Digest Journal of Nanomaterials and Biostructures Vol. 8, No. 2, April - June 2013, p. 639 - 654

STRONG AND WEAK HYDROGEN BONDS IN Sm/LSm OLIGOMERIC ASSEMBLIES: A COMPARISON OF INTRA- AND INTERCHAIN INTERACTION

BOŽIDARKA L. ZARIĆ*, MILICA V. BUKOROVIĆ, AND SRĐAN Đ. STOJANOVIĆ
ICTM - Department of Chemistry, University of Belgrade, Belgrade, Serbia

The Sm and Sm-like (LSm) proteins are a widespread protein family with members in all kingdoms of life. Sm proteins form complexes engaging in various RNA-processing events. Sm proteins do form and act as oligomeric assemblies whose characteristic is their exceptional stability. This study compares strong and weak hydrogen bonds in the interior of monomers and at interfaces of Sm/LSm proteins in order to better understand the stability of oligomers. According to our results, the stability of oligomeric assemblies is achieved by CH···O, NH···O and CH···N interactions including, NH···N, OH···O,XH···π interactions present in small percentages. Intrachain hydrogen bonds behave in respect to geometry, distances and angles, like interchain hydrogen bonds. It is also shown that amino acids Arg and Lys participate significantly as donors or acceptors in some of the strong or weak interactions at interfaces to a higher extent than in the monomers. There is a trend for most polar amino acids to cross into more solvent exposed position in interfaces, which is not the case for nonpolar or charged amino acids. There is no exclusive preference for particular secondary structure both for intrachains and for interfaces.

(Received December 14, 2012; Accepted April 12, 2013)

Keywords: Sm/LSm proteins; Hydrogen bonds; Secondary structure preferences; Solvent accessibility

1. Introduction

Hydrogen bonding plays a key role in structure and function of proteins including features such as protein folding, ligand recognition, hydration, as well as local architecture, enzymatic activity and molecular dynamics [1,2]. The hydrogen bonds are manifested in a variety of strengths and geometries. In hydrogen bonds, hydrogen atoms of O-H, N-H or S-H groups (known as hydrogen bond donors), interact with nonbonding electrons of acceptor atoms (for example O, N, or S). The bonding energies of such hydrogen bonds are lower than energies of covalent interactions [3,4,5]. Accordingly, hydrogen bonds like O-H···O, N-H···O, O-H···N and N-H···N may be considered to be strong, whereas interactions like C-H···O, C-H···N, O-H···π, N-H···π and C-H···π are weak [6]. The importance of conventional interactions such as hydrogen bonds (mentioned above as strong hydrogen bonds), salt bridges and hydrophobicity in protein oligomers are well established [1,7].

A set of weaker interactions have also been recognized to play an important role in the stability and structure of proteins [8,9]. The existence of weak hydrogen bonds was already previously well documented but their importance was not timely appreciated [10]. Only in recent years importance of weaker interactions in various processes have been recognized [8,11,12]. Sets of these weak hydrogen bonds include C-H···π, N-H···π and O-H···π interactions, as well as interactions between aromatic side-chains; C^{α}-H···O=C, and C-H···N interactions. It has been reported that typical energies of covalent bonds are 100–200 kcal/mol, depending on the extent of

*Corresponding author: bzaric@chem.bg.ac.rs

unsaturation in the bonds. Weak interactions play a modest individual influence on chemical structures, however their cumulative effect can be profound and has a large influence on the conformational stability of a biomolecule [13,14].

What contributes to the stability of protein oligomers is the delicate balance between a variety of weak and strong non-covalent interactions. Hydrogen bonds, salt bridges, and hydrophobic interactions are major determinants of structural stability. The weak hydrogen bonds have been shown to be of much greater importance than previously thought [15,16].

The Sm and Sm-like (LSm) proteins are an ancient and widespread protein family together with members in all living kingdoms. *Arcaeabacteria* harbour between one or two Sm/LSm proteins. The *Escherichia coli* Hfq protein and its ortologues represent a family in several bacterial lineages. Genomes of eukaryotes contain a minimum of 24 Sm/LSm genes. Phylogenetic distribution suggests that the family underwent an explosive diversification with the advent of eukaryotes [17,18].Those proteins generally mediate RNA-protein interactions. Members of this protein family are small (9-29 kDa) proteins which lack other domains but may contain N or C terminal extensions [17,18]. Their characteristic is that they all form homo or hetero oligomeric rings which contain six, seven or eight subunits [19- 26,].Individual proteins are characterized by the conserved bipartite Sm fold, composed of Sm motif 1 and Sm motif 2. Solved structures of members in this family, do show that the fold is highly conserved and this is defined by an N-terminal helix followed by a five stranded anti-parallel β sheet. Strands β1- β3 are a part of Sm motif 1 and strands β4 and β5 are implemented in the Sm motif 2. The five stranded β sheet is strongly bent in the middle and the conserved hydrophobic residues form the hydrophobic core [27]. All Sm proteins form structures of a higher order which can be defined or none defined. In general, they are very stable and sometimes the presence of chaotropic agent is necessary for their disruption [28, 29]. We have previously reported [30] contributions of interface hydrophobic interactions, hydrogen bonds and salt bridges to the stability of Sm oligomers. Stabilization centres (SC) of Sm proteins and contribution of non-canonical interactions to the stability of interfaces have been also analyzed [31]. In our another work [32], we showed that the hot spots of Sm/LSm proteins are located within densely packed interface regions, they are highly conserved and have large energy contributions to the interface interactions.

In this current study we further elaborated studies on Sm/LSm oligomeric assemblies in an effort to understand the origin of their stability. In addition, we have systematically analyzed all strong and weak hydrogen bonds (NH···O, OH···O, NH···N, OH···N, CH···O, CH···N, NH···S, OH···S, CH···S, NH···π, OH···π, CH···π). and their various sub-types in Sm proteins. We analyzed and compared all mentioned hydrogen bonds in the interior of monomers and at interfaces. Protein data set used for this work was the same as in our previous studies [30,31].

2. Experimental

For this study we used the Protein Data Bank (PDB) 19. June 2010 list of 68266 structures. The following criteria were employed to assemble the set: (1) no theoretical model structures and no NMR structures were accepted, (2) only crystal structures with the resolution of 3.0 Å or better and a crystallographic R-factor of 25.0% or lower were accepted, (3) crystal structures of proteins containing Sm-like fold (SCOP Classification, version 1.75) without RNA binding were accepted. If not already present, all hydrogen atoms were added and optimized using the program REDUCE [33] with default settings.

To reduce biased statistics, caused by the lack of hetero-oligomer proteins in the dataset, we did not divide dataset into homo and hetero subdatasets. After the dataset had been assembled, several proteins that contained ligands were rejected, leaving 15 Sm/LSm proteins that were actually used as the dataset in our analysis (Table 1).

Table 1. Dataset of the Sm/LSm proteins.

