CTG	CAG	GGA	GCA	GAG	TTT	GGC	CAG	GCG	GAA	TCA	GCA	GAC	ATT	GAT	GCC	AGC	TCA	1080
M	S	L	Q	G	A	E	F	G	Q	A	E	S	A	D	I	D	A	S	S	360
GCT	ATG	GAC	ACA	TCC	GAG	CCA	GCC	AAG	GAG	GAG	GAT	GAT	TAC	GAC	GTG	ATG	CAG	GAC	CCC	1140
A	M	D	T	S	E	P	A	K	E	E	D	D	Y	D	V	M	Q	D	P	380
GAG	TTC	CTT	CAG	AGT	GTC	CTA	GAG	AAC	CTC	CCA	GGT	GTG	GAT	CCC	AAC	AAT	GAA	GCC	ATT	1200
E	F	L	Q	S	V	L	E	N	L	P	G	V	D	P	N	N	E	A	I	400
CGA	AAT	GCT	ATG	GGC	TCC	CTG	GCC	TCC	CAG	GCC	ACC	AAG	GAC	GGC	AAG	AAG	GAC	AAG	AAG	1260
R	N	A	M	G	S	L	A	S	Q	A	T	K	D	G	K	K	D	K	K	420
GAG	GAA	GAC	AAG	AAG	TGA	GAC	TGG	AGG	GAA	AGG	GTA	GCT	GAG	TCT	GCT	TAG	GGA	CTG	CAT	1320
E	E	D	K	K	*															440
GGG	GGA	ATT	C																	1330

Fig. 1. Nucleotide and inferred amino acid sequence of a human cDNA that encodes subunit 5a of the human 26S protease. The amino acid sequence deduced from the S5a cDNA is shown in the one-letter code. Regions of the conceptual translation product that match peptides sequenced directly from erythrocyte S5a are shown in white on black.

