Journal of Molecular Graphics and Modelling 28 (2009) 88–94

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Structural features for homodimer folding mechanism

Sajitha Lulu [a], Abishek Suresh [b], Velmurugan Karthikraja [b], Mohanapriya Arumugam [a], Rajarathinam Kayathri [b], Pandjassarame Kangueane [a,b,*]

[a] School of Biotechnology, Chemical and Biomedical Engineering, VIT University, Vellore 632014, Tamil Nadu, India
[b] Biomedical Informatics, 17A Irulan Sandy Annex, Pondicherry 607402, India

ARTICLE INFO

Article history:
Received 27 December 2008
Received in revised form 30 March 2009
Accepted 11 April 2009
Available online 19 April 2009

Keywords:
Homodimer
Folding
Binding
Mechanism
Structure

ABSTRACT

The homodimers have essential role in catalysis and regulation. The homodimer folding mechanism through 2-state without stable intermediate (2S), 3-state with monomer intermediate (3SMI) and 3-state with dimer intermediate (3SDI) is fascinating. 23MI and 3SDI constitute 3-state (3S). Hence, it is important to differentiate 2S, 3SMI and 3SDI homodimers using structural features. We used the dataset of Li et al. [L. Li, K. Gunasekaran, J.G. Gan, C. Zhanhua, P. Shapshak, M.K. Sakharkar, P. Kangueane, Structural features differentiate the mechanisms between 2S and 3S folding of homodimers, Bioinformation 1 (2005) 42–49] consisting of twenty-five 2S, ten 3SMI and six 3SDI homodimer structures for the study.

Interface to total (I/T) residues ratio is large for 2S than 3SMI and 3SDI. Interface to total residues ratio is similar for 3SMI (mean monomer length (ML) = 208) and 3SDI (mean monomer length (ML) = 404) despite difference in mean monomer size. Interface residues correlate with monomer size in 2S (Pearson's correlation coefficient (r); $r^2 = 0.41$) and 3SMI ($r^2 = 0.52$). This is not true for 3SDI with interface residues and monomer length ($r^2 = 0.17$). Interface area (B/2) does not correlate with interface residues ($r^2 < 0.001$) and monomer size ($r^2 = 0.023$) in 2S. This is despite a relationship with interface residues and monomer size ($r^2 = 0.41$) in 2S. However, this is not true for 3SMI ($r^2 = 0.61$ with interface residues and $r^2 = 0.25$ with monomer size). In 3SDI, a different relationship is seen ($r^2 = 0.28$ with interface residues and $r^2 = 0.09$ with size).

The mean hydrophobicity factor (H_f) is 3-fold less in 3S than 2S. H_f does not correlate with interface area in 2S ($r^2 = 0.03$) and 3SDI ($r^2 = 0.0$). However, a weak causal relation is seen in 3SMI ($r^2 = 0.23$). Hydrophilic amino acid residues (E, R, K, S and Q) are prominent in 2S than 3S. Charged negative amino acid residues (D, E) are more than positive amino acid residues (R, K, H) in 2S and charged positive amino acid residues (R, K, H) are more than negative amino acid residues (D, E) in 3S. These features help to distinguish 2S, 3SMI and 3SDI providing insights to homodimer folding and binding.

1. Introduction

Homodimers have an important role in catalysis and regulation. The mechanism of homodimer folding without stable intermediate (2-state) and with stable intermediate (3-state) is intriguing. Homodimer binding has implications in interaction network for drug target definition and validation. The commercial importance of a number of homodimers has been shown recently (2008) with granted patent (US ID: 17656367; 7387886; 7105308 and 6946543) rights [1]. In addition, a couple of homodimers have been exploited as drug targets [2,3]. Tanaka et al. showed RXR (retinoid receptor) homodimer as a target for anti-cancer ligand (rexinoid) [2]. Schülke et al. showed prostate-specific membrane antigen (PSMA) ECTO homodimer as a known target for prostate cancer [3]. Therefore, it is of interest to study homodimer folding and binding using structural data with known folding information. Neet and Timm reviewed 17 homodimers with known unfolding data (denatured species fraction is measured by spectroscopic techniques such as fluorescence, circular dichroism, NMR, and absorption) [4]. They noticed that some dimers unfold through a step forming intermediates (3-state—3S), while others do not (2-state—2S). Tsai et al. investigated 187 stable and 57 symmetry

Abbreviations: 2S, 2-state; 3S, 3-state; 3SMI, 3-state with monomer intermediate; 3SDI, 3-state with dimer intermediate; B/2, interface area; H_f, hydrophobicity factor; IAR, interface amino acid residue; I/T, interface to total residues; NMR, nuclear magnetic resonance; ML, monomer length; SLL, squared loop length.

* Corresponding author at: Biomedical Informatics, 17A Irulan Sandy Annex, Pondicherry 607402, India. Tel.: +91 413 2633 722; fax: +91 413 2633 589.
E-mail address: kangueane@bioinformation.net (P. Kangueane).

1093-3263/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2009.04.002

related oligomeric structural interfaces [5]. They observed that 2S interfaces are similar to protein cores and 3SMI interfaces resemble monomer surfaces in structural elements.