Protein	Genetic source	Number of subunits	Number of amino-acid residues in single subunit	Resolution (Å)	PDB Code
SmD1D2	Human	1 1	119 (D1) 118 (D2)	2.50	1b34
SmD3B	Human	6 6	75 (D3) 91 (B)	2.00	1d3b
PA-Sm1	*Pyrococcus abyssi*	28	71	1.90	1h64
HFQ	*Escherichia coli*	6	74	2.15	1hk9
AF-Sm1	*Archaeoglobus fulgidus*	28	77	2.50	1i4k
Sm	*Pyrobaculum aerophilum*	7	81	1.75	1i8f
Mth649	*Methanobacterium thermautotrophicum*	7	86	1.85	1jbm
HFQ	*Staphylococcus aureus*	12	77	1.55	1kq1
SmAP3	*Pyrobaculum aerophilum*	28	130	2.00	1m5q
Sm	*Methanobacterium thermoautotrophicum*	7	83	1.70	1mgq
SmF	*Saccharomyces cerevisiae*	7	93	2.80	1n9r
Sm	*Sulfolobus solfataricus*	14	81	1.68	1th7
HFQ	*Pseudomonas aeruginosa*	6	82	1.60	1u1s
LSm5	*Cryptosporidium parvum*	2	121	2.14	2fwk
LSm3	*Saccharomyces cerevisiae*	2	96	2.50	3bw1

Interface areas and interface (interchain) residues were calculated using the "Protein interfaces, surfaces and assemblies service PISA" at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html; [34]).For calculation of various types of hydrogen bonds and their properties, the HBAT program [6] with default settings was used. In order to assign secondary structure preferences for amino acids involved in strong and weak hydrogen bonds, we used a homemade program. The information about secondary structures and solvent accessibility of the proteins were obtained using the program DSSP [35]. Solvent accessibility was divided into three classes, buried, partially buried and exposed, indicating respectively the least, moderate and high accessibility of the amino acid residues to the solvent [36]. The empirical Bayesian method was used to calculate amino acid conservation scores by the ConSurf server [37].Homologues were collected from SWISS-PROT, max. number of homologues = 50, number of PSI-BLAST iterations = 1 (PSI-BLAST E-value = 0.001), and conservation scores ranged from 9 (conserved) to 1 (variable). For testing statistical significance of mean differences we used non-parametric Kolmogorov-Smirnov two-sample test.

3. Results and discussion

In order to better understand the stability of Sm/LSm oligomeric assemblies, we analyzed distribution and characteristics of hydrogen bonds, donor and acceptor role of amino acids, secondary structure preferences of amino acids which participate in hydrogen bonds as well as solvent accessibility of amino acids involved in hydrogen bonds.

All analysis, except the lengths, angles and the last analysis have been performed separately for interior of monomers and for interfaces in order to recognize possible differences and their importance for the stability of oligomeric assemblies.

3.1. Distribution of strong and weak hydrogen bonds

The present study focuses on the strong and weak hydrogen bonds, contributing to the global stability of the Sm/LSm proteins. The number of amino acid residues was correlated with the number of the strong and weak hydrogen bonds in the considered set of Sm/LSm proteins. It

could be inferred that the correlation (r=0.853) is somewhat higher than in the interface hydrogen bonds (r=0.763) [30].

Furthermore, the average number of interface hydrogen bonds per residue is 0.22 whereas the collective contribution of the intrachain hydrogen bonds is 1.26 per residue. This difference in the number of hydrogen bonds does indicate that the contribution to the overall stability of Sm/LSm proteins is not dictated by the number of amino acids. The significantly larger number of the intrachain hydrogen bonds is due to the fact that protein interiors consist mostly of residues which form well defined α-helices and β-sheets.

The percentage contribution of various types of strong and weak hydrogen bonds in the intrachain of Sm/LSm proteins and at interfaces (interchain) in our dataset is shown in Figure 1. The hydrogen bond abbreviation consists of three parts: hydrogen bond type, donor, acceptor. B stands for backbone, S is side-chain, D is donor, and A is acceptor. For example {CHO BD SA} denotes a CH···O hydrogen bond involving a backbone CH donor and a side-chain O-atom acceptor.

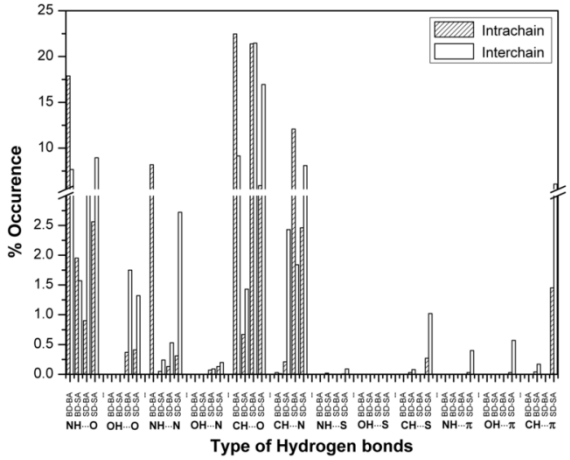

Fig. 1. Distribution of analyzed strong and weak hydrogen bond types in the Sm/LSm proteins.

The distribution of hydrogen bonds, on the basis of data in Sm/LSm proteins in our dataset, in a total of 127,893 hydrogen bonds are shown in Figure 1, that is, on average there are 761 H-bonds present in each chain. The present dataset was divided into intrachain and interchain subdatasets. The population of intrachain hydrogen bonds from the backbone and the side-chain donor is 51.5% and 48.5% respectively, while for hydrogen bonds from the backbone and the side-chain acceptor is 83.5% and 16.5% respectively. The significantly larger number of the backbone acceptors is due to the fact that carbonyl oxygen from peptide bonds mostly form weak CH···O hydrogen bonds (Figure 1). This is not surprising, since most C–H donor groups belong to the C^{ali}–H class surrounding backbone peptide bonds. This suggests that the weak hydrogen bonds could contribute significantly to the stability of the Sm/LSm proteins. Similarly, the population of interchain hydrogen bonds from the backbone and the side-chain donor is 22.8% and 77.2% respectively, while for hydrogen bonds from the backbone and the side-chain acceptor is 48.6% and 51.4% respectively. In the case of interchain hydrogen bonds, the backbone groups are the less frequently involved, because their atoms are not as accessible as the side-chain atoms and also because the backbone groups are involved in CH···O interactions to a substantial extent. The intrachain CH···O interactions are the most frequently involved (50.4%), followed by NH···O interactions (23.3%), CH···N interactions (14.8%), and NH···O interactions (8.7%). The higher percentage of CH···O interactions may be explained in terms of the larger abundance of CH groups and therefore, many investigations of CH···O interactions focus on the CH groups as donors [38]. The large number of the strong hydrogen bonds is due to the fact that protein interiors consist mostly of residues which form well defined α-helices and β-sheets. Among interchain hydrogen bonds, we found that 49.0% of the interactions were CH···O interactions, 23.4% of the interactions were NH···O interactions, 12.4% of the interactions were CH···N interactions, and 3.5%

interactions were NH···N interactions. The contribution from interchain hydrogen bonds with π-acceptors was predominant in CH···π interactions (6.2%). The small percentage of NH···π and OH···π interactions is probably a consequence of the tendency of NH and OH groups to be involved in classic hydrogen bonds. We observed a very small percentage of weak hydrogen bonds involving sulphur atoms. Since oxygen is more electronegative than nitrogen, there are substantially less OH···π and OH···S interactions. This is in agreement with the data for XH···π interactions in the proteins, where a small number of the OH···π interactions were found [39].