Human	M V L E S T M V C V D N S E Y M R N G D F L P T R L Q A Q Q D A V N I V C H S K	40
Drosophila	M V L E S T M I S F D N S D F Q R N G D Y F P T R L I V Q R D G I N L V C L T K	40
Arabidopsis	M V L E A T M I C I D N S E W M R N G D Y S P S R L Q A Q T E A V N L L C G A K	40
S. cerevisiae	M V L E A T V L V I D N S E Y S R N G D F P R T R F E A Q I D S V E F I F Q A K	40
Human	T R S N P E N N V G L I T L A - N D C E V L T T L T P D T G R I L S K L H T V Q	79
Drosophila	L R S N P E N N V G L M T L S - N T V E V L A T L T S D A G R I F S K M H L V Q	79
Arabidopsis	T Q S N P E N T V G I L T M A G K G V R V L T T P T S D L G K I L A C M H G L D	80
S. cerevisiae	R N S N P E N T V G L I S G A G A N P R V L S T F T A E F G K I L A G L H D T Q	80
Human	P K G K I T F C T G I R V A H L A L K H R Q G K N H K M R I I A F V G S P V E D	119
Drosophila	P K G E I N L L T G I R I A H L V L K H R Q G K N H K M R I V V F V G S P I N H	119
Arabidopsis	V G G E I N L T A A I Q I A Q L A L K H R Q N K N Q R Q R I I V F A G S P I K Y	120
S. cerevisiae	I E G K L H M A T A L Q I A Q L T L K H R Q N K V Q H Q R I V A F V C S P I S D	120
Human	N E K D L V K L A K R L K K E K V N V D I I N F G E E E V N T - - E K L T A F V	157
Drosophila	E E G D L V K Q A K R L K K E K V N V D I V S F G D H G N N - - E I L T A F I	157
Arabidopsis	E K K A L E I V G K R L K K N S V S L D I V N F G E D D D E E K P Q K L E A L L	160
S. cerevisiae	S R D E L I R L A K T L K K N N V A V D I I N F G E I E Q N T - - E L L D E F I	158
Human	N T L N G K D G T G S H L V T V P P G P S L - A D A L I S S P I L - - A G E G G	194
Drosophila	N A L N G K D G T G S H L V S V P R G S V L - S D A L L S S P I I - - Q G E D G	194
Arabidopsis	T A V N N - - N D G S H I V H V P S G A N A L S D V L L S T P V F - - T G D E G	196
S. cerevisiae	A A V N N P Q E E T S H L L T V T P G P R L L Y E N I A S S P I I L E E G S S G	198
Human	- A M - - - - - L G L G A S D F E F G V D P S A D P E L A L A L R V S M E E	226
Drosophila	M G G - - - - - A G L G G N V F E F G V D P N E D P E L A L A L R V S M E E	227
Arabidopsis	A S G Y V S A A A A A A G G D F D F G V D P N I D P E L A L A L R V S M E E	236
S. cerevisiae	M G A F G G S G G D S D A N G T F M D F G V D P S M D P E L A M A L R L S M E E	238
	* * * * O O O O ● ●	
Human	Q R Q R Q E E E R R R A A A A S A - - - - - - - - - - - A E A G	247
Drosophila	Q R Q R Q E S E Q R R A N P D G A P P T G G D A G G G G V S G S G P G N E E S	267
Arabidopsis	E R A R Q E A A A K K A A D E A G - - - - - - - - - - - Q K D K	257
S. cerevisiae	E Q Q R Q E - - R L R Q Q Q Q Q Q D Q P E Q S E Q P E Q H Q D K	268
	● ● ● ● ● ● ● ● ● ●	
Human	I A T T G T E D S D D A L L K M T I S - Q Q E F G R T G L P D L S S M T E E E E Q	286
Drosophila	A G A E N E A N T E E A M L Q R A L A L S T E T P E D N L P D F A N M T E E E E Q	307
Arabidopsis	D G D T A S A S Q E T V - - A R T T D - K N A E P M D E D S A L L D Q A I A M S	294
Human	I A Y A M Q M S L Q G A E F G Q A E S A D I D A S S A M D T S E P A K E E D D Y	327
Drosophila	I A F A M Q M S M Q D A P - D D S V T Q Q A K R P K T D E A N A P M D V D E D Y	346
Arabidopsis	V G - D V N M S E A A D E - D Q D L A L A L Q M S M S G E E S S E A T G A G N -	331
Human	- D V M Q D P E F L Q S V L E N L P G V D P N N E A I R N A M G S L A S Q A T K	366
Drosophila	S E V I G D P A F L Q S V L E N L P G V D P Q S E A V R D A V G S L N K D K D K	386
Arabidopsis	- N L L G N Q A F I S S V L S S L P G V D P N D P A V K E L L A S L P D E S K R	370
	* * * * O O O O ● ●	
Human	- - - - D G K K D K K E E D K K	378
Drosophila	- - - - - K S D G K D S Q K K	396
Arabidopsis	T E E E S S S K K G E D E K K	386
	● ● ● ● ● ● ● ● ● ●	

Fig. 2. Comparison of S5a sequences from Human (S5a), *Drosophila* (54 kDa), *Arabidopsis* (MBP1) and *Saccharomyces* proteins. Positions at which a majority of the residues are identical have been denoted by white letters on black. The *Drosophila* sequence is from ref. [27]; the *S. cerevisiae* sequence is in Genbank (Accession U00030) and the *Arabidopsis* sequence will be published in ref. [23].

similarity to MBP1 were identified, all were incomplete, consisting at most of several hundred nucleotides. Therefore, a synthetic oligonucleotide corresponding to one of the partial human sequences was used to screen a HeLa cell cDNA library (see section 2). This resulted in the isolation of a cDNA that encodes a 378 residue protein (Fig. 1). The open reading frame begins with a reasonable consensus for translational initiation [25] and terminates in a UGA stop codon. The

deduced protein contains three regions, shown white on black in Fig. 1, identical to peptides directly sequenced from human 26S protease S5a [19]. A calculated M_r of 40.3 kDa and pI of 4.6 for the putative translation product are almost identical to *Arabidopsis* MBP1 and consistent with the observation that human S5a and *Arabidopsis* MBP1 migrate very similarly on two-dimensional gels [23]. Similarly, the C-terminus of both MBP1 and the protein predicted by the HeLa cell cDNA are

