Published in final edited form as:

Proteomics. 2012 June ; 12(11): 1806–1814. doi:10.1002/pmic.201100260.

Chemical cross-linking, mass spectrometry and *in silico* modeling of proteasomal 20S core particles of the haloarchaeon *Haloferax volcanii*

Ivanka Karadzic¹, Julie Maupin-Furlow², Matt Humbard², Laurence Prunetti², Pragya Singh³, and David R. Goodlett⁴

Ivanka Karadzic: ikaradzic@med.bg.ac.rs

¹Department of Chemistry, School of Medicine, University of Belgrade, Visegradska 26, Belgrade, Serbia

²Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida, USA

³Lawrence Berkeley National Laboratory, Berkeley, California, USA

⁴Department of Medicinal Chemistry, School of Pharmacy, University of Washington, Seattle, Washington, USA

Abstract

A fast and accurate method is reported to generate distance constraints between juxtaposed amino acids and to validate molecular models of halophilic protein complexes. Proteasomal 20S core particles (CPs) from the haloarchaeon *Haloferax volcanii* were used to investigate the quaternary structure of halophilic proteins based on their symmetrical, yet distinct subunit composition. Proteasomal CPs are cylindrical barrel-like structures of four-stacked homoheptameric rings of α - and β -type subunits organized in $\alpha_7\beta_7\beta_7\alpha_7$ stoichiometry. The CPs of *H. volcanii* are formed from a single type of β subunit associated with α_1 and/or α_2 subunits. Tandem affinity chromatography and new genetic constructs were used to separately isolate $\alpha_1\beta_7\beta_7\alpha_1$ and $\alpha_2\beta_7\beta_7\alpha_2$ CPs from *H. volcanii*. Chemically cross-linked peptides of the *H. volcanii* CPs were analyzed by high-performance mass spectrometry and an open modification search strategy to first generate and then to interpret the resulting tandem mass spectrometric data. Distance constraints obtained by chemical cross-linking mass spectrometry (CXMS), together with the available structural data of non-halophilic CPs, facilitated the selection of accurate models of *H. volcanii* proteasomal CPs composed of α_1 -, α_2 -, and β -homoheptameric rings from among several different possible PDB structures.

Keywords

cross-linking; haloarchaea; modeling; proteasomes; tandem mass spectrometry

1 Introduction

Protein structural data combined with genetics and genomics is important in elucidating gene function. Recent advances in the genome sequencing and the development of genetic systems in haloarchaea have made this group of extreme halophiles ideal for providing

*Corresponding authors: Julie Maupin-Furlow, Professor, Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611-0700, USA, Fax: 352 392-5922, Phone: 352 -392-4095, jmaupin@ufl.edu. David R. Goodlett, Professor, Department of Medicinal Chemistry, School of Pharmacy, University of Washington, Seattle, Washington 98195-7610, USA, Fax: 206-685-3252, Phone: 206-616-4586, goodlett@uw.edu.

insights into many basic cellular functions (*e.g.*, basal transcription factors, protein glycosylation, twin arginine translocation (TAT) of proteins, DNA repair and ubiquitin-proteasome systems) [1]. While the genetics of haloarchaea is advanced, halophilic proteins are notoriously difficult to characterize at the biochemical level thus limiting the full potential of haloarchaea for functional genomic studies.

The haloarchaeon *Haloferax volcanii* is a model system for investigating the function of the ubiquitin-proteasome system in archaeal cells. Proteasomes are self-compartmentalized nanomachines, composed of AAA+ ATPases and 20S core particles (CPs) that are important in proteolysis and often essential for growth [2]. The proteasomal CP is a cylindrical complex of four-stacked heptameric rings of α - and β -type subunits organized in $\alpha_7\beta_7\beta_7\alpha_7$ stoichiometry. The α -type subunits form the outer rings and the β -type subunits form the two inner rings which harbor the proteolytic active sites [3,4]. *H. volcanii* synthesizes two different α subunits, α_1 and α_2 , having the potential to make three different CPs: two symmetric ($\alpha_1\beta\beta\alpha_1$, $\alpha_2\beta\beta\alpha_2$) and one asymmetric ($\alpha_1\beta\beta\alpha_2$) [4]. Of these, the $\alpha_1\beta\beta\alpha_1$ and $\alpha_1\beta\beta\alpha_2$ CPs have been characterized [4], while the $\alpha_2\beta\beta\alpha_2$ CP has yet to be isolated. Since the α_1 and α_2 proteins share only 55.5% identity, significant structural differences in the homoheptameric rings formed by α_1 and α_2 is predicted [4].

In the post-genomic era, sensitive and high-throughput techniques such as mass spectrometry can play a significant role in the large-scale analysis of protein structure. A chemical-cross-linking-mass spectrometry (CXMS) analytical approach based on high-performance mass spectrometry to generate tandem mass spectra and the open modification search strategy to interpret the data has been reported [5] and validated using complexes such as a bacterial T3SS needle [6].

A large and increasing set of chemical cross-linkers used as molecular rulers can provide information on distances between cross-linked amino acid residues that are relevant to both the tertiary and quaternary structure of proteins. Homobifunctional N-hydroxysuccinimide (NHS) esters primarily target the ϵ -amino group of lysine residues, but can differ in the length of spacer arm [7, 8, 9]. In contrast to most bifunctional reagents, which introduce a bridge between cross-linked residues, zero-length carbodiimide cross-linkers mediate the formation of a covalent bond between carboxylate and amine groups without an intervening linker, allowing a direct evaluation of contact interactions between protein surfaces [7].