Unfolding of desulfoferridoxin homodimer using GuHcl was shown by Apiyo et al. [6]. They observed a high thermodynamic stability of desulfoferridoxin indicating the formation of intermediates in unfolding (showing evidence for 3S folding mechanism). Mazzini et al. observed dimer disassociation before unfolding rather than the reverse in bovine odorant binding protein [7]. Tiana and Broglia studied the evolution of two identical (homodimer) 20 letter residue chains within the framework of a lattice model using Monte Carlo simulation [8]. They proposed that 3S binds as it folds, while 2S folds and then binds. Levy et al. grouped 2S and 3SMI dimers based on the relationship between intramolecular/intermolecular contacts ratio and interface hydrophobicity using information gleaned from 11 homodimers with known unfolding data [9]. They suggested that the native protein 3D structure is the major factor governing the choice of homodimer folding and binding mechanism.

Mei et al. reviewed 32 homodimers with structural data (class A—18 (2S); class B—10 (3SMI); class C—4 (3SDI)) with known unfolding data [10]. They defined interface amino acid residue (IAR) and squared loop length (SLL) to group 2S, 3SMI and 3SDI. They performed the grouping of 2S, 3S mechanism based on mass and stable monomeric intermediate in the folding–unfolding pathway. Li et al. used 41 homodimer structures with known folding mechanism data to distinguish 2S from 3S [11]. Their analysis showed that small proteins with large interface area and high interface hydrophobicity are 2S (80%). However, 3S proteins are large with small interface area and low interface hydrophobicity (60%). A better understanding of homodimer folding and binding is still not available despite these reports. Hence, we used the dataset of Li et al. (2005) consisting of 2S (25 dimers) and 3S (16 dimers) to probe into the features governing homodimer folding and binding.

2. Methodology

2.1. Dataset

We used the dataset of Li et al. consisting of 2S (25 dimers) and 3S (16 dimers) homodimers in this study [11]. We then compiled the individual function for each of the protein dimer using KEYWORD search from literature source (Table 1). These proteins have both regulatory and catalysis functions. The dataset including mean, minimum, maximum, medium and standard deviation is presented in Table 2. Table 2 describes the characteristics of the dataset under study.

2.2. Data analysis tools

The PROTPARAM tool from EXPASY (http://www.expasy.ch/cgi-bin/protparam) is used to identify hydrophobic (V, F, M, L, A, I, P, G) and hydrophilic (D, E, R, K, H, S, T, N, Q, W, C, Y) residues in 2S and 3S proteins. The PSAP (http://iris.physics.iisc.ernet.in/cgi-bin/psap/index.pl) server was used to identify and count positively and negatively charged residues. Interface area and hydrophobicity factor (ratio of hydrophobicity difference between interface to surface and interior to surface) data for each entry were taken from Li et al. described elsewhere [11]. PROTORP server (http://www.bioinformatics.sussex.ac.uk/protorp/) was used to count interface residues in the dataset.

3. Results

Fig. 1 shows the distribution of interface to total (I/T) residues ratio for 2S (Fig. 1a), 3SMI (Fig. 1b) and 3SDI (Fig. 1b) proteins. The ratio in 2S is larger than 3S. This implies that a major portion of the monomer protein in 2S is at the interface. This is not true in 3SMI and 3SDI as only a small fraction of the monomer protein is associated with the interface (Fig. 1b). Interface to total (I/T) residues ratio is similar for both 3SMI (mean ML = 208) and 3SDI (mean ML = 404) despite difference in mean monomer size.

Fig. 2 shows the relationship between monomer length (ML) and interface residues in the dataset. Interface residues correlate fairly with monomer length in 2S ($r^2 = 0.41$) and 3SMI ($r^2 = 0.52$). This is not true for 3SDI ($r^2 = 0.17$). Fig. 3 shows that interface area does not correlate with interface residues ($r^2 < 0.001$) in 2S. However, this is not true for 3SMI ($r^2 = 0.61$) and 3SDI ($r^2 = 0.28$). Fig. 4 shows that interface area does not correlate with monomer length ($r^2 = 0.02$) in 2S. However, this is not the case for 3SMI ($r^2 = 0.25$) and 3SDI ($r^2 = 0.09$).

The mean hydrophobicity factor (H_f) in the dataset is 3-fold less in 3S than 2S (Table 2). Fig. 5 shows that H_f does not correlate with interface area in 2S ($r^2 = 0.03$) and 3SDI ($r^2 < 0.01$). However, a causal relation is seen in 3SMI ($r^2 = 0.23$) with interface area. Fig. 6 shows that hydrophilic residues (E, R, K, S, Q) are prominent in 2S than 3S proteins. Nonetheless, a difference was not seen for

Table 1
Homodimers with known structure, function and folding mechanism.

	Molecular function	Classification
2S		
2cpg	DNA binding	Gene regulating protein
1arr	DNA binding	Gene regulating protein
1rop	DNA binding	Transcription regulation
5cro	DNA binding	Gene regulating protein
1bfm	DNA binding	Histone protein
1a7g	DNA binding	Transcription regulation
1vqb	DNA binding	DNA binding protein
1b8z	DNA binding	DNA binding protein
1ety	DNA binding	Transcription regulation
1y7q	DNA binding	Transcription regulation
1a8g	Aspartic type endopeptidase activity	Acid proteinase/inhibitor complex
1siv	DNA gyrase inhibitor activity	Plasmid
1vub	DNA gyrase inhibitor activity	Plasmid
1cmb	DNA binding	DNA binding protein
3ssi	None	Serine protease inhibitor
1wrp	DNA binding	DNA binding protein
1bet	Growth factor	None
1oh0	Steroid delta isomerase activity	Isomerase
2gsr	Glutathione transferase activity	Transferase/inhibitor complex
1gsd	Glutathione transferase activity	Transferase
1gta	Glutathione transferase activity	Transferase
2bqp	Sugar binding	Sugar binding protein
1hti	Triose phosphate isomerase activity	Isomerase
1ee1	ATP binding	Ligase
3SDI		
1mul	DNA binding	DNA binding protein
1hqo	None	Signaling protein
1psc	Zinc ion binding	Hydrolase
1cm7	3-Isopropylmalate dehydrogenase activity	Oxidoreductase
1aoz	Oxidoreductase activity	Oxidoreductase
1nl3	ATP binding	Protein transport
3SMI		
1a43	None	Viral protein
1qll	Calcium ion binding	Neurotoxin
1dfx	None	Electron transport
1yai	Metal ion binding	Oxidoreductase
1spd	Metal ion binding	Oxidoreductase
1run	DNA binding	DNA binding protein
11gs	Glutathione transferase activity	Transferase
1tya	RNA binding	Ligase
1nd5	Acid phosphatase activity	Hydrolase
2crk	Catalytic activity	Transferase

