



## ORIGINAL RESEARCH

# Search and Analysis of Identical Reverse Octapeptides in Unrelated Proteins

Konda Mani Saravanan, Samuel Selvaraj \*

*Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli 620024, India*

Received 14 August 2012; revised 20 November 2012; accepted 21 November 2012

Available online 21 March 2013

**KEYWORDS**Reverse octapeptide;  
Octapeptide search;  
Dipole and quadrupole;  
Volume and surface area;  
Protein stability

**Abstract** For the past few decades, intensive studies have been carried out in an attempt to understand how the amino acid sequences of proteins encode their three dimensional structures to perform their specific functions. In order to understand the sequence-structure relationship of proteins, several sub-sequence search studies in non-redundant sequence-structure databases have been undertaken which have given some fruitful clues. In our earlier work, we analyzed a set of 3124 non-redundant protein sequences from the Protein Data Bank (PDB) and retrieved 30 identical octapeptides having different secondary structures. These octapeptides were characterized by using different computational procedures. This prompted us to explore the presence of octapeptides with reverse sequences and to analyze whether these octapeptides would adopt similar structures as that of their parent octapeptides. Our identical reverse octapeptide search resulted in the finding of eight octapeptide pairs (octapeptide and reverse octapeptide) with similar secondary structure and 23 octapeptide pairs with different secondary structures. In the present work, the geometrical and biophysical characteristics of identical reverse octapeptides were explored and compared with unrelated octapeptide pairs by using various computational tools. We thus conclude that proteins containing identical reverse octapeptides are not very abundant and residues in the octapeptide pairs do not contribute to the stability of the protein. Furthermore, compared to unrelated octapeptides, identical reverse octapeptides do not show certain biophysical and geometrical properties.

**Introduction**

With the vast amount of knowledge gained from the three dimensional (3D) structures of proteins, one might expect to

easily understand sequence-structure-function relationships. However this is not the case and many unsolved problems remain. Various attempts have been made to study the stability and folding of proteins and peptides by using reduced representation of proteins as well as by reading the protein sequences backward. Several theoretical and experimental approaches toward studying retro proteins have been carried out and give contradictory results [1].

Skolnick and coworkers made a systematic study by generating retro proteins and found a very low number of similar 3D structures between inversely similar proteins and they also reported that short helices keep their conformations even when the sequence is inverted [2]. Systematic search in public domain

\* Corresponding author.

E-mail: [selvarajsamuel@gmail.com](mailto:selvarajsamuel@gmail.com) (Selvaraj S).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.



Production and hosting by Elsevier

sequence databases revealed that a protein with identical composition but with backward read primary structure should fold under native conditions into a similar structure compared with the original sequence [3].

Since the retro protein has the same amino acid composition and hydrophobicity profiles as the native protein, all structure prediction methods based on amino acid composition may fail. The assumption that the retro protein would adopt the mirror image of the native protein is very unlikely, because right-handed helices would have to be replaced by left-handed helices. Changes in the secondary structure of the retro protein are significant because the packing of the native topology is similar to the packing of the original protein [4]. Nebel et al proposed that the abundance of retro proteins primarily can be explained by the fact that they display the same repeat structures and amino acid propensity of existing proteins [5].

In contrast to proteins, studies of smaller peptide fragments with inverted sequences have provided fruitful clues toward understanding protein sequence-structure relationships. Several studies on reversed proteins and peptides consisting of D-amino acids which form the mirror image structures of the respective L-amino acid protein have been carried out [6–8] and validated experimentally by synthesizing D-human immunodeficiency virus protease which showed reciprocal chirality [9].

There are several other interesting studies of retro peptides. One such example is a hairpin peptide mimic of the FcRI chain that inhibits IgE-FcRI interactions both in its native and retro form, which has potential for the treatment of allergic disorders [10]. Rai has used the S peptide from ribonuclease S as a model system and explored the relationship between native and reverse conformations of the peptide by using circular dichroism (CD) [7].

Huang and coworkers synthesized a novel peptide by reversing the sequence of the human metallothionein-2 $\alpha$  domain to analyze its chemical and spectroscopic properties. The results of their analysis indicate that the reversion of the amino acid sequence for  $\alpha$  domain does not change its foldability and metal-binding capacity, suggesting that the order of its sequence is not essential for the formation of a critical metal-tetrathiolate nucleus [11]. In our previous work, we performed a systematic analysis of the role of long-range contacts in homologous families of proteins in determining the final native structure [12]. Our analysis indicated the importance of cooperative long-range interactions in determining the final conformations of two proteins with 88% sequence identity but adopting different folds and functions [13]. The importance of cooperative long-range interactions for protein stability was further verified experimentally by studying the fragments of chymotrypsin inhibitor 2 in barley [14]. In our previous paper, we have characterized 30 identical octapeptides adopting different conformations by computing difference in the number of long-range contacts of residues [15].