With regard to CXMS investigation of *H. volcanii* CPs, several points must be emphasized. First, haloarchaeal proteins are typically halophilic (salt-loving) and quite specific with regard to amino acid composition. Haloarchaeal proteins have an extremely high content of acidic residues that are often surface exposed and a low number of basic amino acid residues (*e.g.*, lysine, a target of many crosslinkers) [10]. In addition, the *H. volcanii* CPs do not contain any cysteine residues. Both of these factors limit and complicate the investigation of the halophilic CPs by CXMS since lysine and cysteine residues are often targets of commercial cross-linkers. Furthermore, there are only a few crystal structures resolved to date of haloarchaeal proteins and proteasomal CPs (with the latter derived only from non-halophilic archaea or eukarya). In spite of these limitations with regard halophilic protein structure, CXMS coupled with protein modeling, could emerge as a powerful approach to predict the structure of these unusual, acidic proteins.

Comparative modeling of protein complexes results in a number of possible structures that may be optimized and/or evaluated for validity through distance constraints generated by CXMS [7, 11,12]. In particular, comparative protein modeling, which uses previously solved structures as templates, can be a very effective starting point. It has been shown that even a small number of intermolecular cross-link constraints was sufficient to validate the

topology prediction of a protein complex [7, 13, 14], implying the ultimate goal to be acquisition of cross-linking distance constraints of protein assemblies on a routine basis [7]. The gap between CXMS and protein structure modeling has been recently addressed by a software platform called MSX-3D which facilitates validation of theoretical models based on CXMS data [15].

Structural characterization of proteasomes from eukarya and yeast by CXMS have been recently published [16, 17]. Although electron microscopy of the proteasomal CPs from *H. volcanii* reveals a four-stacked ring structure with a central channel [3], the structural details for these protein complexes are limited. Here we report the first investigation that focuses on the structure of the *H. volcanii* CP complexes by using chemical cross-linkers to determine distance constraints between juxtaposed amino acids by high-accuracy MS. The CXMS data were used in two ways. First, intramolecular cross-links were used to validate the predictions of the 3D-structures of protein subunits. Next, intermolecular distance constraints were used to assemble the proteins into the quaternary CP structure.

2 Material and Methods

2.1 Strains, media and plasmids

Strains and plasmids are summarized in Table 1. *Escherichia coli* DH5 α was used for DNA cloning. *E. coli* GM2163 was used as the *dam*⁻ strain for isolation of plasmid DNA for transformation of *H. volcanii*. *E. coli* strains were grown in Luria-Bertani medium (37°C, 200 rpm). *H. volcanii* strains were grown in complex medium (ATCC 974) (42°C, 200 rpm). Media were supplemented with 100 mg of ampicillin or 0.1 mg of novobiocin per liter as needed. Plasmid pJAM2564 (used for co-production of α 2-His6 and β -StrepII in *H. volcanii*) was constructed by blunt-end ligation of a PCR product encoding α 2 with a C-terminal His6-tag downstream of a BlnI site in pJAM816. PCR product was generated using Phusion DNA polymerase with *H. volcanii* genomic DNA and the following primer pairs (α 2_{wt}_{xba}_{nde}: 5'-GCTCTAGACATATGaaccgaaacgacaagc-3' and α 2_{his}_{blnI}_{rev} 5'-CCGATGCTGAGC-TTAGTGGTGGTGGTGGTgctctcgctgctgcgctcgcgga-3'). Clones were screened for directionality of insert, and the fidelity was confirmed by DNA sequencing using the dideoxy termination method with capillary array sequencers (Genomics Core, Interdisciplinary Center for Biotechnology Research, University of Florida).

2.2 Proteasome Purification

Synthetic operons encoding affinity-tagged CPs were expressed in *H. volcanii* parent H26 and proteasomal mutant strains. *H. volcanii* strains GZ114-pJAM2545 and GZ130-pJAM2564 were used for synthesis and purification of α 1 β β α 1 and α 2 β β α 2 proteasomes, respectively. Proteasomes were purified by tandem Ni²⁺-Sephacel (HiTrap chelating; Amersham Biosciences) and Streptactin (Qiagen) chromatography as previously described [20].

2.3 Cross-Linking Reactions

Before initiating the cross-linking reactions, proteasome preparations were dialyzed against 20 mM phosphate buffer pH 7.2 (buffer A) supplemented with 2 M NaCl. All dialysis steps were performed overnight at 4°C using a Slide-A-Lyzer™ Dialysis Cassette (MWCO 3500, Pierce, ThermoScientific, USA). Proteasomes were cross-linked with: (i) bis[sulfosuccinimidyl] glutarate-d₀ (BS2G) and (ii) 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride (EDC). All cross-linking reactions were carried out according to protocols from the manufacturer (Pierce, ThermoScientific, USA), with slight modifications as described below.

For BS2G cross-linking, proteasomes (50 µg in 100 µL) were incubated for 60 min at room temperature with 1 µL of 50 mM BS2G dissolved in DMSO. The reaction was terminated by addition of ammonium bicarbonate (NH₄HCO₃) to a final concentration of 20 mM. Excess salt was removed by dialysis against buffer A.

For EDC cross-linking, proteasomes (50 µg in 100 µL) were incubated at room temperature for 2 h with 8 mM EDC (added from a 100 mM EDC stock, freshly prepared in ultrapure water). The reaction was terminated by dialysis against buffer A.