Table 2
Dataset statistics of homodimer structures with known folding data.

State	Num	Monomer size (residues)					Interface area (Å^2)		Hydrophobicity factor	
		Min	Max	Mean	Median	SD	Mean	SD	Mean	SD
2S	25	45	271	125.67	102.5	66.347	1509.12	475.13	1.9560	2.8550
3S										
3SMI	10	72	381	208.3	175.5	106.877	1067.5	468.16	0.4521	0.6951
3SDI	6	90	835	404.5	346	258.797	1705.5	357.73	1.0927	0.7616
Total	41			188.15	123	153.383	1430.14	498.91	1.2433	2.3042

Num = number of proteins in each subset; Min = minimum length in residues; Max = maximum length in residues; SD = standard deviation about the mean.

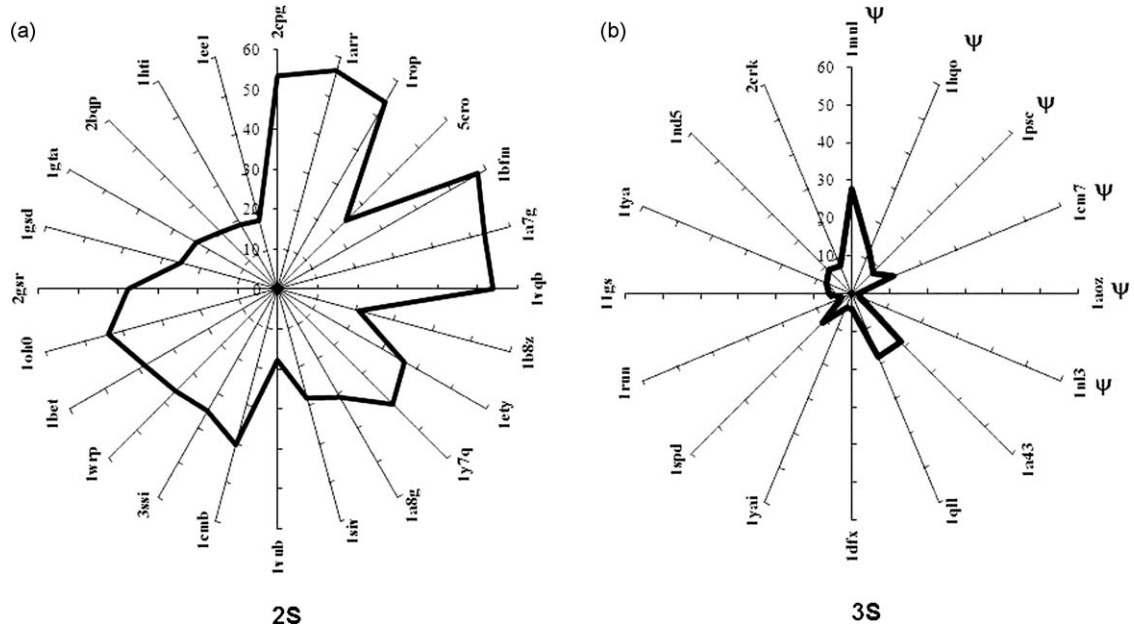

Fig. 1. Radar diagram showing the distribution of interface to total residue ratio is given. The mean value is 36% for 2S and 9.5% for 3S.