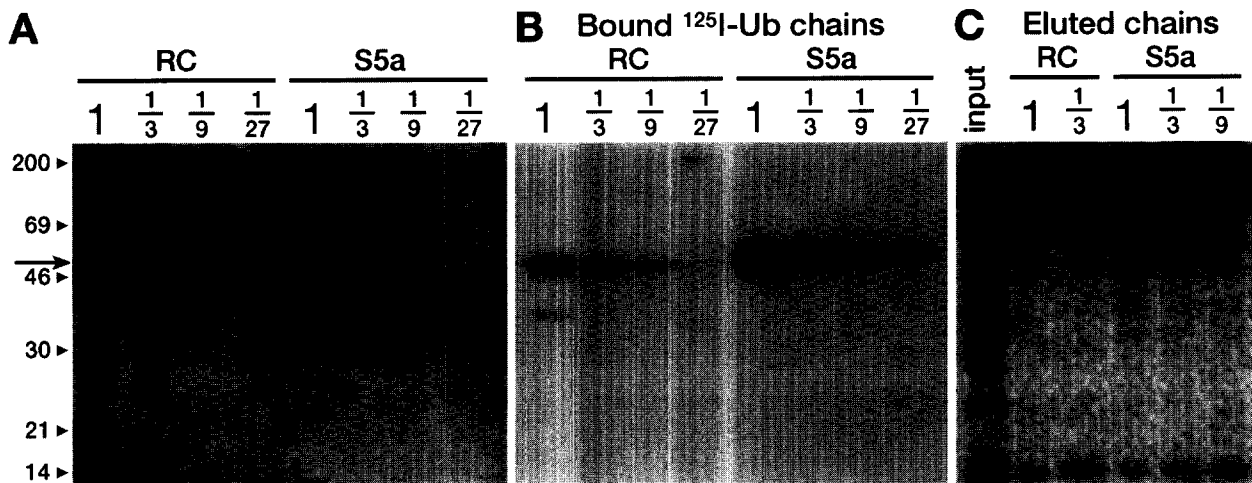


Fig. 3. Multiubiquitin binding properties of recombinant S5a. The human S5a cDNA was inserted into pET16b behind a decahistidine extension. *E. coli* cells transformed with this construct were induced with IPTG, harvested and sonicated as previously described [19]. The histidine-S5a recombinant protein was purified by metal chelate chromatography. In Panel A, 1 μ g of the largely purified decahistidine tagged S5a protein (S5a, right four lanes) or for comparison 20 μ g of human erythrocyte regulatory complex (RC, left four lanes) were subjected to 3-fold dilutions, electrophoresed on a 10% SDS polyacrylamide gel, transferred to nitrocellulose and the filter was stained with Ponceau S. Panel B is an autoradiograph of the nitrocellulose filter after incubation in the presence of [125 I]multiubiquitin chains. Panel C is a phosphorimage of the bound multiubiquitin chains eluted from the nitrocellulose filter shown in panel B and separated on a 10% SDS polyacrylamide gel. Approximately 1000 cpm were electrophoresed from lanes 1 and 1/3 of RC and lanes 1, 1/3 or 1/9 of recombinant S5a; insufficient radioactivity was recovered from lanes 1/9, 1/27 of RC and lane 1/27 of S5a for analysis of the size distribution of multiubiquitin chains bound at these dilutions. As estimated by Ponceau S staining intensity (panel A) and binding activity (panel B), lane 1 of RC and lane 1/9 of recombinant S5a contain similar amounts of the S5a protein. As observed in panel C lanes 1 of RC and 1/9 of S5a, both regulatory complex and recombinant S5as exhibit similar selection for long multiubiquitin chains. Note that histidine tagged S5a migrates slightly slower than the S5a from the regulatory complex. However, recombinant S5a lacking the histidine extension migrates identically to S5a from the regulatory complex, and it also exhibits similar multiubiquitin chain binding properties (data not shown).

enriched in 'alternating' lysines (K) and glutamates (E). These KEKE regions have been proposed to function in protein-protein association [26].