2.4 In Solution Trypsin Digestion

To denature the protein, urea (36 mg) was added to 100 µL of the cross-linked protein solution. After the urea was completely dissolved, 7.5 µL of 1.5 M Tris buffer at pH 8.8 was added, and the sample was incubated for 1 h at room temperature. The sample was reduced by addition of 2.5 µL of 200 mM tris(2-carboxyethyl)phosphine (TCEP) and incubation for 1 hr at 37°C. Protein was alkylated by addition of 20 µL of 200 mM iodoacetamide and incubation in the dark for 1 h at room temperature. Finally, 20 µL of 200 mM of dithiothreitol was added, and the sample was incubated at room temperature for 1 h. To dilute the urea, 800 µL of 25 mM NH₄HCO₃ and 100 µL methanol (HPLC grade) were added to each tube. Protein was digested with trypsin (Promega Sequencing Grade modified Trypsin dissolved in 25 mM ammonium-bicarbonate) at a ratio of 1 trypsin: 50 protein (2 µg trypsin for 100 µg protein). Digestion was performed overnight at room temperature and concentrated to near dryness in a Speedvac. Protein was desalted by UltraMicro Spin mini C18 columns (SS18V, The Nest Group, MA, USA) according to manufacturer protocol.

2.5 Mass Spectrometry and HPLC

Peptide digests were analyzed by electrospray ionization in the positive mode on a hybrid linear ion trap-Orbitrap instrument (LTQ-Orbitrap, Thermo Fisher, San Jose, CA, USA). Peptides were separated by nanoflow HPLC (NanoAcquity, Waters Co., Milford, MA, USA). Homemade precolumns (100 µm i.d. x 25 mm long) packed with 200 Å C₁₈ stationary phase (5 µm, C18AQ, Michrom) were used for peptide trapping. Analytical columns (75 µm x 210 mm long) packed with 100 Å C₁₈ stationary phase (5 µm, C18AQ, Michrom) were coupled to the mass spectrometer.

HPLC-MS methods were designed according to Singh *et al.* [5]. Peptide mixtures obtained after tryptic digestion (0.5 µg) were applied to the precolumn at 4 µL·min⁻¹ in 5% (v/v) acetonitrile with 0.1% (v/v) formic acid. Peptides were eluted by a linear gradient of A (water, 0.1 % formic acid) and B (acetonitrile, 0.1% formic acid), as follows: 0 min- A (95%), B (5%); 55 min- A (65%), B (35%); 60 min- A (15%), B (85%); 65 min- A (5%), B (95%); 75-90 min- A (95%), B (5%). Ion source conditions were optimized with a tuning and calibration solution according to instrument provider. All MS survey scans were performed from m/z 400-2000, at a resolution of 60,000 (m/z) and ion population of 5 x 10⁵. For tandem MS, resolution was set to 7500, ion population to 2 x 10⁵ and precursor isolation width to 4 m/z units. Data dependent analysis was performed by selection of the five most abundant precursors, rejecting singly, doubly, triple charged ions. Data redundancy was minimized by dynamic exclusion of previously selected precursor ions (-0.1/1.1 Da) for 45 s before being selected again for fragmentation.

2.6 MS Data Processing

Tandem mass spectral data were converted to .dta files and deconvoluted to the 2+ charge state precursor and 1+ charge state fragments by an in-house written Perl script (<http://goodlett.proteomics.washington.edu>). A database was formed using xComb v1.1[19], setting parameters as follows: sequence was submitted in UniProt FASTA format; cleavages: by

trypsin with up to 2 missed cleavages; intra- or inter-protein crosslinks. Deconvoluted spectra were searched by Phenyx (GeneBio SA, Geneva, Switzerland) to identify cross-linked peptides using data bank (DB) formed by xComb. File in .mgf format was submitted to Phenyx and search parameters were set as follows: DBs created by xComb were added; taxonomy-root; scoring model-ESI-LTQ-Orbitrap (CID_LTQ_scan_Orbitrap_6 ppm); parent charge- 1,2,3,4; modifications including methionine in oxidized and reduced forms and cysteine alkylated by iodacetamide; enzyme- do not cleave, missed cleavages-0; parent tolerance- 10 ppm; peptide thresholds: length 6, score 4.0, p-value 1.0 E-5; AC score of 4; turbo scoring: tolerance of 10.0 ppm, coverage 0.1, series of b;b++;y;y++ were used. The MS/MS fragmentation of cross-linked peptides obtained from the Phenyx search was analyzed to assign ion peaks, using MS2Assign with threshold of 50 ppm [22]. According to the classification of cross-linking products interpeptide, intrapeptide and dead-end modifications were considered [22].

2.7 Protein structure validation

Theoretically calculated models of $\alpha 1$, $\alpha 2$ and β protein subunits were obtained from the publicly available MODBASE (<http://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi>). Structures designated as $\alpha 1$ (Model ID: 4de8eea6b134a909401c34d144ce816c), β (Model ID: b0e4 c2ed79c3a4a1971f2bb8e71bd170) and $\alpha 2$ (Model ID: 2aecb9b89471cbe8bca4af 05785aa78b) were used. To analyze 3D protein models using mass spectrometry data, MSX-3D, version 3.4.23 (<http://proteomics-pbil.ibcp.fr/cgi-bin/msXsetup.pl>) was used [15]. Pairwise comparison of protein structures was established using DaliLite workbench software (<http://www.ebi.ac.uk/Tools/dalilite/index.html>) [23]. The coordinates of heptameric complexes of $\alpha 1$, $\alpha 2$ and β were obtained by DaliLite software using PDB entries 1j2p, 1ryp, and 1fnt (for $\alpha 1$ and $\alpha 2$) and 1ryp, 1j2q, and 1fnt (for β). Coordinates of the PDB structures with high Z scores and a strong match to $\alpha 1$, β , and $\alpha 2$ subunits were used to assemble proteins into heptameric ring structures of $\alpha 1_7$, β_7 , and $\alpha 2_7$ and double heptameric rings of $\beta_7\beta_7$ by Vega ZZ molecular modeling software. Molecular modeling package Vega ZZ was used to add hydrogen onto the ring structures and to calculate charge and potential, using Charm for force field and Gasteiger for charge [24].

