IDENTIFICATION AND STRUCTURE-FUNCTIONAL ANALYSIS OF THE SPECIFICITY DETERMINING RESIDUES OF THE ALPHA SUBUNITS OF THE PROTEOSOMAL COMPLEX

***Baryshev P.B.*[*1], *Afonnikov D.A.*[1, 2], *Nikolaev S.V.*[2]**
[1] Novosibirsk State University, Novosibirsk, 630090, Russia; [2] Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia
[*] Corresponding author: e-mail: pavel1983@ngs.ru

Key words: proteosome, protein-protein interactions, evolution, gene duplication, specificity determining positions

SUMMARY

Motivation: Proteosomes are polyenzymatic proteolytic structures that provide the degradation of the bulk of cytoplasmic proteins to oligopeptides. The proteosomal genes in the eukaryotes all arose by duplication of a single ancestral gene encoding the proteosomal subunits in the bacteria. The analysis of evolutionary events after duplication may be useful for discovering new information about proteosomal structural and functional properties.

Results: We confine our study here to the detection of the positions of the α-subunits whose amino acid substitutions are specific to particular subunits of the proteosomal alpha-rings. We detected a set of the α-subunit positions whose substitutions are specific to the genes that encode the various proteosomal subunits. It was demonstrated that these specific amino acid substitutions are the features of residues that form the subunit contacts in the α-ring of the proteosomes.

Availability: The proteosomal sequences, multiple sequence alignments and phylogenetic tree used in analysis are available upon request.

INTRODUCTION

According to the current concepts, the active moiety of the proteosome (20S) results in self-assembly of the subunits, a ring of seven α-subunits is assembled first, then a ring of β-subunits is built in (Kopp *et al*., 1997). It has been suggested that the order of the subunits in the proteosomal complex is fixed, i.e. each subunit in the ring occupies strictly defined place. Like in the case of self-assembly the proteosome, the order is defined by complementary interaction of the subunits, dependent on the spatial and physicochemical complementarity of the interacting parts of the macromolecules.

The evolutionary history of the α-subunit encoding genes in the eukaryotes is that they all arose by duplication of a single ancestral gene encoding the α-subunit in the bacteria (DeMartino, Slaughter, 1999); genome early during eukaryotic phylogeny. After duplication, as a result of a divergent evolution, each paralog gave rise to a group of orthologs, with each coding for 1 to 7 subunits that form the α-proteosomal ring in eukaryotes, including yeast and mammals (Bouzat *et al*., 2000). This model of evolution, based on the phylogenetic analysis of protein sequences, underlies the currently accepted

classification of the α-subunits in the paralogous groups. The model also implies that, after duplication, the α-subunit encoding genes kept accumulating mutations under selection pressure designed to maintain the stable ordering of this multi-subunit-structure (Nikolaev, Afonnikov, 2004). Thus, detection of such mutations and their analysis would give important information about how the features of this multisubunit structure might have formed.

We confine our study here to the detection of the positions of the α-subunits whose amino acid substitutions are specific to particular subunits of the proteosomal alpha-rings. To this end, we used the method implemented in the SDP program (Kalinina *et al*., 2004). We detected a set of the α-subunit positions whose substitutions are specific to the genes that encode the various proteosomal subunits. It was demonstrated that these specific amino acid substitutions are the features of residues that form the subunit contacts in the α-ring of the proteosomes.

MATERIALS AND METHODS

The sequences of the 20S proteosome subunits were retrieved from the SWISS-PROT database (Boeckmann *et al.*, 2003). An additional database search of homologous sequences was done using the BLASTP program (Altschul *et al.*, 1997). As a result, additional members of the proteosomal α-subunit family were chosen.

The CLUSTALW program (Thomson *et al*., 1994) was applied for the multiple alignment of the sequences. Analysis of the phylogenetic tree built by CLUSTALW program allowed us to assign the α-subunit sequences to paralogous groups. The groups include sequences from species exemplifying all the seven subunits types. After group assignment, the yielded multiple alignment was used to assess the conservation/variability at protein positions.

To define the positions with the subunit-specific mutations we used the SDPPred program (Kalinina *et al*., 2004). To estimate the significance of the positions the SDPPred uses the mutual information values. The values express the relation between the amino acid type at a given position and the index of paralog group (in our case it was the subunit index in the proteosomal ring from A to G) calculated as $I_i = \sum_{x,y} f(x_i, y) \cdot \log \frac{f(x_i, y)}{f(x_i) \cdot f(y)}$, where $f(x_i)$ is the occurrence rate of the amino acid x at the position i of the multiple sequence alignment, $f(y)$ is the fraction of the proteins assigned to the paralog group y, $f(x_i, y)$ is the occurrence frequency of the amino acid type x at the position i of the proteins in the paralog group y.