The HeLa cell cDNA sequence is highly similar to *Arabidopsis* MBP1 as well as to a 50-kilodalton 26S protease subunit from *Drosophila* [27] and an ORF from *Saccharomyces* (accession no. YSCH9998). Several interesting features emerge from a comparison of the four sequences. All show a high degree of similarity over the first 190 amino acids (50–69% identity). This relatively well conserved N-terminal portion is followed by a glycine-rich region in each protein (residues 190–205 for *Drosophila* and human and residues 190–215 for the yeast and plant sequences). Just C-terminal to the glycine rich region is the most conserved segment in the four sequences. The 27 amino acid sequence that reads... EFGVDP-SADPELALALRVSMEEQRQRQE... in the human amino acid sequence varies among the four proteins at only 18 out of 108 positions. In the *Arabidopsis* protein a portion of this region is repeated almost exactly (DQDLALAL...) further toward the C-terminus. And in the three higher eukaryotic proteins, the tetrapeptide GVDP is repeated near the C-terminus. In fact, one can speculate that GVDP is part of a larger repeat characterized by GVDP-X_{5–7}-LALAL or AMGSL followed by a mixed charge run (e.g. EEQRQRQEERRR at residues 225–237 or KDGGKDKKEEDKK at residues 366–378 in the human sequence). The proposed longer repeats are denoted in Fig. 2 by two series of asterisks, open circles and closed circles placed beneath the aligned sequences. Although the exact boundaries of the proposed repeated motifs are not readily apparent, it is clear that the C-terminal halves of the human, fly and plant proteins contain repeated sequences. The yeast sequence, however, is considerably shorter than the

other three proteins, and it lacks the proposed sequence repeats. One might assume that the predicted yeast protein is prematurely truncated due to sequencing error, but conceptual translation of the remaining two reading frames does not reveal further homology to S5a. Moreover, antiserum specific for MBP1 reacts with a yeast protein that migrates on SDS-PAGE gels with an apparent molecular weight considerably less than human S5a or MBP1 and consistent with that predicted by the genomic yeast sequence (data not shown). Thus, it appears that yeast S5a is smaller than its counterparts in higher eukaryotes.

To determine if the identified HeLa cell cDNA encodes a protein that binds multiubiquitin chains, several concentrations of human 26S protease regulatory complex and recombinant his-tagged S5a were separated by SDS-PAGE and transferred to nitrocellulose filters. The filters were then incubated in the presence of [125 I]multiubiquitin chains, rinsed and subjected to autoradiography. As shown in panels A and B of Fig. 3, both S5a from the regulatory complex and recombinant S5a bound [125 I]multiubiquitin chains. The two proteins were shown to exhibit similar selection for longer multiubiquitin chains by excising pieces of the nitrocellulose blot that contained [125 I]multiubiquitin-S5a complexes, boiling them in SDS-PAGE buffer and analyzing the size distribution of [125 I]-multiubiquitin chains on a second SDS-polyacrylamide gel. It can be seen from the autoradiogram in Fig. 3C that S5a molecules from the regulatory complex or from *E. coli* select principally for long multiubiquitin chains (compare eluted chains to input in Fig. 3C).

Different concentration of regulatory complex or recombinant protein were analyzed in Fig. 3 because we have observed variation in the lengths of bound multiubiquitin chains

depending upon the amounts of S5a on the filter. Although in Fig. 3C the concentration-dependent differences in chain length are slight, they can be discerned. For example, intermediate sized chains (MW between 45 and 70 K) are more abundant when large amounts of S5a are present on the nitrocellulose filter (compare lanes 1 vs. 1/9 in subpanel S5a of Fig. 3C). In view of these concentration effects, we cannot be certain that recombinant S5a binds multiubiquitin chains in exactly the same way as S5a obtained from the regulatory complex. They are, at the least, very similar.