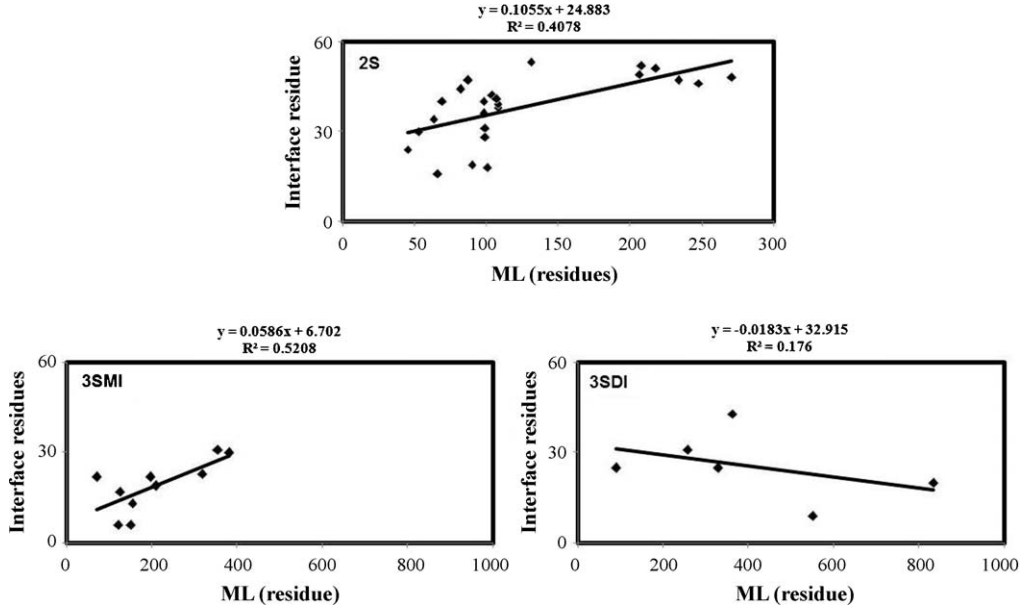

Fig. 2. The relationship between monomer size and interface residues is given for 2S, 3SMI and 3SDI.

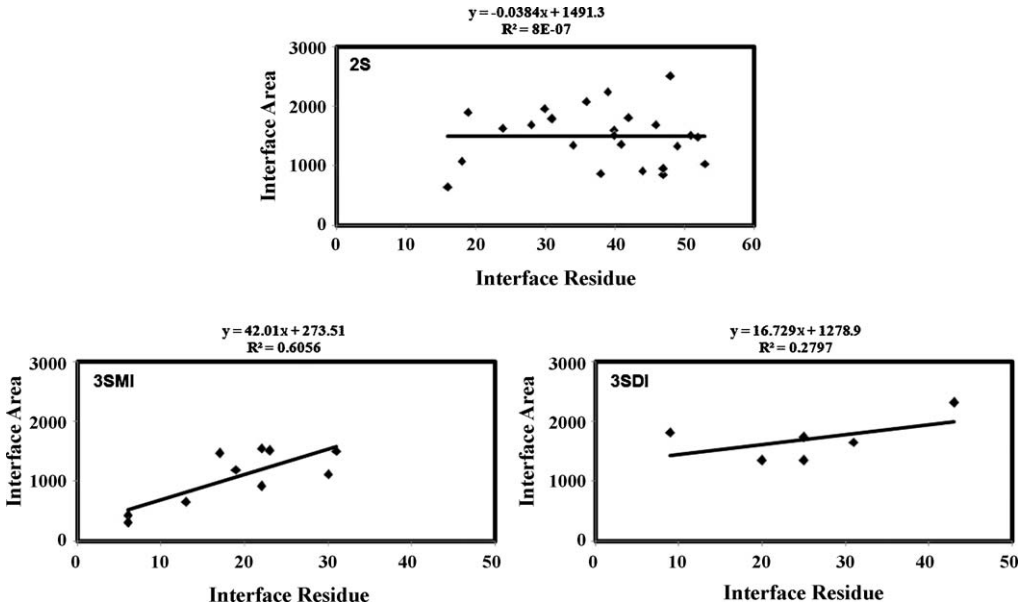

Fig. 3. The relationship between interface area and interface residues is given for 2S, 3SMI and 3SDI.