The identified positions were mapped to the proteosomal 3D structure (Unno *et al.*, 2002; PDB identifier 1IRU). The program iMoltalk (http://i.moltalk.org) was used to determine inter-subunit contact positions.

RESULTS AND DISCUSSION

In the course of the preliminary search, we choose 193 sequences of 35 species, of which 4 belonged to bacteria, 7 to archea and other to eukaryotes. Based on the surveyed phylogenetic tree we choose 7 paralogous groups. Each group corresponded to the homologs of the particular subunit of the proteosomal α-ring. This group assignment is based on the idea that the family genes resulted from single or series of duplications followed by sequence divergence. Therefore, homology among the sequence of different members of the family are supposed the result of sequence divergence after specific

events (the orthologous genes) or after the duplication of genes within an ancestral species (the paralogous genes) (Bouzat *et al*., 2000).

A SDPPred program was used to analyze the set of aligned sequences of the α-subunit sequences chosen as described above. SDPPred detected 25 positions at which the amino acid residues were conserved among the orthologs and different among the paralogous groups. Using the iMoltalk program we obtained that every subunit has at least 8 contact positions with the other α-subunits (positions 48, 53, 54, 72, 105, 209, 215, 224 of the multiple sequence alignment). Obtained contact positions are shown on Fig. 1. Moreover, certain amino acid residues form associations with the β-subunits. Conservatism of the remaining positions allow us to assume their importance for formation of the spatial proteosome structure. Mutations at such positions can result in incorrect folding of subunits and disrupt proteosomal complex formation. The chi-square test was applied to determine the significance of the positions detected by SDPPred program and the structural data (Table 1). We estimated the significance between the specific fixation the amino acid residues with respect to the subunit index, with the involvement of such position in the protein-protein interface for each of the α-subunit protein chain. The results shows, that the significance varies between subunits approaching the 90 % significance level.

The results suggested that during early phylogenesis, duplication in the subunit sequences was followed by mutations of residues that forms protein-protein interface and were important for the specific packing of subunits in the proteosomal machine.

Table 1. List of positions assigned as specificity-determining by SDPPred program and amino acids specific to the subunit sequence in the structure 1IRU. CP: 15 positions forming inter-subunit contacts; FP: 10 positions are not in contact with other subunits, likely responsible for the formation of the secondary and tertiary subunit structure

No.	SDP	Paralogous groups						
	CP	Chain A	Chain B	Chain C	Chain D	Chain E	Chain F	Chain G
1	48	S	E	S	S	R	R	G
2	53	R	F	S	R	R	N	L
3	54	H	S	R	A	G	D	S
4	55	I	L	T	I	V	V	A
5	61	E	S	E	D	E	Q	D
6	72	K	A	E	E	E	E	K
7	76	Q	G	H	K	L	Q	N
8	105	G	A	A	G	T	S	C
9	208	S	M	K	K	A	T	V
10	209	Q	Q	Q	Q	L	Q	H
11	215	A	G	G	N	−	Y	S
12	224	M	V	K	R	S	R	V
13	260	E	E	K	−	−	D	V
14	290	F	W	W	W	C	C	Y
15	301	K	N	N	S	A	R	A
	FP	Chain A	Chain B	Chain C	Chain D	Chain E	Chain F	Chain G
1	63	R	K	R	H	R	R	R
2	65	Y	V	Y	F	F	H	F
3	113	T	T	A	V	V	A	−
4	116	Q	E	R	K	R	R	E
5	159	M	Y	V	F	M	I	V
6	358	A	A	V	A	I	A	I
7	362	L	L	T	V	V	T	V
8	363	S	K	M	V	M	L	H
9	364	T	E	D	Q	E	P	D
10	410	K	G	V	L	N	E	N