4. Discussion

Isolation and sequencing of a cDNA with high similarity to plant MBP1 has allowed us to deduce the amino acid sequence of human S5a (Fig. 1). There is good evidence that we have obtained the entire coding region for the human subunit. First, if one takes into account the extra N-terminal histidines, one can conclude that the expressed protein migrates on SDS-PAGE very much like S5a obtained directly from the 26S protease (Fig. 3A). Second, a set of 14 human ESTs overlap the S5a open reading frame starting from base pair (bp) 142 and continuing into a poly A tail beyond base pair 1330. Third, the deduced amino acid sequence is similar in length to sequences for S5a homologs present in *Arabidopsis* and *Drosophila* (Fig. 2). Fourth, the nucleotide sequence presented in figure 1 is virtually identical to a human pituitary cDNA proposed to encode a protein termed antisequestory factor (AF) [28]. Although the deduced amino acid sequences for human S5a and AF diverge at the C-terminus, this may be explained by sequencing errors or lack of fidelity in DNA synthesis during cloning. In fact, the final 24 residues reported for AF are present in an alternate reading frame of the cDNA encoding human S5a.

AF was identified as a pituitary protein able to inhibit intestinal fluid loss in rats injected with cholera toxin [29]. Johansson et al. [28] obtained a cDNA for AF by screening a human pituitary expression library with antibodies to porcine AF. The apparent identity of AF and the 26S protease subunit S5a is very surprising. Free S5a subunits can inhibit ubiquitin mediated proteolysis [30], so AF might function by reducing the turnover of membrane bound receptors or possibly speeding the turnover of G proteins within intestinal cells. It remains difficult to explain how S5a secreted from the pituitary could enter intestinal cells to function in this capacity. The finding that AF is apparently identical to a ubiquitin-conjugate binding subunit of the 26S protease clearly requires additional documentation.

We have shown that human S5a expressed in *E. coli* binds and selects for longer multiubiquitin chains (Fig. 3). Selection of multiubiquitin chains by the his-tagged recombinant S5a is very similar to that seen with S5a obtained directly from the human erythrocyte 26S protease. Still, we cannot rule out subtle differences between recombinant S5a molecules and those present in the regulatory complex. In this regard it is noteworthy that Haracska and Udvardy [27] suggest that a 50 kDa subunit from the *Drosophila* 26S protease is glycosylated; the *Drosophila* subunit is now known to be S5a from the data presented here and in ref. [23].

How the 26S protease recognizes multiubiquitin chains remains an interesting and only partially solved problem. Our previous discovery of a specific 26S protease subunit that

binds multiubiquitin chains [22] and more recent identification of sequences for homologs of this subunit ([23,27] and this report) provide information that should prove helpful in elucidating the molecular mechanism for multiubiquitin chain recognition. The highly conserved segment FGVPDS...RQE is likely to be an important element for binding ubiquitin polymers. In fact, current deletion analyses implicate this portion of S5a in multiubiquitin binding (Q. Deveraux et al., in preparation). Under results, we emphasized the probable existence of repeated sequences in the C-terminus of S5a. We did so because we believe such repeats can explain the ability of S5a to selectively bind longer polymers of ubiquitin. We speculate that repeated sequence elements in S5a bind repetitive hydrophobic patches on multiubiquitin chains [31]. Ongoing deletion analyses should provide an experimental test of the repeated sequence or multiple 'loop' model for multiubiquitin chain recognition by the 26S protease [19].

Acknowledgements: These studies were supported by National Institutes of Health Grant GM37009 to Martin Rechsteiner and a grant from the Lucille P. Markey Charitable Trust. We would like to thank Cecile Pickart for providing the multiubiquitin chains, Wolfgang Dubiel and Patrick Young for helpful comments and Chris Jensen for proofreading the sequencing gels. We also thank Linda Van Orden for expert word processing.