An example of hydrogen bonds between subunits (C and V) from *Archaeoglobus fulgidus* Sm core domain is shown in Figure 2. There are two strong NH···O and seven weak CH···O hydrogen bonds in that interface.

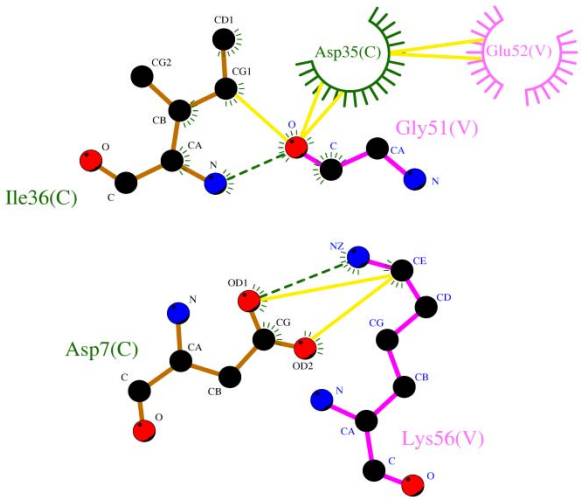

Fig. 2. View of the interface between subunits (C and V) from Archaeoglobus fulgidus Sm core domain (PDB ID code 1i4k). The flattened diagram places atoms and bonds on the 2D page to minimize the overlap of atoms and the crossing of bonds in the final diagram. Strong hydrogen bonds (NH···O) are indicated by dashed green lines between the atoms involved. Corresponding atoms involved in weak CH···O hydrogen bonds are represented by yellow lines between the atoms involved. There are 33 amino acids in the interface, and two of them are involved in NH···O hydrogen bonds (CIle36(N)···VGly51(O); VLys56(NZ)··· CAsp7(OD1)). There are seven weak CH···O hydrogen bonds in that interface CAsp35(CA)···VGly51(O); CAsp35(CA)···VGlu52(OE1); CAsp35(CA)···VGlu52(OE2); CAsp35(CB)···VGly51(O); CIle36(CG1)···VGly51(O); VLys56(CE)···CAsp7(OD1); VLys56 (CE)···CAsp7(OD2). This figure was prepared using program LigPlot+ v.1.0.5 [40].

3.2. Hydrogen bond geometry: lengths and angles

The interaction geometry of the most abundant type of hydrogen bonds (intrachain CHO) in the total Sm/LSm proteins in our dataset is shown in Fig. 3.

The CH···O interactions include {CHO BD BA}, {CHO SD SA}, {CHO BD SA} and {CHO SD BA}. For {CHO BD BA} the angle distribution has two distinct maxima at 105° and 140° with a narrow range of linearity. The metrics of the other CH···O interactions are surprisingly consistent. Also, it is similar for {CHO BD SA; CHO SD BA}, the maxima are still around 115° and 140° with variable geometry. In both cases, the lower angle maxima distribution corresponds generally to multifurcated geometries (The lower area in Figures 3a and 3g) [5]. For {CHO SD SA} the maxima occur at 130° and 145°,whereas at interfaces maxima occur at 125°,140° and 160°-165°[31].

The median CH···O distances, *d*, in all the above cases are <3.0 Å. For {CHO BD BA}, *d* is 2.4 Å. For other CH···O interaction types {CHO SD SA} and {CHO BD SA; CHO SD BA}, the

median distances, are 2.8 Å and 2.9 Å, respectively. Similar median CH···O distances have been found when we considered interface non-canonical interactions [31].

The inverse length–angle correlations are also well behaved in all these cases. To summarize, the main-chain CH···O interactions {CHO BD BA}, might be slightly more linear than the side-chain interactions, and they have somewhat shorter median distances (similar as in interfaces, [31]).