In the present work, we performed a systematic search of identical reverse octapeptides in the sequences of the Protein Data Bank (PDB) [16] to explore the structural and functional features of the protein fragments in reversed form. The results from our analysis imply that protein fragments of eight residues of a given protein and their reversed form in another unrelated protein are not very abundant. The geometric and biophysical properties, such as volume, surface area, dipole and quadrupole moments and continuous symmetry measure in such octapeptides, were examined in detail. Also the gene ontology (GO)

terms of the proteins containing reverse octapeptides are extracted to understand the structure-function relationship.

## Results

### Reverse octapeptide search

While searching for the reverse form of the overlapping octapeptide segments in each of the proteins of the 3124 PDB sequences, we found 8 octapeptide pairs (octapeptide and reverse octapeptide) with similar secondary structure (Table S1) and 23 octapeptide pairs with different secondary structures (Table 1). The occurrence of proteins containing identical reverse octapeptides with similar and different secondary structures is very low, which was 0.5% and 1.4%, respectively, while 98% of proteins do not contain identical reverse octapeptides. These data demonstrated the low abundance of such inverse fragments in a large number of unrelated protein sequences. However, it is not surprising that an octapeptide and its reverse sequence adopt similar secondary structures, due to the similar amino acid composition and certain similar biophysical properties such as hydrophobicity, hydrophilicity, etc. Interestingly, we have found eight octapeptide pairs with similar secondary structure, which are mainly made up of helices.

The occurrence of each of the 20 amino acid residues in reverse octapeptides having similar and different conformations is shown in Figure 1. Leucine (L) and glutamic acid (E), which are generally considered as helix favoring residues, appear with high percentage, whereas in peptides adopting different secondary structures, their occurrence was relatively low (~10% and 9%, respectively). Since glutamic acid and its salts perform important functions such as neurotransmitting and flavor enhancing, understanding the role of glutamic acid in octapeptides may provide fruitful insights in peptide and biomaterials design [17]. The low occurrence of tryptophan (W) may be explained by the fact that the L-stereoisomer of tryptophan is present in proteins, but the D-stereoisomer is only occasionally found in naturally-produced peptides [18].

The amino acid residues in the octapeptide pairs with similar secondary structure are mainly occupied by hydrophilic residues and they are generally surrounded by very minimum contacting residues [19–20]. Since hydrophobic residues tend to be buried in the core of proteins, the low occurrence of hydrophobic residues in octapeptide pairs with similar secondary structure suggests that octapeptide fragments would be found on the protein surface and contribute to various interactions with other molecules. On the other hand, majority of the residues in octapeptide pairs with different secondary structures are neither hydrophilic (D, E, R, K and H) nor hydrophobic residues (F, I, L, M, V, W, A and P). Instead, these octapeptide pairs are mainly made up of residues such as G, S, T, C, N, Q and Y.

### Volume and surface area of octapeptide pairs

The functional specificity of a peptide arises from its unique physico-chemical properties such as volume and surface area, which are important geometrical quantities associated with macromolecular structures and motions. Hence, we performed a systematic analysis of surface area and volume of octapeptide pairs with similar conformations (Table S1) and those with different conformations (Table 1). The correlation coefficient

**Table 1** Biophysical and geometrical properties of octapeptides with inverse sequences having different secondary structures