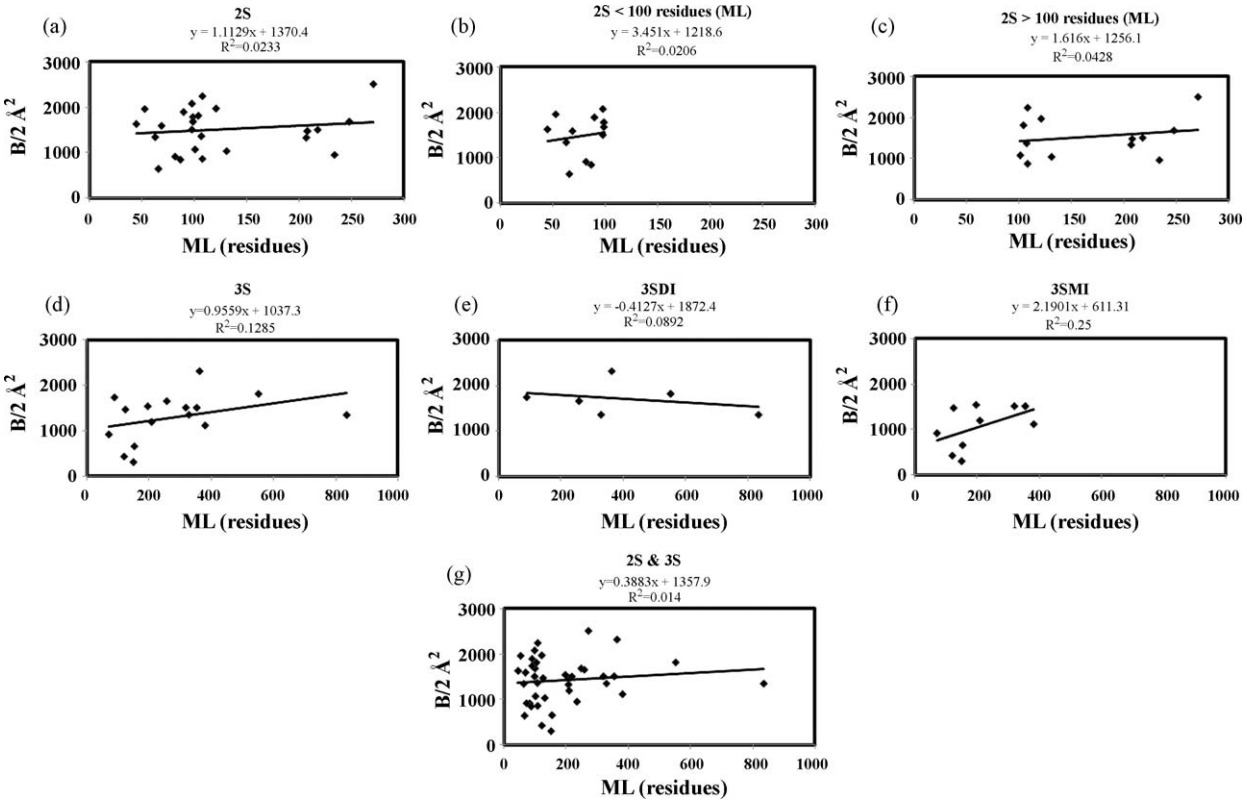

Fig. 4. The relationship between interface area and monomer size is given for 2S, 3SMI and 3SDI.

hydrophobic residues in 2S and 3S proteins. Fig. 7 shows that charged negative residues (D, E) are more than positive residues (R, K, H) in 2S proteins and charged positive residues (R, K, H) are more than negative residues (D, E) in 3S proteins.

4. Discussion

Dimers have an essential role in both catalysis and regulation. The network of interactions inside the cellular system is certainly beyond our realization. The combination of interactions is huge, yet specific and sensitive. The formation of homodimers for catalysis and regulation is fascinating. The mechanism of formation for such homodimer interfaces is more intriguing. The known three types (2S, 3SMI, 3SDI) of folding mechanism associated with the binding of homodimers are attractive for scientific investigation. This has implication for selecting homodimers as drug targets [2,3]. Homodimer folding mechanism is postulated by measuring denatured species fraction using spectroscopic techniques such as

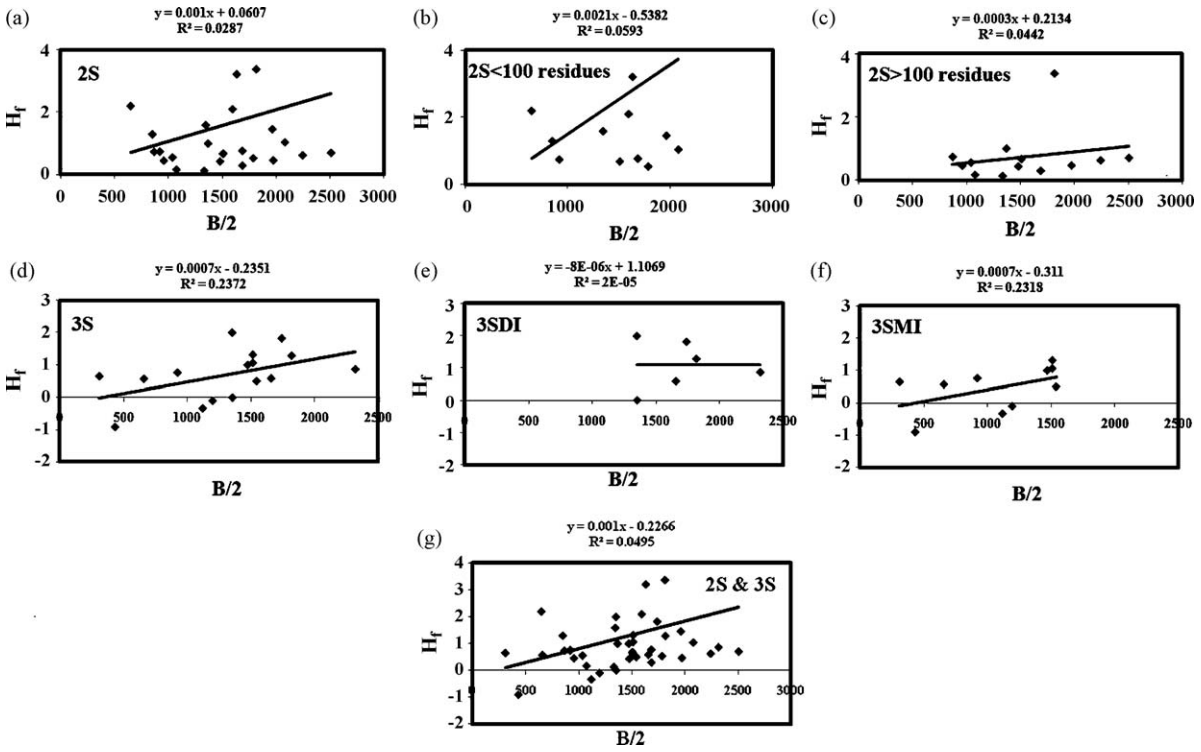

Fig. 5. The relationship between interface area and hydrophobicity factor is given for 2S, 3SMI and 3SDI.

fluorescence, circular dichroism, NMR and absorption [4,6,7]. The folding mechanism of several homodimers has been documented in the literature and this data is compiled by few investigators [4,9–11]. 3-Dimensional structures for some of these homodimers are available at the protein data bank (PDB). The list of homodimer structures with known folding data is compiled and documented elsewhere [5,9–11]. The size and nature of the dataset varies in each of these compilations. Here, we used the dataset (41