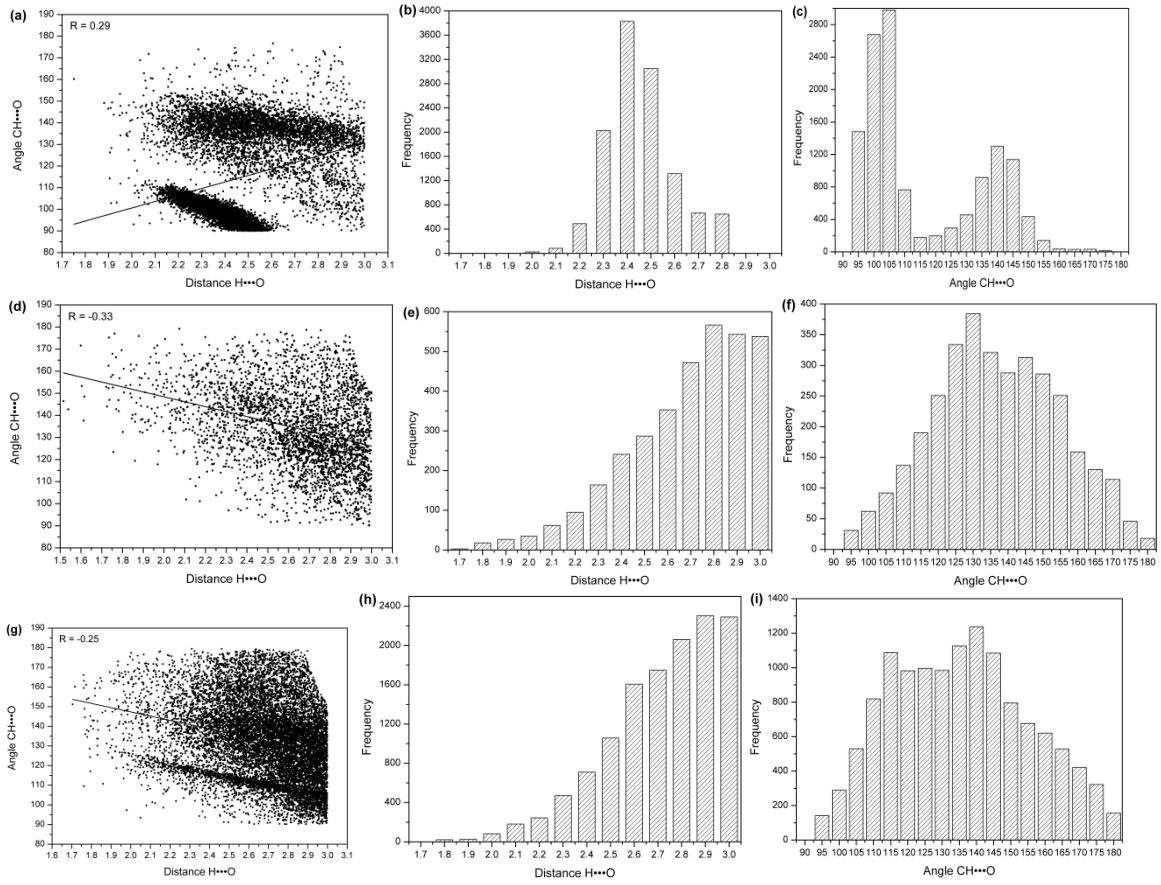

Fig. 3. Hydrogen bond geometry for the intrachain {CHO BD BA} (a-c), {CHO SD SA} (d-f) and {CHO BD SA; CHO SD BA} (g–i). In each case the inverse length-angle scatterplot is followed by histograms of distances and angular distributions.

The geometries for other hydrogen bonds observed in the Sm/LSm proteins (data not shown) are consistent and fall within acceptable limits. For strong hydrogen bonds, the median distances, *d*, are less than 2.4 Å. The angular distributions for strong hydrogen bonds are similar with maxima in the range of 170–175°. Strong hydrogen bonds show better linearity and shorter distances compared with weak hydrogen bonds. The weak hydrogen bonds have variable geometry.

These observations are in agreement with the fundamental property of hydrogen bonds, namely linearity and holds by and large for all categories in macromolecular structures [5,9].

3.3. Donor and acceptor role of amino acids in intra and interchain hydrogen bonds

The percentage contribution of each of the amino acid residues as donor and acceptor for each type of interaction in intra and interchain area was calculated as the ratio of the occurrence of specific amino acid involved in the particular type of interaction and they are tabulated in Table 2 and 3, respectively.

Table 2 Percentage contribution of different amino acids in a particular type of hydrogen bonds in intrachains.

	NH···O		**OH···O**		**NH···N**		**OH···N**		**CH···O**		**CH···N**		**NH···S**		**OH···S**		**CH···S**		**NH···π**		**OH···π**		**CH···π**	
	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc
Nonpolar																								
Gly	5,9	5		3,9	8,2	11,7		0,9	3	6,2	0,4	5,1	18,7											
Ala	4,9	5,4		2,1	7,3	5,6		6,3	3,5	4,3	4,5	4,6											1,4	
Val	10	9,8		13	7	5		1,8	17,2	11,4	23,6	10,4					14,3						17,1	
Leu	11,5	9,8		5	9,8	9,6		4,5	15,1	9,8	12	10					27,4						23,4	
Ile	6,2	5,5			3,6	4			9,4	5,9	13,5	6,8					16,1						5,7	
Met	2	0,9		0,9	1,6	1,6			1,6	1,6	1,1	0,5		37			13,1	70					5,3	
Pro		1,9		3,9		2,2		5,4	1,7	1,7	2	1,4					1,2						0,2	
Phe	3,3	3,6		1,6	3,9	3			4,4	3,2	4,8	4,4					9,5						9,2	63
Trp																								
Polar																								
Ser	4,6	5,2	43,4	1,8	5,3	3,8	45,1		3,2	6	1,6	4,1	6,2				0,6				6,7		1,6	
Cys	0,4	0,5		0,2	0,3	0,3			0,4	0,3	0,1	0,07		63		100		30						
Thr	2,3	2,4	28,1	5,5	2,7	3,1	4,5	4,5	2,8	3,3	3	1,7					1,8						2,9	
Asn	9,4	7,6		10,3	11,6	9,8		4,5	4,7	7,9	3,6	6,7	44		100									
Gln	3	2,6		4,1	3	3,7		0,9	2,2	3,1	2,2	4	6,2				6,6						7,9	
Tyr	1,5	2,4	28,5	1,1	1,3	1,6	49	9	3,6	3,7	3,4	2,1					0,6			75	93,3		2,5	18,4
Charged																								
Lys	8,7	4,4		3,4	8	7,7		1,8	7,7	5	4,8	6,4	12,5				5,4						5,5	
Arg	12	6,7		4,3	8,9	6,8		21	7,7	6,2	6	15,8	6,2				0,6		75					
His	3,3	1,1		0,2	4	4,6		38	2,1	2,1	2,5	4,2					3		25	25		100	1,7	19
Asp	3,7	12,1		12,8	4,8	8,5		0,9	2,7	7,3	3,5	4,9	6										0,7	
Glu	7,2	13		25,8	8,7	7,6	1,8	0,9	6,8	11	7,2	6,7											15	

Don, donor; Acc, acceptor

Table 3. Percentage contribution of different amino acids in a particular type of hydrogen bonds in interchains

	NH···O		**OH···O**		**NH···N**		**OH···N**		**CH···O**		**CH···N**		**NH···S**		**OH···S**		**CH···S**		**NH···π**		**OH···π**		**CH···π**	
	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc	Don	Acc
Nonpolar																								
Gly	0,2	2,4			0,4	0,8			2,1	2,1	2,8	0,6											0,6	
Ala	0,6	0,2		1,3					3,4	0,5	2,6												1,7	
Val	8,8	17		0,9	6,1	0,4			14	13	9,5	0,3											20	
Leu	2,6	5,8		1,7					5,9	5,5	14	0,2					3,6						22	
Ile	6,6	11			0,4	0,4			11	12	12	1,4					3,6						0,2	
Met	0,7	0,7		1,3					0,8	0,3	4			100				100					5,5	
Pro		0,8						9,1	11	1,5	7,6	1,3												
Phe	2,7	2,3				11			8,8	3,2	6,3	2,9								87		4,7	18	60
Trp																								
Polar																								
Ser	1,6	5,2	48	4,3			45		3,7	12											35		4,2	
Cys						0,4			0,1		0,4													
Thr	0,1	1,6	14	11			32		2,9	4,2	3,7	0,1									12		3	
Asn	9,2	7,1		2,6	19	18		14	2,5	10	1,2	26											5,5	
Gln	2,8	1,6		18	0,8	1,1		4,6	1,3	3,6	0,1	4,1					18		17				7,8	
Tyr	4	2,2	38	17			23		8,5	7,4	8,3	4,7					2,4				54	49	4	24
Charged																								
Lys	9,6	2,3		12	17	14		4,6	6,6	3,1	7,6	4,9							6				0,8	
Arg	45	2		16	56	51		18	9,8	2,6	8,5	44	100				72		77				1,7	
His	5,4	1,2			0,4	1,1		45	4,4	0		6,7								13			2,5	15,6
Asp		11		0,9		0,4			1,5	6,9	9,4	2,8											1,3	
Glu	0,4	26		13		1,9		4,6	1,9	12	1,8												1,9	

It could be inferred from the Table 2 that only in the NH···O, NH···N, CH···O, and CH···N interactions, most of the amino acid residues serve as donor and acceptor in strong and weak hydrogen bonds. In CH···S and CH···π there is greater diversity of amino acids among donors and only two and three (respectively) amino acids are acceptors. Among NH···π interaction donors, there are two amino acids Arg and His. Acceptors in NH···π interactions are His and Tyr. In case of OH···π, interactions acceptor is His, while donors are Ser and Tyr.