No.		PDB ID	Octapeptide	Secondary structure	Stabilizing residues	Hphil (%)	Hpho (%)	Others (%)	Attribute	RMSD	CSM	Volume (Å <sup>3</sup> )	Surface area (Å <sup>2</sup> )	Quadrupole (Debyes)
1	P1	1A8D_A(52-59)	VPGINGKA	ECCCCCEE	No	12	25	62	Basic	4.3	96.655	971	621.9	81
	P2	1YZY_A(88-95)	AKGNIGPV	CCCCHHHH	No	12	25	62	Basic	4.3	94.398	949	630.15	42
2	P1	1D8W_A(418-425)	YEKEILSR	HHHHCCCC	No	30	20	50	Neutral	0.6	83.482	1260	718.53	100
	P2	2E52_A(244-251)	RSLIEKEY	HHHHHHHH	No	30	20	50	Neutral	0.6	85.753	1305	772.85	24
3	P1	1DM9_B(103-110)	KLNALTMP	HHHHCCCC	No	38	0	62	Basic	3.0	99.401	959	580.7	71
	P2	1W8K_A(137-144)	PMTLANLK	CEEHHHHH	No	38	0	62	Basic	3.0	92.286	1118	653.28	45
4	P1	1ED1_A(47-54)	AESLLENK	CHHHHHCH	No	20	13	67	Acidic	1.6	80.533	1117	740.55	163
	P2	1R6X_A(20-27)	KNELLSEA	HHHHHHHH	No	20	13	67	Acidic	1.6	77.565	1129	719.77	110
5	P1	1G4M_A(106-113)	KKLGEHAY	HHCCCCEE	No	50	25	25	Basic	1.0	90.681	1178	802.44	248
	P2	2UVK_A(276-283)	YAHEGLKK	CCCCCCCC	No	50	25	25	Basic	1.0	99.514	1102	757.62	27
6	P1	1GJW_A(2-9)	LLREINRY	HHHHHHHH	No	38	50	12	Basic	3.5	81.161	1378	849.18	43
	P2	2ODI_A(231-238)	YRNIERLL	EEEEEECC	No	38	50	12	Basic	3.5	59.256	1363	973.29	234
7	P1	1ITX_A(354-361)	QTCTGGSS	CCCCEECC	No	0	0	100	Neutral	1.4	96.532	864	659.05	88
	P2	2V3I_A(20-27)	SSGGTCTQ	CCCCEEEE	No	0	0	100	Neutral	1.4	87.245	960	637.51	109
8	P1	1JL1_A(139-146)	AAAAMNPTL	HHHHCCCC	No	40	20	40	Neutral	0.04	95.226	972	660.34	71
	P2	2NVO_A(383-390)	LTPNMAAA	CCHHHHHH	No	40	20	40	Neutral	0.04	93.055	967	580.17	80
9	P1	1N1B_A(481-488)	YHDILCLA	CCHHHHHH	No	9	18	73	Neutral	2.7	86.665	1140	686.48	105
	P2	2FA1_A(142-149)	ALCLIDHY	EEEEEEEE	No	9	18	73	Neutral	2.7	90.691	1154	797.82	53
10	P1	1O2D_A(131-138)	VVEIPTTA	EEEEEECC	132V	20	27	53	Acidic	1.6	93.583	1023	724.32	87
	P2	1TG7_A(626-633)	ATTPIEVV	CCEEEEE	No	20	27	53	Acidic	1.6	93.519	1026	735.4	49
11	P1	1QOY_A(177-184)	AYAGAAAG	HHHCCCCC	No	0	12	88	Neutral	3.2	99.551	787	550.79	57
	P2	1RWR_A(209-216)	GAAAGAYA	CCCCCEE	No	0	12	88	Neutral	3.2	99.532	776	590.71	67
12	P1	1QUS_A(277-284)	GQAPGLPN	CCCCCCCC	No	20	20	60	Neutral	2.5	67.633	918	649.24	134
	P2	1QWO_A(286-293)	NPLGPAQG	CCCCHHHH	No	20	20	60	Neutral	2.5	75.013	886	558.93	91
13	P1	1SU8_A(269-276)	IVSVSKEM	HHHHHHHC	No	19	25	56	Neutral	3.4	81.048	1126	667.56	53
	P2	1W77_A(75-82)	MEKSVSVI	CCCCEEEE	80S, 82I	19	25	56	Neutral	3.4	89.778	1112	786.99	103
14	P1	1SVF_A(177-184)	SPAITAAN	HHHHHHCC	No	0	12	88	Neutral	4.1	90.993	907	555.91	23
	P2	1YNF_A(110-117)	NAATIAPS	HCEEEEEC	No	0	12	88	Neutral	4.1	84.159	896	667.46	64
15	P1	1UA4_A(409-416)	IKEGIGEV	CCCCEEEE	No	38	38	25	Acidic	1.9	52.912	1048	737.5	48
	P2	1X2I_A(48-55)	VEGIGEKI	CCCCHHH	No	38	38	25	Acidic	1.9	46.389	1043	706.78	19
16	P1	2ANE_A(68-75)	LFTVGTVA	CCCEEEEE	73T, 75A	0	50	50	Neutral	1.5	67.694	1000	724.57	170
	P2	3B8D_A(263-270)	AVTGVTFI	CCCEEEEC	268T	0	50	50	Neutral	1.5	80.077	1012	734.76	76
17	P1	2BKX_A(202-209)	KAEAVRKL	HHHHHHHH	No	50	25	25	Basic	3.6	66.488	1204	797.6	44
	P2	2BZ1_A(3-10)	LKRVAEAK	EEEEEEEE	No	50	25	25	Basic	3.6	47.348	1173	873.21	260
18	P1	2CHH_A(4-11)	GVFTLPAN	CEECCCCC	No	0	38	62	Neutral	2.2	99.601	1000	716.51	97
	P2	2Q2R_A(21-28)	NAPLTFVG	CCCEEEEE	No	0	38	62	Neutral	2.2	99.613	1010	712.11	154
19	P1	2ISB_A(111-118)	EEVVEAMR	HHHHHHHC	No	47	20	33	Acidic	2.9	93.375	1092	668.75	39
	P2	2JDJ_A(85-92)	RMAEVVEE	HHEEEEE	No	47	20	33	Acidic	2.9	99.661	1002	773.7	117
20	P1	2NY1_A(253-260)	PVVSSQLL	CCCCCEE	257S, 258Q	0	50	50	Neutral	4.1	99.461	1087	756.28	14
	P2	2A40_B(220-227)	LLQSSVVP	HHHHCCCC	No	0	50	50	Neutral	4.1	99.342	1076	688.67	39
21	P1	2ZBL_A(14-21)	EQETDRIF	HHHHHHHH	No	21	14	64	Acidic	1.1	82.528	1230	707.03	39
	P2	1M0W_A(271-278)	FIRDTEQE	ECCCCCEE	No	21	14	64	Acidic	1.1	84.776	1221	736.33	133
22	P1	1R6X_A(20-27)	KNELLSEA	HHHHHHHH	No	29	29	43	Acidic	1.6	77.567	1129	719.77	110
	P2	1ED1_A(47-54)	AESLLENK	CHHHHHCH	No	29	29	43	Acidic	1.6	80.533	1117	740.55	163
23	P1	2E1Y_A(171-178)	KMEAVAAG	HHHHHHCC	No	25	25	50	Neutral	0.4	93.512	941	619.3	125
	P2	1TWD_A(232-239)	GAAVAEMK	HHHHHHHH	No	25	25	50	Neutral	0.4	91.508	805	545.36	30