3 Results and Discussion

3.1 Isolation and structural characterization of proteasomal CP complexes

Two distinct and symmetric proteasomal CP subtypes were purified to homogeneity from *H. volcanii* for CXMS analysis. One CP subtype was composed only of $\alpha 1$ and β ($\alpha 1\beta\alpha 1$) and the other CP was only of $\alpha 2$ and β subunits ($\alpha 2\beta\alpha 2$). A two-step affinity method was used to purify the $\alpha 1\beta\alpha 1$ CP from *H. volcanii* GZ114-pJAM2545 as previously described [20] (see Table 1 for genetic details). This two-step method was also applied to a new genetic construct of *H. volcanii* (GZ130-pJAM2564, this study) that enabled purification of $\alpha 2\beta\alpha 2$ CPs for the first time. The CP subtypes ($\alpha 1\beta\alpha 1$ and $\alpha 2\beta\alpha 2$) separately purified from *H. volcanii* are presumed to be assembled as a cylinder of four-stacked heptameric rings in $\alpha_7\beta_7\beta_7\alpha_7$ stoichiometry based on analogy to CPs with detailed X-ray crystal structures [25, 26] and the previous observations as follows. The *H. volcanii* CPs are of a 600 kDa molecular mass (based on gel filtration) corresponding to a 28 subunit complex of 14 α -type and 14 β -type subunits (based on the masses calculated from polypeptide sequence). The α - and β -type subunits of *H. volcanii* CPs are in a 1:1 stoichiometric ratio as determined by Coomassie staining of the CP proteins separated by reducing SDS-PAGE. The CPs are cylindrical barrel-like structures of four-stacked protein rings with the α -type proteins associating as heptameric rings as observed by electron microscopy of purified complexes [3, 4, 20].

3.2 Identification of juxtaposed amino acids by CXMS and protein structure validation

To analyze the two symmetric CP subtypes of *H. volcanii* by CXMS the following strategy was used. First, CP complexes ($\alpha 1\beta\alpha 1$ and $\alpha 2\beta\alpha 2$) were treated with the homobifunctional lysine-reactive cross-linker BS2G and the zero length EDC cross-linkers. Efficiency of cross-linking reactions was monitored by SDS PAGE. Second, MS/MS spectra of the cross-linked CPs were analyzed using the open modification search strategy to confirm the identity and to evaluate the type of cross-linkages [5]. Briefly, according to this method, all data obtained at high mass accuracy on LTQ-Orbitrap mass spectrometer, were searched with respect to the type of cross-linker with Phenix, using a DB formed by xComb [19], and finally the MS/MS spectra of cross-linked peptides were annotated using MS2Assign [22]. Third, distance constraints obtained by CXMS were used for validation of theoretical models of $\alpha 1$, β , and $\alpha 2$ subunits by MSX-3D [10]. In addition to MSX-3D, VEGA ZZ program was used to display, analyze, and manage the three dimensional (3D) structure of the protein complexes [24]. The cross-linked peptides of the $\alpha 1\beta\alpha 1$ and $\alpha 2\beta\alpha 2$ CP subtypes that were identified after treatment with BS2G and EDC, as well as the predicted intra- and inter-molecular distances between juxtaposed amino acids within each monomer and between the same residues in the oligomeric complex, based on theoretical models are shown in Table 2. The predictions of cross-linked peptides were validated by visualizing them in 3D structures of protein complexes of $\alpha 1$, $\alpha 2$, and β subunits obtained by homology modeling according to the PDB templates: 1j2p (α -ring of the proteasome from *Archaeoglobus fulgidus*), 1ryp (20S proteasome from the yeast *Saccharomyces cerevisiae*), and 1fnt (20S proteasome from the yeast *Saccharomyces cerevisiae* in complex with the proteasomes activator PA26 from *Trypanosoma brucei*). A stoichiometry of seven $\alpha 1$, $\alpha 2$, and β subunits assembled in a homoheptameric rings was considered, and the coordinates of this α -, $\alpha 2$ -, and β -ring were obtained by DaliLite software. The DaliLite method is used routinely to compare newly solved structures against those in the PDB database and to compare predicted model structures to X-ray crystal structures based on sensitive measurements of geometrical similarity between protein structures [23]. Homoheptameric ring ensembles generated according to the PDB templates 1j2p, 1ryp, and 1fnt chosen based on Z-score and amino acid sequence identity (list of PDB files available on request, in Supplemental Information).