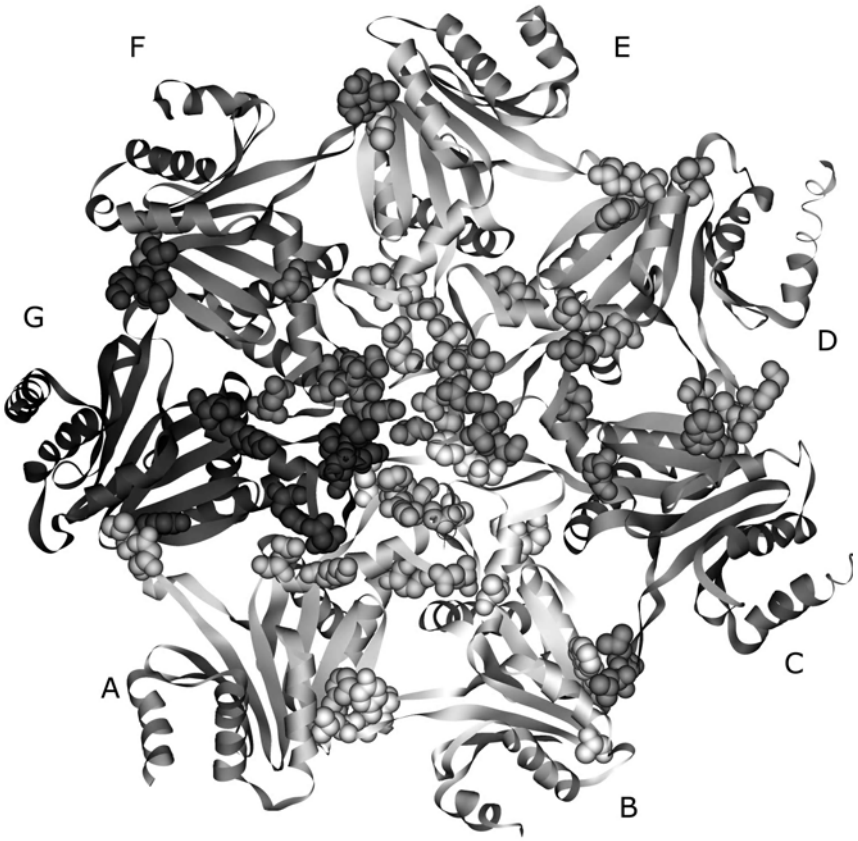

Figure 1. Spatial structure of the α-subunit ring. Subunit indices A-G are shown. SDP residues are shown in ball representation.

ACKNOWLEDGEMENTS

The work is supported by the U.S. Civilian Research & Development Foundation for the Independent States of the Former Soviet Union (CRDF) within the Basic Research and Higher Education Program (Y1-B-08-20, REC-008), the Ministry of Education grant РНП.2.1.1.4935, Russian Foundation for Basic Research (05-04-49141-a).

REFERENCES

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.*, **25**, 3389–3402.

Boeckmann B., Bairoch A., Apweiler R., Blatter M.-C., Estreicher A., Gasteiger E., Martin M.J., Michoud K., O'Donovan C., Phan I., Pilbout S., Schneider M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucl. Acids Res*., **31**, 365–370.

Bouzat J.L., McNeil L. K., Robertson H.M., Solter L.F., Nixon J.E., Beever J.E., Gaskins H.R., Olsen G., Subramaniam S., Sogin M.L., Lewin H.S. (2000) Phylogenomic analysis of the α proteosome gene family from early-diverging eukaryots. *J. Mol. Evol.*, **51**, 532–543.

DeMartino G.N., Slaughter C.A. (1999) The proteosome, a novel protease regulated by multiple mechanisms. *J. Biol. Chem.*, **274**, 22123–22126.

Kalinina O.V., Mironov A.A., Gelfand M.S., Rakhmaninova A.B. (2004) SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins. *Nucl. Acids Res.*, **32**, W424–W428.

Kopp F., Hendil K.B., Dahlmann B., Kristensen P., Sobek A., Uerkvitz W. (1997) Subunit arrangement in the human 20S proteosome. *Proc. Natl. Acad. Sci. USA*, **94**, 2939–2944.

Nikolaev S.V., Afonnikov D.A. (2004) Inter-subunit contacts of the proteosomal alpha-subunits as determinants of paralog groups. *Proceedings of the Fourth International Conference on Bioinformatics of Genome Regulation and Structure*, Novosibirsk, Russia, 2004, **1**, p. 319–322.

Thompson J.D., Higgins D.G., Gibson T.J. (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap-penalties and weight matrix choice. *Nucl. Acids Res*., **22**, 4673–4680.

Unno M., Mizushima T., Morimoto Y., Tomisugi Y., Tanaka K., Yasuoka N., Tsukihara T. (2002) The structure of the mammalian 20S proteosome at 2.75 A resolution. *Structure*, **10**, 609–618.