Note: P1 and P2 refer to the octapeptide and its reverse octapeptide, respectively. Hphil, Hpho and others column indicates the occurrence (in percentage) of hydrophilic residues, hydrophobic residues and other residues in the octapeptide, respectively. No in Stabilizing residues column indicates that there are no stabilizing residues; if a stabilizing residue exists, number and stabilizing amino acid in a single letter code are indicated. CSM, continuous symmetry measure; RMSD, root mean square deviation.

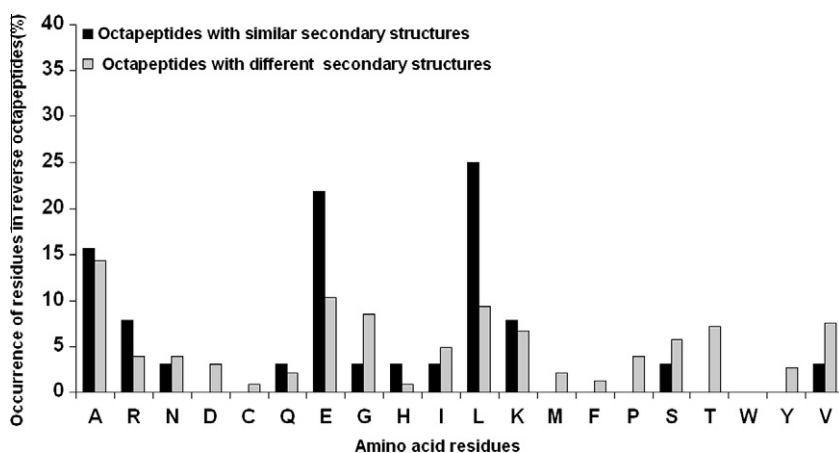


Figure 1 Occurrence of each amino acid residue in identical reverse octapeptide pairs with similar and different conformations