During a cross-linking reaction, the cross-linker can react with two amino acid residues within a single subunit, or with residues from two subunits to form a covalent dimer, which can further react to form an oligomeric complex, as was observed by means of SDS-PAGE (Figure S-1. Supplemental Information). Several $\alpha 1$ -specific linkages were detected by tandem MS after treatment of the $\alpha 1\beta\alpha 1$ CP subtype with the homobifunctional lysine-reactive BS2G ($\alpha 1$ K54-K68 and K44-K47 cross-linkage) and zero-length EDC ($\alpha 1$ K54-E67 and K163-E62 cross-linkages) (Table 2). With the CXMS approach, the $\alpha 1$ -specific peptides 44-TPEGVVLAAADKR-55 and 58-SPLMEPTSVEKIH-70 were clearly cross-linked with BS2G through lysine residues K54 and K68 (Fig. 1). Of all the lysine residues within $\alpha 1$, only the intramolecular distance between K54 and K68 (9.4 Å) was found (Fig. 2d) to accommodate the length of the BS2G cross-linker spacer arm (7.7 Å) and generate an $\alpha 1$ -interpeptide linkage. Although the BS2G cross-linker length and distance between the reactive amino acid residues differ, this discrepancy is consistent with previous CXMS analyses of proteins with known crystal structures in which the span between reactive side chain atoms is often slightly greater than the length of the cross-linker used to connect them [7, 27]. In accordance with the close proximity of the $\alpha 1$ lysine residues K54 and K68 in space, the $\alpha 1$ peptides in the region cross-linked by BS2G were also cross-linked with zero-length EDC (K54 and E67 were linked within $\alpha 1$ -peptides 44-TPEGVVLAAADKR-55 and 58-SPLMEPTSVEK-68), (Fig. S-2a., Supplemental Information). An intramolecular distance of 6.9 Å between K54 and E67 was measured and is acceptable for this EDC-

mediated linkage [28]. The $\alpha 1$ peptide 58-SPLMEPTSVEK-69 was also found connected to a peptide in the central region of $\alpha 1$, 150-LYETDPSGTPYEWKAVSIG-171, with fragment ions consistent with a linkage between K163 and either E62 or E67 (Fig. S-2b., Supplemental Information). While an intramolecular distance of greater than 30 Å between K163 and E62/67 was found, intermolecular cross-links between two $\alpha 1$ subunits were measured to be: 4.8 Å, 6.7 Å, and 9.7 Å, for models based on PDB coordinates of 1j2p, 1fnt, and 1ryp, respectively, as it is shown in Table 2. The EDC cross-link identified in the heptameric $\alpha 1$ -ring between K163 and E62 was best modeled using 1j2p as a template (Fig. 2a) and was estimated to have a length of 4.8 Å (Fig. 2b). The DaliLite Z-score was 41 with a sequence identity of 51% suggesting a high geometrical similarity between the *H. volcani* $\alpha 1$ ring and template 1j2p.

Interpeptide linkages of $\alpha 2$ analogous to $\alpha 1$ K54-K68 side chains cross-linked by the homobifunctional lysine reactive BS2G were not detected for the $\alpha 2\beta\alpha 2$ CP subtype. When the primary amino acid sequences of $\alpha 1$ and $\alpha 2$ were aligned it was found that $\alpha 1$ K54 was exchanged with $\alpha 2$ R53 (Fig. 3). Although $\alpha 2$ harbors a lysine at K67 (which corresponds to K68 of $\alpha 1$), $\alpha 2$ is missing the lysine residue that would correspond to K54 from $\alpha 1$ and thus cannot form a crosslink with BS2G at the $\alpha 2$ R53-K68 residues analogous to $\alpha 1$ K54-K68.

In contrast to the BS2G results, two new cross-links were observed in $\alpha 2$ compared to $\alpha 1$ after treatment of the $\alpha 2\beta\alpha 2$ CP subtype with EDC (Table 2). The EDC-derived $\alpha 2$ interpeptide linkage between K88 from 88-KLVDFAR-94 and D119 from 116-TITDNIQUESTQSGGTRPYGASLLIGGVENGSGR-148 was observed (Fig. S-2d., Supplemental Information). Between cross-linked $\alpha 2$ K88-D119 residues an intramolecular distance of 18.5 Å was measured, while intermolecular average distances of the same residues within the homoheptameric ring of $\alpha 2$ modeled based on coordinates of the 1j2p, 1fnt, and 1ryp PDB structures (as above for $\alpha 1$) were found to be: 3.1 Å, 4.8 Å, and 5.5 Å. Considering the flexibility of the side chains between which zero-length cross-links can form all measured distances are acceptable, but $\alpha 2$ model, with the best Z-score of 29.3 and the highest sequence identity of 42 %, based on the 1j2p structure was preferred, with an average distance of 3.1 Å for the K88-D119 linkage. In addition, in $\alpha 2$ an intrapeptide linkage within 68-LHKLDDALGAATAGHVADARK-88 was detected which generated fragment ions clearly suggesting a cross-link between K70 and D72/D73. MSX-3D revealed that the K70-D73 cross-linkage is preferred in the $\alpha 2$ model based on its shorter distance of 3.2 Å compared to the 6.6 Å for K70-D72. The expected $\alpha 2$ interpeptide K162-E61/E66 cross-link between 149-LFATDPSGTPQEWKAVAIGGHR-170 and 54-STPSELMEAEISIEKLHK-70, which corresponds to the $\alpha 1$ peptides cross-linked through K163-E62/E67 of $\alpha 1$, was not identified. While this does not definitively prove the absence of the $\alpha 2$ K162-E61/E66 cross-link, this cross-linkage was below the threshold of detection in our experiments.