The percentage contribution of amino acids as donor and acceptor for each type of interactions between different chains (interchain) is somewhat different (Table 3).

In OH···O type of interactions, some amino acids which play a role of acceptors (mainly non-polar) are excluded from interchains, like Cys, Gly, His, Ile Phe, Pro. In NH···N interactions the predominant donor and acceptor is Arg in interchains, whereas percentage distribution is more uniformed in case of intrachains. This is in agreement with observation [30] that Arg is more represented in interfaces of the Sm proteins than in the interior of the monomers, and reason for this is the involvement of Arg in strong and weak hydrogen bonds in interfaces and salt bridges.

Regarding OH···N interactions, some amino acids are excluded as acceptors like Ala, Asp, Gly, Ile, Leu, probably because they prefer interior of monomers [30]. In building of CH···O interactions in interchains, most of amino acids are represented as donors and acceptors, like in intrachains with somewhat different distribution. For example, Ala is an 8 folds more present acceptor in intrachains. Similar observation holds for CH···N interactions where Arg is the predominant acceptor in interchain and 2.8 folds is more included in this type of interactions.

. Predominant donor for CH···S interactions in interchains is Arg which is not the case in intrachains where more different amino acids are involved.

In CH···π interactions most of amino acids which are donors in interchains are donors in intrachains, with some exceptions like depletion of Arg, Asn and Pro. Amino acids which are donors and acceptors in NH···π intrachain interactions differ from amino acids in interchains, with some amino acids in common like Arg and His. In case of OH···π interactions, Ser is present as acceptor in interchain and not in intrachains. Concerning acceptors, His is the only acceptor in intrachain OH···π interactions, while in the interchain other amino acids like Phe and Tyr play that role.

In several type of interactions (NH···O, NH···N, CH···N, NH···S, CH···S) Arg is significantly more present in interchains, despite its charge. Arg contributes about 10% of the accessible surface area and the surface buried at interfaces [41].Another significant contributor to the interfaces is Lys and percentage values of donors or acceptors for some interactions (NH···..N, OH···N, CH···O, and CH···N) are higher in interchain hydrogen bonds, but differences are not so pronanunced as in the case of Arg.

Despite the fact that Leu is abundant at interfaces [41] our calculations (Table 2 and 3) show that this amino acid is more occupied by building of intrachain non canonical interactions. Similar observation holds for Glu, Arg and Lys. Glu is more represented at the interfaces of Sm proteins [30] and involvement of the first two amino acids in hydrogen bonds results from their presence in interchain area. Glutamic acid is probably more involved in salt bridges formation which are more represented in interfaces of Sm proteins than in test set (Binding Interface Database) [30]. Ala, Leu, Gly, Ile, Met, and Cys are with higher percentage involved in building of intrachain hydrogen bonds, which is corroborated further by our published data [30] that these amino acids prefer to be in the interior of monomers.

4. Solvent accessibility and conservation score of residues involved in hydrogen bonds

In this study, we have estimated the solvent accessibility of all residues that are involved in various types of hydrogen bonds with the aid of DSSP [42]. The relation between the amino acid residues in these interactions and solvent accessibility is illustrated in Table 4. The solvent accessibility of amino acid residues has been categorized as buried (0–20%), partially buried (20–50%), and exposed (>50%).

Table 4.Solvent accessibility preference for the amino acid involved in hydrogen bonds.

	NH···O		OH···O		NH···N		OH···N		CH···O		CH···N		NH···S		OH···S		CH···S		NH···π		OH··π		CH··π	
	*	#	*	#	*	#	*	#	*	#	*	#	*	#	*	#	*	#	*	#	*	#	*	#
Nonpolar																								
Gly	B	P	B		B	P	B		B	B	B	B	B											B
Ala	B	B	B	B	B		B		B	B	B	B											B	B
Val	B	B	B	B	B	B	B		B	B	B	B					B						B	B
Leu	B	B	B	B	B		B		B	B	B	B					B	B					B	B
Ile	B	B			B	B			B	B	B	B					B	B					B	B
Met	B	B	B	P	B				B	B	B	B	B	P			P	P					B	B
Pro	B	P	B		B		B	P	B	B	B	B					B						B	
Phe	B	B	B		B	B			B	B	B	B					P			B		B	B	B
Trp																								
Polar																								
Ser	P	E	P	E	P		P	E	P	P	B						P				P	E	P	E
Cys	B	P	B		B	P			B	P	B	P	P		P		P							
Thr	P	E	P	E	P		P	E	B	P	B	P					B					E	E	E
Asn	E	E	E	E	E	E	E	E	P	P	P	E	E		E									E
Gln	P	E	P	E	P	E	P	E	P	P	P	P	P				P	P		E			P	P
Tyr	P	P	P	P	P		P	P	B	P	P	P					P	P	P		P	P	P	P
Charged																								
Lys	E	E	E	E	E	E	E	E	E	E	E	E	E				E			E			E	E
Arg	E	E	E	E	E	E	E	E	E	E	E	E	E	E			E	E	E	E				E
His	P	E	P		P	E	P	E	P	P	P	P					P		P	P	P	P	P	P
Asp	E	E	E	E	E	E	E		P	P	E	E	E										E	E
Glu	E	E	E	E	E	E	E	E	E	E	E	E											E	E

B, buried (0–20% ASA); P, partially buried (20–50% ASA); E, exposed (>50% ASA);
* Asteriks indicates intrachain interactions; # Number sign indicates interchain interactions.
Blank space shows that the particular amino acid does not participate in the specific interaction.

Most of the other amino acid residues that were involved in hydrogen bonds prefer to be in the solvent excluded environment, especially when the interaction involves main-chain (intrachain) atoms. The data indicate that the most charged amino acid residues prefer to be solvent exposed when they are involved in hydrogen bonds. We found that, of the different amino acids that were involved in strong hydrogen bonds; Lys, Arg, Asp and Glu were in the exposed regions, irrespective whether they are involved in the intra or interchain hydrogen bonds. Polar amino acids, as well as His, were in partially buried regions, when they are involved in the intrachain hydrogen bonds. Polar interchain residues such as: Ser, Cys, Thr and Gln preferred to be in the exposed region. The general trend in case of polar amino acids is crossing of particular amino acid to the position more exposed to the solvent, in the same type of interaction. This observation is quite reasonable in the sense that most of the interchain residues tend to be exposed or partially buried. The nonpolar amino acid residues were in the buried regions no matter whether they are in the intra or interchain hydrogen bonds. Although, the solvent accessibility patterns for both CH···O and CH···N interacting residues were almost similar, it was interesting to find that Asn and Asp residues that were involved in CH···π interactions are more exposed. Met and Cys residues that were involved in sulphur hydrogen bonds were in partially buried regions.