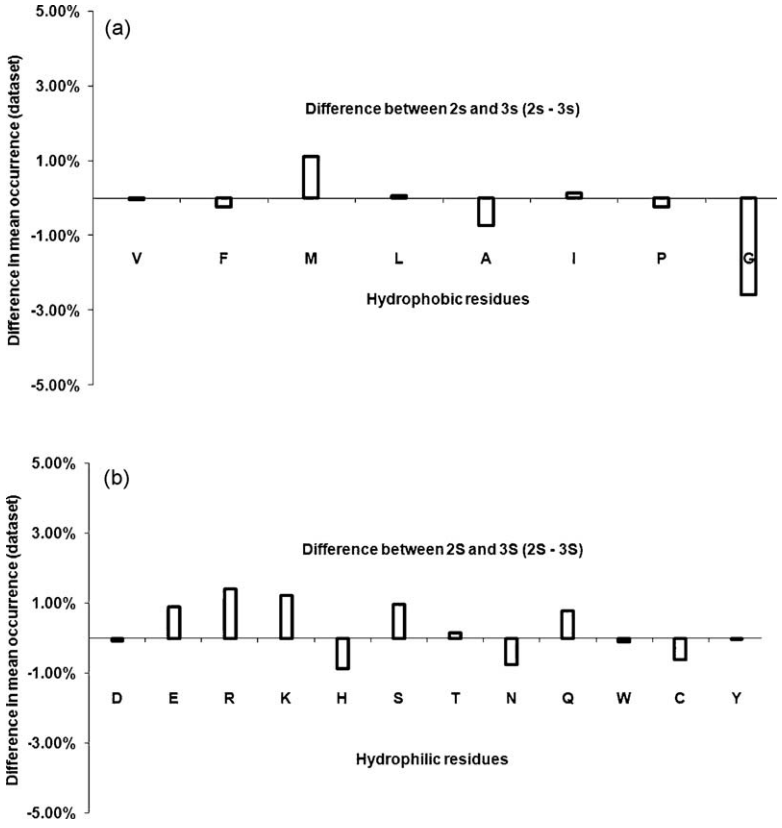

Fig. 6. The difference in distribution of hydrophobic and hydrophilic residues in 2S and 3S is presented.

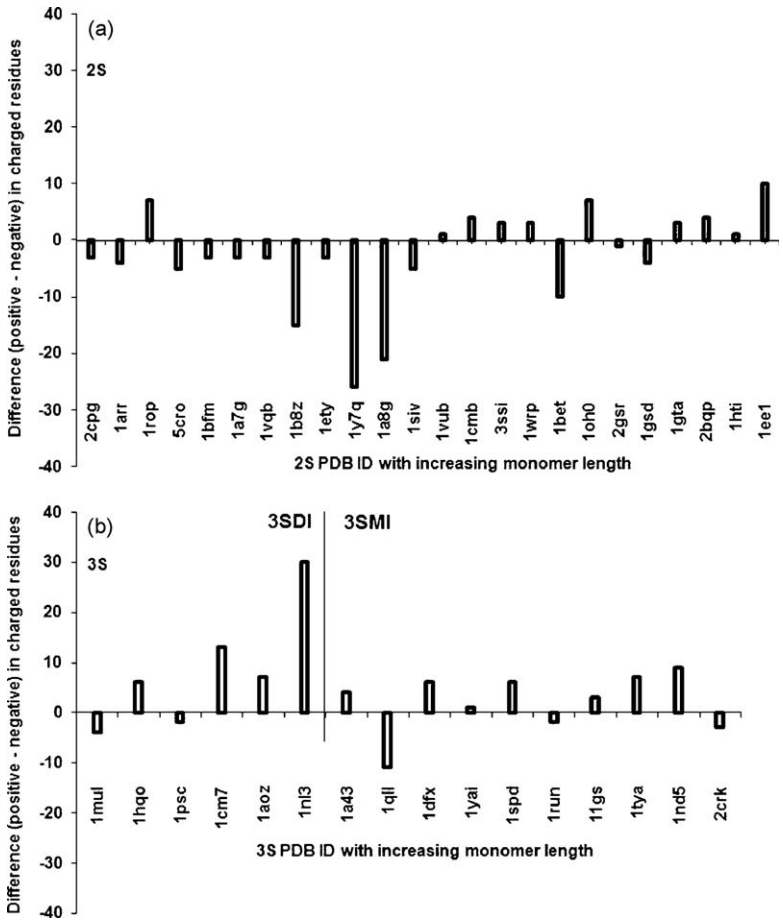

Fig. 7. The difference in distribution between positive and negative charged residues is given for 2S, 3SMI and 3SDI.

homodimers) of Li et al. consisting of twenty-five 2S, ten 3SMI and six 3SDI to probe into the features governing homodimer folding and binding [11].

Several studies have documented the importance of structural parameters in homodimer folding and binding [5,8–11]. Tsai et al. observed that 2S interfaces are similar to protein cores and 3SMI interfaces resemble monomer surfaces in structural elements [5]. Tiana and Broglia developed a framework of a lattice model using Monte Carlo simulation to propose 3S binds as it folds, while 2S folds and then binds [8]. Levy et al. grouped 2S and 3SMI using relationship between intramolecular/intermolecular contacts ratio and interface hydrophobicity and suggested that the native protein 3D structure is the major factor governing the choice of homodimer folding and binding mechanism [9]. Mei et al. defined interface amino acid residue (IAR) and squared loop length (SLL) to group class-1 (2S), class-2 (3SMI) and class-3 (3SDI) to propose high stability of class-2 homodimers with large IAR and SLL [10]. Li et al. used monomer length (ML), interface area (B/2) and hydrophobicity factor (H_f) to differentiate 2S, 3SMI and 3SDI [11]. However, no study is available relating structural features such as ML, B/2, I/T and H_f.