With BS2G as a cross-linker, four dead-end modifications were detected in $\alpha 1$ tryptic peptides, suggesting that the highly-reactive lysine residues (K68, K71, K54 and K163) are on the surface of $\alpha 1$. In contrast, only two dead-end modifications were observed in $\alpha 2$ (K67, K70) (Table S1, Supporting Information).

Using both cross-linkers, the same interpeptide and intrapeptide cross-linkages of subunit β were identified in both the $\alpha 1\beta\alpha 1$ and $\alpha 2\beta\alpha 2$ CP subtypes (Table 2). An intrapeptide BS2G cross-link was detected between residues K44 and K47 of the β peptide 38-ADELGDKETKTGTTTGVIGKTEEGVVLATDMRASMGYMVSSK-78. The distance between β K44 and K47 using BS2G, was measured to be a short intramolecular distance of 6.1 Å by MSX-3D. Between β K78 from 69-ASMGYMVSSKDVQK-82 and β E212 from

209-SAVER-213, cross-linked with EDC (Fig. S-2c., Supplemental Information), the large intramolecular distance of >30 Å was measured. Intermolecular distances in the homoheptameric β_7 -ring and doublet of $\beta_7\beta_7$ -rings, modeled according to the 1ryp, 1fnt, and 1j2q (20S proteasome from *Archaeoglobus fulgidus*) PDB templates, were calculated and found to differ significantly. The intermolecular distance between the β K78-E212 residues within *two* rings $\beta_7\beta_7$, was measured to be 4.6 Å in 1fnt for the configuration shown in Fig. 2c., and 22.1 Å for 1ryp-based structure. In contrast, distances of 29.4 Å, 16.4 Å, and 26.5 Å were measured between the same residues within *one* ring for 1fnt-, 1j2q-, and 1ryp- based structures, respectively. Thus, the model based on 1fnt appears more accurate for the packing of the $\beta_7\beta_7$ -rings than the models based on 1ryp and 1j2q.

4 Concluding remarks

The paucity of known crystal structures and the extremely acidic properties of haloarchaeal proteins complicate their structural analysis. The CP subtypes ($\alpha_1\beta\beta\alpha_1$ and $\alpha_2\beta\beta\alpha_2$) from *H. volcanii* are presumed to be associated as four-stacked heptameric rings in an $\alpha_7\beta_7\beta_7\alpha_7$ symmetry based on analogy to non-halophilic archaeal CPs for which detailed X-ray crystal structures are known. In this paper, we demonstrate that CXMS (previously performed only for non-halophilic proteins of relatively neutral pI) coupled with atomic structural data (determined experimentally for non-halophilic protein homologs) provides useful distance constraints for extremely acidic halophilic protein complexes (proteasomal CPs from the archaeon *H. volcanii*). Several observed cross-links were used to validate the predicted 3D-structures of the *H. volcanii* proteasomal α - and β -type subunits, and intermolecular distance constraints were used to assemble these proteins (α_1 , α_2 , β) into the quaternary structure of the CP complexes. Distance constraints obtained by this CXMS study facilitated selection of an accurate model from several possible PDB proteasomal models and assisted in the determination of the arrangement of the α_1 -, α_2 -, and β -homoheptameric rings in the complex protein structures

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded in part by: NIH R01 GM057498 and DOE DE-FG02-05ER15650 to JMF; 1S10RR023044-01 and 1U54AI57141-01 to DRG; and Fulbright Association and MNRS-III 43004 to IMK. Authors thank Dr. Priska von Haller (Proteomics Resource, University of Washington) for her assistance with mass spectrometry analysis. Special thanks to Yi-Hsuan Tsai (University of North Carolina) for assistance with bioinformatics and Branko Drakulic (University of Belgrade) for homology modeling.

Abbreviations

CXMS	cross-linking mass spectrometry
CP	core particle
T3SS	type three secretion system
NHS	N-hydroxysuccinimide
BS2G	bis[sulfosuccinimidyl] glutarate- d_0
EDC	1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
DB	data bank