We found that amino acids which were involved in CH···π interactions, Ser, Asp, Glu, Lys and Arg were in the exposed regions, His, and Tyr were in partially buried regions, and nonpolar residues were in the buried regions. According to [43] CH···π interactions involving aromatic residues either as donor or acceptor groups are found mostly in the interior of the protein and tend to be buried in nature. These might be one of the reasons for their solvent accessibility nature. Furthermore, we found that most of the polar amino acid residues involved in NH···π and OH···π

interactions was solvent exposed and most of the nonpolar residues involved in NH···π and OH···π interactions were excluded from the solvent.

It is considered that structurally conserved residues are important in protein stability and folding [44]. We found that most of the amino acids making hydrogen bonds are highly conserved: most of them had a conservation score of 9, the highest number on the scale. The calculated average conservation score is 6.8 ± 1.8 (mean ± standard deviation). These data indicate to a similar importance of all hydrogen bonds in Sm/LSm proteins.

5. Secondary structure preferences for amino acids building strong and weak hydrogen bonds

The occurrence of these weak interactions has been observed at the terminus of the secondary structural units in particular α-helix and β-sheet [16]. These interactions have been proposed to have a definitive role in stabilizing these secondary structural scaffolds of proteins. The propensity of the amino acid residues to favour a particular conformation is well described. Such conformational preference is not only dependent on amino acid alone but as well on the local amino acid sequence [14]. We have analyzed secondary structure preference for each amino acid that participates in different types of hydrogen bonds, separately for intrachain and for interchain interactions (Table 5 and 6).

Table 5.Secondary structure preferences for amino acids involved in intrachain interactions

	NH···O		**OH···O**		**NH···N**		**OH···N**		**CH···O**		**CH···N**		**NH···S**		**OH···S**		**CH···S**		**NH···π**		**OH···π**		**CH···π**	
	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A
Nonpolar																								
Gly	CS	S		S	C	CT		S	CH	S	CS	HS	T											
Ala	HS	H		H	H	H		T	HS	HS	S	H											H	
Val	S	S		H	HS	H		S	S	S	S	S					S						S	
Leu	HS	S		H	HT	H		S	S	S	S	S					S						HS	
Ile	S	S			CH	HT			S	S	S	S					S						S	
Met	C	S		S	C	C			S	S	S	S					CS	S					S	
Pro		H		HS		H		S	ST	HS	H	CH					C						CS	
Phe		H		HS		H		S	ST	HS	H	CH					C						S	CS
Trp																								
Polar																								
Ser	ST	CS	CT	T	ST	T	S		S	S	CS	S	C				C				S		C	
Cys	S	S		S	CH	H			S	S	S	S		H		H		S						
Thr	CS	S	CS	S	HT	CH	C	S	S	CS	S	S					S						S	
Asn	CH	C		T	CH	HT		CH	CT	CH	C	CT	H		H									
Gln	HS	HS		S	HT	TH		S	CS	ST	CS	CS	C				H						T	
Tyr	S	S	S	S	H	H	S	S	S	S	S	S					C		C				S	S
Charged																								
Lys	HS	HS		S	HS	HT		S	SH	HS	CHS	ST	S	C			S						S	
Arg	CS	CS		C	CH	CH		HS	CS	CS	CS	CS	S				S		C					
His	HST	H		S	T	HT		H	CHT	TS	TH	ST					H		C	C		H	H	ST
Asp	CS	C		S	H	HT		T	CS	CS	CH	C	S										S	
Glu	HS	HS		C	CH	HT	S	C	S	S	CS	S											S	

D, donor; A, acceptor; H, helix; C, coli; S, strand; T, turn; Blank space shows that the particular amino acid will not participate in that interaction.

Members of the Sm protein family are characterized by the conserved bipartite Sm fold composed of Sm motif 1 and Sm motif 2. Solved structures of this family members, do show that the fold is highly conserved and this is defined by an N-terminal helix followed by a five stranded anti-parallel β sheet. Strands β1-β3 are a part of Sm motif 1 and strands β4 and β5 are parts of Sm motif 2. The five stranded β sheet is strongly bent in the middle and the conserved hydrophobic residues form hydrophobic core. Analysis of the percentage of the secondary structural units in Sm proteins included in this study indicates that the percentage of helices, beta strands, coils, and turns are 15%, 54%, 31% and 26% respectively.

Table 6.Secondary structure preferences for amino acids involved in interchain interactions.

	NH···O		OH··O		NH···N		OH···N		CH···O		CH···N		NH···S		OH···S		CH···S		NH···π		OH···π		CH···π	
	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A
Nonpolar																								
Gly	C	CH			C	C			S	HT	C	C											C	
Ala		CS		T					S	CS	HS												C	
Val	S	S		S	H	C			S	S	S	S											C	
Leu	S	S		S					ST	S	S	S				S							H	
Ile	S	S			S	C			S	S	S	S				S							S	
Met	S	S		S					CS	S	C			C									S	
Pro		C						T	HS	C	H	H												
Phe	S	S				C			S	S	S	S								C		C	S	CS
Trp																								
Polar																								
Ser	C	S	S	S			S		S	S											T		T	
Cys						H			H		H													
Thr	C	ST	T	S			S		S	S	S	C									S		S	
Asn	CH	CT		C	T	CT		C	C	CH	HT	H											C	
Gln	ST	T		HST	S	S		T	S	ST	S	C				H			S				T	
Tyr	S	S	S	H			S		S	S	H	S				H					S	H	S	H
Charged																								
Lys	ST	H		S	ST	ST		T	HT	S	CS	ST							C				CH	
Arg	C	S		H	S	S		C	C	S	CH	CS	C			C			S				C	
His	HS	T			C	T		S	S	T		S			T					CT		HS	H	S
Asp		CH		C		T			S	CS	H	C											S	
Glu	S	S		S		C		S	T	CT	CH												H	

D, Don; A, Acceptor; H, helix; C, coli; S, strand; T, turn; Blank space shows that the particular amino acid will not participate in that interaction.

In the whole data set we did not find any exclusive preference for particular secondary structure. The majority of intrachain and interchain hydrogen bonds prefer to occur in strand, irrespective of the amino acid propensity to adopt a particular secondary structure. This is probably

due to the fact that beta strands are more represented than other types of secondary structures, and that interaction interfaces between two monomers are enabled using beta strands (β4 and β5).

Except for the intrachain NH···π B-S, and NH···π S-S the remaining sub types of interactions were found to be not significantly selective to any particular secondary structure. In general, strands are the most represented in different types and sub types and turns are the least involved. In case of interchain interactions, only NH···S, S-S type of interaction shows preference to coil secondary structure elements, although Met which is acceptor in this type of interaction shows preference toward beta strand.