References

- [1] Glotzer, M., Murray, A.W. and Kirschner M.W. (1991) *Nature* 349, 132–138.
- [2] Hershko, A., Ganoth, D., Pehrson, J., Palazzo, R.E. and Cohen, L.H. (1991) *J. Biol. Chem.* 266, 16376–16379.
- [3] Chen, Z., Hagler, J., Palombella, V.J., Melandri, F., Scherer, D., Ballard, D. and Maniatis, T. (1995) *Genes Dev.* 9, 1586–1597.
- [4] Shanklin, J., Jabben, M. and Vierstra, R.D. (1987) *Proc. Natl. Acad. Sci. USA* 84, 359–363.
- [5] Hershko, A. and Ciechanover, A. (1992) *Annu. Rev. Biochem.* 61, 761–808.
- [6] Ciechanover, A. (1994) *Cell* 79, 13–21.
- [7] Rechsteiner, M., Hoffman, L. and Dubiel, W. (1993) *J. Biol. Chem.* 268, 6065–6068.
- [8] Orlowski, M. (1993) *J. Lab. Clin. Med.* 121, 187–189.
- [9] Rivett, A.J. (1993) *Biochem. J.* 291, 1–10.
- [10] Hoffman, L. and Rechsteiner, M. (1992) *J. Biol. Chem.* 267, 22362–22368.
- [11] Udvardy, A. (1993) *J. Biol. Chem.* 268, 9055–9062.
- [12] Peters, J.-M., Franke, W.W. and Kleinschmidt, J.A. (1994) *J. Biol. Chem.* 269, 7709–7718.
- [13] DeMartino, G., Moomaw, C.R., Zagnitko, O.P., Proske, R.J., Chu-Ping, M., Afendis, S.J., Swaffield, J.C. and Slaughter, C.A. (1994) *J. Biol. Chem.* 269, 20878–20884.
- [14] Ganoth, D., Leshinsky, E., Eytan, E. and Hershko, A. (1988) *J. Biol. Chem.* 263, 12412–12419.
- [15] Dubiel, W., Ferrell, K. and Rechsteiner, M. (1995) *Mol. Biol. Reports* 21, 27–34.
- [16] Tanaka, K. (1995) *Mol. Biol. Reports* 21, 21–26.
- [17] Boldin, M.P., Mett, I.L. and Wallach, D. (1995) *FEBS Lett.* 367, 39–44.
- [18] Lurquin, C., Van Pel, A., Mariamé, B., De Plaen, E., Szikora, J.-P., Janssens, C., Reddehase, M.J., Lejeune, J. and Boon, T. (1989) *Cell* 58, 293–303.
- [19] Deveraux, Q., Jensen, C. and Rechsteiner, M. (1995) *J. Biol. Chem.* 270, 23726–23729.
- [20] Dubiel, W., Ferrell, K., Dumdey, R., Standera, S., Prehn, S. and Rechsteiner, M. (1995) *FEBS Lett.* 363, 97–100.
- [21] Kominami, K., DeMartino, G.N., Moomaw, C.R., Slaughter, C.A., Shimbara, N., Fujimuro, M., Yokosawa, H., Hisamatsu, H., Tanahashi, N., Shimizu, Y., Tanaka, K. and Toh-e, A. (1995) *EMBO J.* 14, 3105–3115.
- [22] Deveraux, Q., Ustrell, V., Pickart, C. and Rechsteiner, M. (1994) *J. Biol. Chem.* 269, 7059–7061.

- [23] van Nocker, S., Deveraux, Q., Rechsteiner, M. and Vierstra, R. (1996) PNAS in press.
- [24] Ausubel, F. M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1992) Current Protocols in Molecular Biology, Green Publishing Associates, Wiley-Interscience, New York.
- [25] Kozak, M. (1986) Cell 44, 283–292.
- [26] Realini, C., Rogers, S.W. and Rechsteiner, M. (1994) FEBS Lett. 348, 109–113.
- [27] Haracska, L. and Udvardy, A. (1995) Eur. J. Biochem. 231, 720–725.
- [28] Johansson, E., Lönnroth, I., Lange, S., Jonson, I., Jennische, E. and Lönnroth, C. (1995) J. Biol. Chem. 270, 20615–20620.
- [29] Lönnroth, I. and Lange, S. (1986) Biochim. Biophys. Acta 883, 139–144.
- [30] Deveraux, Q., van Nocker, S., Mahaffey, D., Vierstra, R. and Rechsteiner, M. (1995) J. Biol. Chem. 270, 29660–29663.
- [31] Beal, R., Deveraux, Q., Gang, X., Rechsteiner, M. and Pickart, C. (1996) Proc. Natl. Acad. Sci. USA, in press.