References

1. Soppa J, Baumann A, Brenneis M, Dambeck M, et al. Genomics and functional genomics with haloarchaea. *Arch Microbiol.* 2008; 190:197–215. [PubMed: 18493745]
2. Zhou G, Kowalczyk D, Humbard MA, Rohatgi S, et al. Proteasomal components required for cell growth and stress responses in the haloarchaeon *Haloferax volcanii*. *J Bacteriol.* 2008; 190:8096–8105. [PubMed: 18931121]
3. Wilson HL, Aldrich HC, Maupin-Furlow JA. Halophilic 20S proteasomes of the archaeon *Haloferax volcanii*: purification, characterization, and gene sequence analysis. *J Bacteriol.* 1999; 181:5814–5824. [PubMed: 10482525]
4. Kaczowka SJ, Maupin-Furlow JA. Subunit topology of two 20S proteasomes from *Haloferax volcanii*. *J Bacteriol.* 2003; 185:165–174. [PubMed: 12486053]
5. Singh P, Shaffer SA, Scherl A, Holman C, et al. Characterization of protein cross-links via mass spectrometry and an open-modification search strategy. *Anal Chem.* 2008; 80:8799–8806. [PubMed: 18947195]
6. Sanowar S, Singh P, Pfuetzner R, André I, et al. Interactions of the transmembrane polymeric rings of the *Salmonella enterica* serovar typhimurium type III secretion system. *MBio.* 2010; 1:e00158–10. pii. [PubMed: 20824104]
7. Leitner A, Walzthoeni T, Kahraman A, Herzog F, et al. Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol Cell Proteomics.* 2010; 9:1634–49. [PubMed: 20360032]
8. Petrotchenko EV, Borchers CH. Crosslinking combined with mass spectrometry for structural proteomics. *Mass Spectrom Rev.* 2010; 29:862–76. [PubMed: 20730915]
9. Singh P, Panchaud A, Goodlett DR. Chemical cross-linking and mass spectrometry as a low-resolution protein structure determination technique. *Anal Chem.* 2010; 82:2636–42. [PubMed: 20210330]
10. Karadzic I, Maupin-Furlow JA. Improvement of two-dimensional gel electrophoresis proteome maps of the haloarchaeon *Haloferax volcanii*. *Proteomics.* 2005; 5:354–359. [PubMed: 15627962]
11. Cohen FE, Sternberg MJ. On the use of chemically derived distance constraints in the prediction of protein structure with myoglobin as an example. *J Mol Biol.* 1980; 137:9–22. [PubMed: 7365797]
12. Ngai SM, Sönnichsen FD, Hodges RS. Photochemical crosslinking between native rabbit skeletal troponin C benzoylbenzoyl-troponin I inhibitory peptide residues 104–115. *J Biol Chem.* 1994; 269:2165–2172. [PubMed: 8294472]
13. Schulz DM, Kalkhof S, Schmidt A, Ihling C, et al. Annexin A2/P11 interaction: new insights into annexin A2 tetramer structure by chemical cross-linking, high-resolution mass spectrometry, and computational modeling. *Proteins.* 2007; 69:254–269. [PubMed: 17607745]
14. Balasu MC, Spiridon LN, Miron S, Craescu CT, et al. Interface analysis of the complex between ERK2 and PTP-SL. *PLoS One.* 2009; 4:e5432. [PubMed: 19424502]
15. Heymann M, Paramelle D, Subra G, Forest E, et al. MSX-3D: a tool to validate 3D protein models using mass spectrometry. *Bioinformatics.* 2008; 24:2782–2783. [PubMed: 18826958]
16. Amy Lam Y, Lawson T, Velayutham M, Zweier, et al. A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal. *Nature.* 2006; 441:763–767. [PubMed: 11961560]
17. Kao A, Chiu C, Vellucci D, Yang Y, et al. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Mol Cell Proteomics.* 2011; 10:1–17.
18. Wendoloski D, Ferrer C, Dyal-Smith ML. A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. *Microbiology.* 2001; 147:959–964. [PubMed: 11283291]
19. Allers T, Ngo HP, Mevarech M, Lloyd RG. Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl Environ Microbiol.* 2004; 70:943–953. [PubMed: 14766575]
20. Humbard MA, Zhou G, Maupin-Furlow JA. The N-terminal penultimate residue of 20S proteasome $\alpha.1$ influences its N^α-acetylation and protein levels as well as growth rate and stress responses of *Haloferax volcanii*. *J Bacteriol.* 2009; 191:3794–3803. [PubMed: 19376868]

21. Panchaud A, Singh P, Shaffer SA, Goodlett DR. xComb: a cross-linked peptide database approach to protein-protein interaction analysis. *J Proteome Res.* 2010; 9:2508–2515. [PubMed: 20302351]
22. Schilling B, Row RH, Gibson BW, Guo X, et al. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom.* 2003; 14:834–50. [PubMed: 12892908]
23. Holm L, Park J. DaliLite workbench for protein structure comparison. *Bioinformatics.* 2000; 16:566–567. [PubMed: 10980157]
24. Pedretti A, Villa L, Vistoli G. VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. *J Mol Graph Model.* 2002; 21:47–49. [PubMed: 12413030]
25. Groll M, Brandstetter H, Bartunik H, Bourenkow G, et al. Investigations on the maturation and regulation of archaeobacterial proteasomes. *J Mol Biol.* 2003; 327:75–83. [PubMed: 12614609]
26. Löwe J, Stock D, Jap B, Zwickl P, et al. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science.* 1995; 268:533–539. [PubMed: 7725097]
27. Jacobsen RB, Sale KL, Ayson MJ, Novak P, et al. Structure and dynamics of dark- state bovine rhodopsin revealed by chemical cross-linking and high-resolution mass spectrometry. *Protein Sci.* 2006; 15:1303–1317. [PubMed: 16731966]
28. Lee JY. Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes. *Mol Biosyst.* 2008; 4:816–823. [PubMed: 18633483]

Fig. 1. Tandem mass spectrum of peptides SPLMEPTSVEKIHK -TPEGVVLAADKR cross-linked with BS2G (m/z=590.12). Fragment ions of cross-linked peptides are marked with an α if belong to the larger peptide (SPLMEPTSVEKIHK) or with β if belong to the smaller peptide (TPEGVVLAADKR). Fragment ions containing cross-linked sequences from both peptide chains were assigned using MS2Assign. Specific b and y ions were observed within 10 ppm measured mass accuracy. More than 50% of the fragment ions were identified by the program.

Fig. 2.

Structural models of the α_{17} -ring and $\beta_7\beta_7$ -rings of *H. volcanii* 20S CPs. a) Heptameric α_{17} -ring modeled according 1j2p PDB template, b) distances between the NZ of K163 and OE1 of E-62 from the two α_1 subunits estimated at 4.7 Å, c) the model of the $\beta_7\beta_7$ -rings was assigned according to the PDB template 1fnt, as H, I, V, and b. The measured distance between K-78 from the b chain and E-212 from the H chain was 4.64 Å, d) MSX-3D predicts that lysine K54 from TPEGVVLAADKRSR cross-linked with lysine K68 from SPLMEPTSVEKIHK is at a distance of 9.45 Å in the model structure of the α_1 subunit.