In general, strands are the most involved in building of interchain strong and weak hydrogen bonds followed by coils, which is in accordance with the percentage share of various secondary structures in Sm proteins, and fact [21,26,27,29] that interaction interface between monomers is via β4 and β5 strands. In interchain area of analyzed Sm proteins, OH···S interactions are not found, whereas in interior of monomers they are represented, although not with high share when compared to other hydrogen bonds.

In order to draw correlation between the occurrences of a particular hydrogen bond to an amino acid adopting a particular secondary structural fold, we have analyzed the percentage occurrence of the interactions in a particular secondary structure, irrespective of the amino acid, and a result is depicted in Figure 4 (right and left panel) for intrachain and interchain interactions.

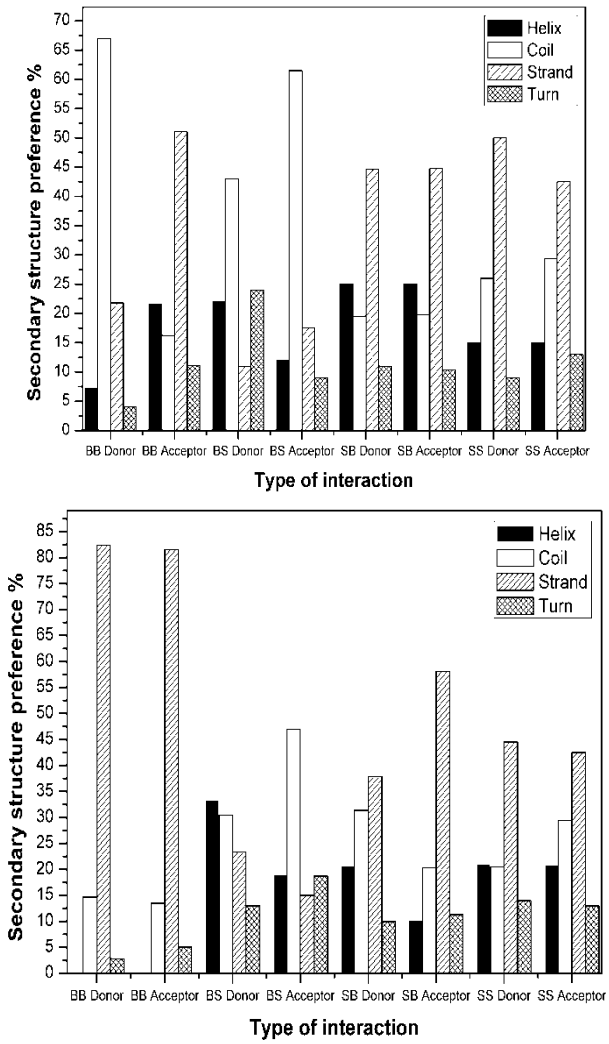

Fig. 4. Percentage of residues in the different secondary structural units that participate in the various types of hydrogen bonds in intrachains (left panel),and in interchains (right panel)

The Donor amino acid residues involved in backbone to backbone intrachain interactions prefer to be in coil conformation. Acceptor amino acids from intrachain backbone to backbone interactions prefer to adopt beta strand conformation. However, both Donors and Acceptors from backbone to side-chain intrachain interactions are mainly in coils. Donors and Acceptors from the side-chain to backbone and side-chain to side-chain intrachain interactions are predominantly from beta strands.

In case of interchain backbone to backbone interactions, both Donors and Acceptors are from strands with high share (~80%). As opposed, to intrachain interactions where a small percentage of BB Donors and BB Acceptors belong to alpha helix, BB Donors and acceptors from interchain interactions are not from alpha helix. BS Donors from interchains are more or less similarly distributed in the three secondary structures and only a turn is represented with lower percentage. BS interchain acceptors are similarly distributed in three secondary structures and coil is represented with higher percentage. Donors and Acceptors from side chain to backbone interactions and side chain to side chain interchain interactions are predominantly from beta strands.

When we compare types of interaction with secondary structure preferences in intrachains and interchains, we can notice that there is similar trend that in both Donors and Acceptors the most represented type of secondary structure is beta strand, which is not surprising because strands represent the majority of secondary structures of Sm/LSm proteins.

4. Conclusions

This study compares strong and weak hydrogen bonds in the interior of monomers and at interfaces of Sm/LSm proteins in order to better understand the stability of oligomeric assemblies. All strong and weak hydrogen bonds and their properties were analyzed separately for the interior of monomers and at interfaces. Similar percentages of CH···O, NH···O and CH···N interactions were found in the interior of monomers and at interfaces. Differences between interior of monomers and interfaces are pronounced in the case of NH··· N, OH···O,XH···π interactions and interactions involving sulphur atoms which are more represented at the interfaces. Although they do not represent predominant type of hydrogen bonds between chains, this may suggest that they contribute to a certain extent to the stability of oligomeric associations. In the case of interchain hydrogen bonds, the backbone groups are less frequently involved. Characteristics of hydrogen bonds in respect of geometry for interchain interactions were previously reported. In this study we found that intrachain hydrogen bonds behave similarly in respect of geometry, distances and angles.We found that Arg is involved in building of NH···N interactions in interchains with high share, whereas in intrachains it is less frequently involved in this type of interaction. Another significant contributor to the interfaces is Lys, acting as Donor or Acceptor for some interactions (NH···N, OH···N, CH···O, and CH···N). These findings are in agreement with our previous analysis where Arg was more abundant at interfaces, and higher occurrence of NH···N hydrogen bonds at interfaces. Solvent accessibility pattern of amino acids involved in the hydrogen bonds analysis indicates that the majority of the amino acid residues prefer to involve in hydrogen bonds only when they are excluded from the solvent. Most of charged amino acids are solvent exposed irrespective whether they are at interfaces or at intrachains. Most of nonpolar amino acids are buried in both cases. In general, there is a trend for most polar amino acids to cross into more solvent exposed positions in interfaces. Based on the analysis we were not able to assign exclusive preference for particular secondary structure both for intra and interchain interactions. The majority of intra- and interchain hydrogen bonds prefer to occur in strand, irrespective of the amino acid propensity to adopt a particular secondary structure. This is probably due to the fact that beta strands are more represented than other types of secondary structures and that interaction interfaces between two monomers are established using beta strands (β4 and β5).Our analysis suggests that beside the most represented types of interactions contributing to the stability of interfaces, exist smaller percentages of other interactions which play an additional role in the stability of interfaces. It should be noted that Arg and Lys play their role in supporting the stability of interfaces, in some cases to a higher extent than to the stability of monomers. The high

conservation score of amino acids that are involved in hydrogen bonds is an additional strong argument for their importance in the stability of both monomers and oligomeric association.

Acknowledgements

This work was supported by the grant No. 172001 from the Ministry of Science and Education, Republic of Serbia.