Fig. 8 gives example structures for 2S (1ARQ), 3SMI (1SPD), 3SDI (1CM7) with structural features and their values. Fig. 8a shows a 53 residues long ARC repressor (PDB ID: 1ARQ) having I/T ratio of 80% and B/2 of 2007 Å^2 is a 2S protein. Fig. 8b shows a 161 residues long Cu/Zn superoxide dismutase (PDB ID: 1SPD) having I/T ratio of 13% and B/2 of 658 Å^2 is a 3SMI protein. Fig. 8c shows a 363 residues long 3-isopropylmalate dehydrogenase (PDB ID: 1CM7) having I/T ratio of 16% and B/2 of 2317 Å^2 is a 3SDI protein. These example structures show that the 2S (PDB ID: 1ARQ) protein is small sized with high I/T ratio and large B/2, 3SMI (PDB ID: 1SPD) protein is moderate sized with low I/T and small B/2 and 3SDI (PDB ID: 1CM7) protein is large sized with low I/T and large B/2. Thus, these structures demonstrate structural features distinguishing 2S, 3SMI and 3SDI. However, such distinct features may not be always observed in a dataset of 2S, 3SMI and 3SDI structures. Nonetheless, there exist common features specific to a subset. Therefore, it is of interest to study the significance of structural features such as I/T, B/2, ML and H_f in a dataset of homodimers with known folding data.

Interface to total (I/T) residues ratio (percent) is large for 2S (mean = 36.08%) than 3S (mean = 9.5%) in the dataset (Fig. 1). This suggests that a major portion of the monomer in 2S is involved in binding at the interface due to the small size of 2S proteins (mean = 125.6 residues). This is similar to the findings that small sized monomers are generally 2S proteins as documented by Mei et al. [10] and Li et al. [11]. However, a similar observation is not seen in 3SMI and 3SDI, where only a small fraction of the monomer is involved in binding at the interface (Fig. 1b). Nevertheless, the interface to total (I/T) residues ratio in 3SMI (mean ML = 208) and 3SDI (mean ML = 404) is similar, despite a 2-fold difference in mean size.

A study by Zhanhua et al. showed the weak relationship of monomer length with interface area ($r^2 = 0.20$) and interface residues ($r^2 = 0.23$) in a large dataset of 170 homodimers [12]. Therefore, it is of interest to study similar relationships within the 2S, 3SMI and 3SDI subsets. Interface residues correlate with monomer length (Fig. 2) in 2S ($r^2 = 0.41$) and 3SMI ($r^2 = 0.52$). The data presented here for 2S and 3SMI is different from the relation shown by Zhanhua et al. [12]. However, this is not true for 3SDI

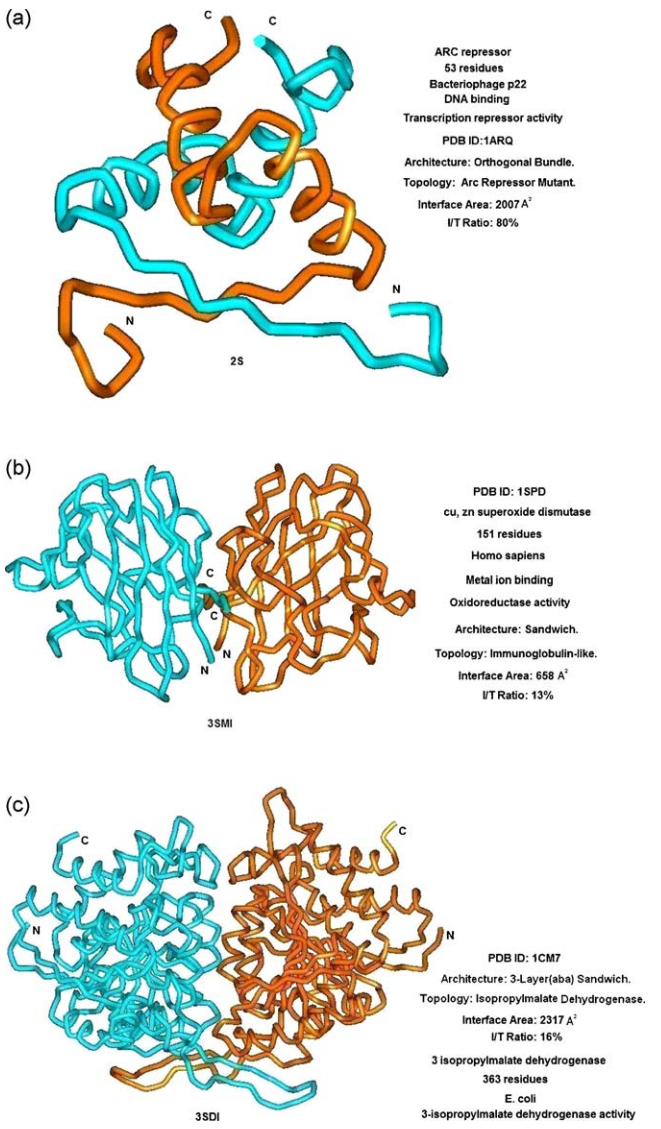

Fig. 8. Example structures for 2S (1ARQ), 3SMI (1SPD), 3SDI (1CM7) is shown with structural features and their values given next to the structures.