Fig. 3. Alignment of the $\alpha 1$ and $\alpha 2$ sequences using ClustalW. Conserved regions are shaded in gray.

Fig. 3. Alignment of the $\alpha 1$ and $\alpha 2$ sequences using ClustalW. Conserved regions are shaded in gray.

Table 1

Strains and plasmids used in this study.

Strain or plasmid	Description	Source or reference
Strain: <i>Escherichia coli</i>		
DH5α	F ⁻ <i>recA1 endA1 hsdR17(r_k⁻ m_k⁺) supE44 thi-1 gyrA relA</i>	Life Technologies
GM2163	F ⁻ <i>ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 rpsL 136 dam13::Tn9 xylA5 mtl-1 thi-1 mcrB1 hsdR2</i>	New England Biolabs
<i>Haloferax volcanii</i>		
DS70	Wild-type isolate DS2 cured of plasmid pHV2	[18]
H26	DS70 Δ <i>pyrE2</i>	[17]
GZ114	H26 Δ <i>psmC</i> (α2 minus)	[2]
GZ130	H26 Δ <i>psmA</i> (α1 minus)	[2]
Plasmid:		
pJAM816	Ap ^r Nv ^r ; pHV2 ori-based plasmid for synthesis of β-StrepII in <i>H. volcanii</i>	[20]
pJAM2545	Ap ^r Nv ^r ; pHV2 ori-based plasmid for synthesis of β-StrepII and α1-His6 in <i>H. volcanii</i>	[20]
pJAM2564	Ap ^r Nv ^r ; pHV2 ori-based plasmid for synthesis of β-StrepII and α2-His6 in <i>H. volcanii</i>	This study

Table 2

Cross-linked peptides identified from $\alpha 1\beta\beta\alpha 1$ and $\alpha 2\beta\beta\alpha 2$ proteasomal CPs after treatment with homobifunctional lysine reactive BS2Gandzero-length EDC.

Measured precursor mass (Da)	Charge state	MMA (ppm)	Cross-linker	Subunit, amino acid residue no.	Sequence 1 —Sequence 2	Cross-linked sites	Distance intramolecular ($\text{\AA}$) ^a	Distance intermolecular ($\text{\AA}$) ^b
$\alpha 1\beta\beta\alpha 1$ CP:								
2945.56	5	7.6	BS2G	$\alpha 1$, 44-55 $\alpha 1$, 58-70	TPEGVVLAA <u>D</u> KR — SPLMEPTSV <u>E</u> KIHK	K54-K68	9.4	37.1 (lj2p) 37.8 (lfnt) 36.3 (lryp)
2453.28	4	5.7	EDC	$\alpha 1$, 44-55 $\alpha 1$, 58-68	TPEGVVLAA <u>D</u> KR — SPLMEPTSV <u>E</u> K	K54-E67	6.9	37.9 (lj2p) 36.9 (lfnt) 37.9 (lryp)
3652.74	4	1.7	EDC	$\alpha 1$, 150-171 $\alpha 1$, 58-68	LYETDPSGTPYEWK <u>A</u> VSIGADR — SPLMEPTSV <u>E</u> K	K163-E62	39.3	4.8 (lj2p) ^b 6.7 (lfnt) 9.7 (lryp)
4418.12 (1 ox)	4	3.9	BS2G	β , 38-78, intrapeptide	ADELGD <u>K</u> ETKTGTITTVGIKTEEGVVVLATDMRASMGYMVSSK	K44-K47	6.1	/
2072.03	4	4.9	EDC	β , 69-82 β , 209-213	ASMGYMVSSK <u>D</u> VQK — SA <u>V</u> ER	K78-E212	30.9	29.4/4.6 ^c (lfnt) 16.4 (lj2q) 26.5/22.1 ^c (lryp)
$\alpha 2\beta\beta\alpha 2$ CP:								
4164.12	4	2.4	EDC	$\alpha 2$, 116-148 $\alpha 2$, 88-94	TITDNIQESTQSGGTRPYGASLLJGGVENGSGR — <u>K</u> LVDFAR	K88- D119	18.5	3.1 (lj2p) 4.8 (lfnt) 5.5 (lryp)
2115.06	4	3.5	EDC	$\alpha 2$, 68-88, intrapeptide	LHK <u>L</u> DDALGAATAGHVADARK	K70-D72	3.2	/
4418.12 (1 ox)	4	4.2	BS2G	β , 38-78, intrapeptide	ADELGD <u>K</u> ETKTGTITTVGIKTEEGVVVLATDMRASMGYMVSSK	K44-K47	6.1	/
2072.03	4	4.9	EDC	β , 69-82 β , 209-213	ASMGYMVSSK <u>D</u> VQK — SA <u>V</u> ER	K78-E212	30.9	29.4/4.6 ^c (lfnt) 16.4 (lj2q) 26.5/22.1 ^c (lryp)

^aDistance between cross- linked amino acid residues within the subunit calculated by MSX-3D

^b Intermolecular distance of cross-linked amino acid residues between subunits calculated as average using modeled structures based on lj2p, lfnt, lryp, and lj2q PDB templates

^c Intermolecular distance between cross- linked amino acid residues from subunits within one ring $\beta 7$ /distance between cross- linked amino acid residues from subunits within two rings $\beta 7\beta 7$, calculated as average using modeled structures based on lj2p, lfnt, lryp, and lj2q PDB templates