References

[1] Jeffrey, G.A., Saenger, W., Hydrogen Bonding in Biological Structures, Springer-Verlag, Berlin (1991).
[2] Sarkhel, S., Desiraju, G.R.. Proteins **54** 247-259.(2004).
[3] Bartlett,G.J., C.T.Porter, N.Borkakoti, and J.M.Thornton. J. Mol. Biol. **324**:105-121. (2002).
[4] Steiner,T. Angew. Chem. Int. Ed **41**:48-76. (2002).
[5] Panigrahi, S.K., Desiraju, G.R., Proteins **67**, 128-141. (2007).
[6] Tiwari, A., Panigrahi, S.K.In Silico. Biol. **7**, 651-661.(2007).
[7] Smith, D.A. Am. Chem. Soc. Symp. Ser. **569**, 82–219.(1994)
[8] Desiraju, G.R., Steiner, T., The weak hydrogen bond in structural chemistry and Biology, Oxford University Press, Oxford.(1999)
[9] Panigrahi, S.K., Amino. Acids **34**, 617-633 (2008)
[10] McPhail, A.T., Sim, G.A., Chem. Commun. 124-125.(1965).
[11] Armstrong, K.M., Fairman, R., Baldwin, R.L., J. Mol. Biol. **230**, 284-291, (1993).
[12] Parkinson, G., Gunasekera, A., Vojtechovsky, J., Zhang, X., Kunkel, T.A., Berman, H., Ebright, R.H., Nat. Struct. Biol. **3**, 837-841.(1996).
[13] Hu,J., L.J.Barbour, and G.W.Gokel Proc. Natl. Acad. Sci. USA **99**:5121-5126.(2002).
[14] Chakkaravarthi, S., Babu, M.M., Gromiha, M.M., Jayaraman, G., Sethumadhavan, R.,. Proteins **65**, 75-86(2006).
[15] Senes, A., Ubarretxena-Belandia, I., Engelman, D.M., Proc. Natl. Acad. Sci. USA **98**, 9056-9061(2001).
[16] Babu, M.M., Kumar, S.S., Balaram, P., J. Mol. Biol. **322**, 871-880(2002).
[17] Anantharaman, V., Koonin, E.V., Aravind, L.,. Nucleic Acids Res. **30**, 1427-1464. (2002).
[18] Anantharaman, V., Aravind, L., BMC. Genomics **5**, 45 (2004).
[19] Collins, B.M., Harrop, S.J., Kornfeld, G.D., Dawes, I.W., Curmi, P.M., Mabbutt, B.C., J. Mol. Biol. **309**, 915-923.(2001).
[20] Mura, C., Cascio, D., Sawaya, M.R., Eisenberg, D.S. Proc. Natl. Acad. Sci. USA **98**, 5532-5537.(2001)
[21] Collins, B.M., Cubeddu, L., Naidoo, N., Harrop, S.J., Kornfeld, G.D., Dawes, I.W., Curmi, P.M., Mabbutt, B.C., J. Biol. Chem. **278**, 17291-17298.(2003)
[22] Sauter, C., Basquin, J., Suck, D. Nucleic Acids Res. **31**, 4091-4098.(2003).
[23] Thore, S., Mayer, C., Sauter, C., Weeks, S., Suck, DJ. Biol. Chem. **278**, 1239-1247(2003).
[24] Zaric, B., Chami, M., Remigy, H., Engel, A., Ballmer-Hofer, K., Winkler, F.K., Kambach, C. J. Biol. Chem. **280**, 16066-16075 (2005).
[25] Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dg, A., Wasney, G.A., Gao, M., Hills, T., Brokx, S., Qiu, W., Sharma, S., Diassiti, A., Alam, Z., Melone, M., Mulichak, A., Wernimont, A., Bray, J., Loppnau, P., Plotnikova, O., Newberry, K., Sundararajan, E., Houston, S., Walker, J., Tempel, W., Bochkarev, A., Kozieradzki, I., Edwards, A., Arrowsmith, C., Roos, D., Kain, K., Hui, R., Mol. Biochem. Parasitol. **151**, 100-110 (2007).
[26] Naidoo, N., Harrop, S.J., Sobti, M., Haynes, P.A., Szymczyna, B.R., Williamson, J.R., Curmi, P.M., Mabbutt, B.C., J. Mol. Biol. **377**, 1357-1371(2008).
[27] Kambach, C., Walke, S., Young, R., Avis, J.M., de la Fortelle, E., Raker, V.A., Luhrmann, R., Li, J., Nagai, K., Cell **96**, 375-387.(1999).

[28] Zaric, B.L. Reconstitution of two human LSm protein complexes reveals aspects of their architecture, assembly and function. Diss.ETHno. 15977, Swiss Federal Institute of Technology, Zurich, (2005).
[29] Fischer, S., Benz, J., Spath, B., Maier, L.K., Straub, J., Granzow, M., Raabe, M., Urlaub, H., Hoffmann, J., Brutschy, B., Allers, T., Soppa, J., Marchfelder, A. J. Biol. Chem. **285**, 34429-34438.(2010).
[30] Zarić,B.L., V.B.Jovanović, and S.Đ.Stojanović. J. Theor. Biol. **271**,18-26.(2011).
[31] Stojanović, S. Đ., Isenovic.ER, Zarić, BL.Molecular Inofrmatics, **5**,430-442 (2011).
[32] Stojanović, S. Đ., Zarić, B.L., Zarić, S.D.J. Mol. Model. **16**, 1743-1751 (2010).
[33] Word, J.M., Lovell, S.C., Richardson, J.S., Richardson, D.C. J. Mol. Biol. **285**, 1735-1747 (1999).
[34] Krissinel,E., andK.Henrick. J. Mol. Biol. **372**,774-797.(2007).
[35] Kabsch, W., Sander, C.,Biopolymers **22,** 2577-2637.(1983).
[36] Gilis, D., Rooman, M., 1997. J. Mol. Biol. **272**, 276-290.(1997).
[37] Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., Ben-Tal, N., Nucleic Acids Res. **33**, W299-W302.(2005)
[38] Novoa,J.J., and F.Mota. Chemical Physics Letters **266**,23-30. (1997).
[39] Steiner, T., Koellner, G.J. Mol. Biol. **305**, 535-557(2001).
[40] Wallace,A.C., R.A.Laskowski, and J.M.Thornton. Protein Eng. **8**,127-134.(1995).
[41] Lo,C.L., C.Chothia, and J.Janin. J. Mol. Biol. **285,**,2177-2198(1999).
[42] Kumarevel, T.S., Gromiha, M.M., Selvaraj, S., Gayatri, K., Kumar, P.K., Biophys. Chem. **99**, 189-198 (2002).
[43] Brandl,M., M.S.Weiss, A.Jabs, J.Suhnel, and R.Hilgenfeld. J. Mol. Biol. **307**,357-377. (2001).
[44] DeLano, W.L., Curr. Opin. Struct. Biol. **1**2, 14-20.(2002)