(r^2 = 0.17). Thus, a low correlation is predominant in 3SDI proteins for size and interface residues. The number of residues forming the binding interface increases with increase in monomer size in 2S and 3SMI. In the case of 3SDI, the number of interface residues does not increase with monomer size, suggesting that a limit is attained despite longer size. Interface area (Figs. 3 and 4) does not correlate with interface residues ($r^2 < 0.01$) and monomer length (r^2 = 0.02) in 2S. This is despite a relationship with interface residues and monomer length (r^2 = 0.41). However, this is not true for 3SMI (r^2 = 0.61 with interface residues and r^2 = 0.25 with size). In 3SDI, a different relationship is seen (r = 0.28 with interface residues and r = 0.09 with size). The mean hydrophobicity factor (ratio of hydrophobicity difference between interface to surface and interior to surface) is 3-fold more in 2S than 3S (Table 2). It does not correlate with interface area in 2S (r^2 = 0.03) and 3SDI (r^2 = 0.0). However, a weak causal relation is seen in 3SMI (r^2 = 0.23).

Tsai et al. observed that 2S interfaces are similar to protein cores and 3SMI interfaces resemble monomer surfaces in structural elements [5]. Levy et al. suggested that the native protein 3D structure is the major factor governing the choice of homodimer folding and binding mechanism [9]. Hence, we studied the distribution of hydrophobic and hydrophilic residues in 2S and 3S monomers (Fig. 6). Hydrophilic residues (E, R, K, S, Q) are prominent in 2S than 3S unlike equal preference for hydrophobic residues. Data also shows that charged negative residues (D, E) are more than positive residues (R, K, H) in 2S and charged positive residues (R, K, H) are more than negative residues (D, E) in 3S (Fig. 7). We believe that these features help to distinguish 2S, 3SMI and 3SDI by providing insights to homodimer folding and binding.

5. Conclusion

Available data on homodimer folding and binding is persuasive in nature. 2S, 3SMI and 3SDI homodimers fold and bind through different mechanisms. Information gleaned from 3D structures of homodimers with known folding data provides insight to their binding mechanism. Our analysis documents trends in parameters (monomer size, interface residues, interface area, hydrophobicity factor, hydrophilic residues and charged residues) for distinguishing 2S from 3S proteins.

Data suggests that interface to total residues ratio is large for 2S than 3S. Interface residues correlate with monomer size in 2S and 3SMI. Interface area correlates with interface residues and monomer size in 3SMI. The mean hydrophobicity factor is 3-fold more in 2S than 3S. Data also shows that hydrophilic residues (E, R, K, S, Q) are prominent in 2S than 3S and charged negative residues (D, E) are more than positive residues (R, K, H) in 2S. It is also shown that charged positive residues (R, K, H) are more than negative residues (D, E) in 3S. Thus, it is possible to differentiate 2S from 3S proteins using these parameters. We hope to revisit this problem using an increased dataset for consistency check and data examination.

Acknowledgments

We wish to express our sincere appreciation to all members of the Biomedical Informatics for many discussions on the subject of this article. A.M. and S.L. thank Dr. Lazar Mathew at VIT University for his organizational support.

References

[1] The United States Patent and Trademark Office Database, USA.
[2] T. Tanaka, K.S. Suh, A.M. Lo, L.M. De Luca, p21WAF1/CIP1 is a common transcriptional target of retinoid receptors: pleiotropic regulatory mechanism through retinoic acid receptor (RAR)/retinoid X receptor (RXR) hetrodimer and RXR/RXR homodimer, J. Biol. Chem. 282 (2007) 29987–29997.
[3] N. Schülke, O.A. Varlamova, G.P. Donovan, D. Ma, J.P. Gardner, D.M. Morrissey, R.R. Arrigale, C. Zhan, A.J. Chodera, K.G. Surowitz, P.J. Maddon, W.D. Heston, W.C. Olson, Prostate-specific membrane antigen $(PSMA)_{ECTO}$ homodimer is a known target for prostate cancer, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 12590–12595.
[4] K.E. Neet, D.E. Timm, Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation, Protein Sci. 3 (1994) 2167–2174.
[5] C.J. Tsai, D. Xu, R. Nussinov, Structural motifs at protein–protein interfaces: protein cores versus 2S and 3S model complexes, Protein Sci. 6 (1997) 1793–1805.
[6] D. Apiyo, K. Jones, J. Gulrdy, P. Wittung-Stafshede, Equilibrium unfolding of dimeric desulfoferridoxin involves a monomeric intermediate: iron cofactors dissociate after polypeptide unfolding, Biochemistry 40 (2001) 4940–4948.
[7] A. Mazzini, A. Maia, M. Parisi, R.T. Sorbi, R. Ramoni, S. Grolli, R. Favilla, Reversible unfolding of bovine odorant binding protein induced by guanidinium hydrochloride at neutral pH, Biochim. Biophys. Acta 1599 (2002) 90–101.
[8] G. Tiana, R.A. Broglia, Design and folding of dimeric protein, Proteins 49 (2002) 82–94.
[9] Y. Levy, P.G. Wolynes, J.N. Onuchic, Protein topology determines binding mechanism, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 511–516.
[10] G. Mei, A. Divenere, A. Rosata, A. Finazzi-Agrò, Importance of being dimeric, FEBS J. 272 (2005) 16–27.
[11] L. Li, K. Gunasekaran, J.G. Gan, C. Zhanhua, P. Shapshak, M.K. Sakharkar, P. Kangueane, Structural features differentiate the mechanisms between 2S and 3S folding of homodimers, Bioinformation 1 (2005) 42–49.
[12] C. Zhanhua, J.G. Gan, L. Li, M.K. Sakharkar, P. Kangueane, Protein subunit interfaces: heterodimers versus homodimers, Bioinformation 1 (2005) 28–39.