between the volumes of the octapeptides and their reverse with similar and different secondary structures is 0.93 and 0.91, respectively. Consequently, the correlation coefficient between surface areas of the identical reverse octapeptide pairs with similar secondary structure and that of the octapeptide pairs with different secondary structures is 0.95 and 0.73, respectively. However, the correlation coefficient between volume and surface area of unrelated octapeptide pairs with similar and different secondary structures seems to be very poor ( $-0.04$  and  $-0.03$ ) (Table S2). The difference in surface areas of unrelated octapeptide pairs implies that the octapeptide and its reverse octapeptide show similar geometric properties such as volume and surface area. Hence, from our results, we suggest that the octapeptides and their reverse have similar spatial arrangements like related octapeptide pairs whereas unrelated octapeptide pairs do not adopt similar volume and surface area. The  $C\alpha$  average root mean square deviation (RMSD) of identical reverse octapeptides is 2.27, which is 3.01 for unrelated octapeptide pairs. This implies that the  $C\alpha$  atoms of the unrelated octapeptides and identical reverse octapeptides may not be accurately superimposed on their pair, probably due to the influence of several other factors such as conformation angle preferences, surrounding residues, position of the octapeptide in the sequence, *etc.*

### Continuous symmetry of octapeptide pairs

Chirality and symmetry are treated as key descriptors of molecules. Continuous symmetry is a quantitative measure used to track the changes in molecules with similar composition during dynamic processes. The values of continuous symmetry measure (CSM) seem high for most of the octapeptide pairs with different secondary structures (Table 1) or with similar secondary structure (Table S1), which implies that the symmetry of the molecule is distorted. The difference between CSM of an octapeptide and its reverse varies between the different octapeptide pairs. Since molecular symmetry can explain many chemical properties, the higher values of CSM in both cases (octapeptide pairs with similar or different secondary structures) imply that the interpretation of physical and chemical properties with respect to their conformations of octapeptide pairs is problematic.

### Dipole and quadraploe moment of octapeptide pairs

The net charge and dipole moment of a protein may affect its binding properties and function [21] and hence we have computed net charge and dipole moment of the octapeptide and their reverse and correlated between them. From our analysis,

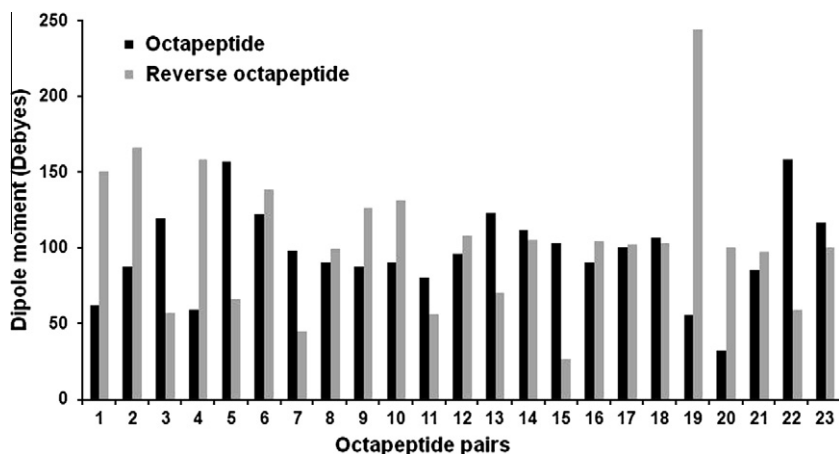


Figure 2 Difference in dipole moment of identical reverse octapeptides with different conformations

no obvious correlation was observed between net charges and dipole moments. The correlation coefficients between the dipole moment of octapeptides and their reverse were  $-0.75$  and  $-0.52$  for those with similar secondary structure and with different secondary structures, respectively. The difference in dipole moments of the identical reverse octapeptides with different conformations is shown in **Figure 2**. The results reveal the importance of dipole moment in adopting different conformations of octapeptide pairs, even though they possess similar residue composition, hydrophobicity, hydrophilicity and volume. A similar trend is observed for the quadrupole moment of the octapeptides and their reverse. These analyses suggested that the dipole and quadrupole moment may play a vital role in protein fragments involved in various binding processes.

### Role of stabilizing residues in reverse octapeptides

It is important to know how proteins are stabilized by various residues and their interactions. In order to understand the role of stabilizing residues in identical reverse octapeptides, we have predicted the stabilizing residues by using the SRide server [22]. The results reveal that residues present in octapeptide pairs with similar secondary structure did not contribute to the stability of a protein. For octapeptide pairs with different secondary structures, four octapeptide pairs contained stabilizing residues and contributed to the stability of the protein (**Table 1**). Our results suggest that the octapeptide pairs with similar secondary structure may not contribute to stability whereas the identical reverse octapeptides with different secondary structures may have biological significance to the stability of the protein. Understanding the mechanism by which protein stability is achieved is a challenging task and may help in protein design experiments [23].

### GO terms of protein pairs

The description of proteins containing identical reverse octapeptides with similar and different secondary structures and their corresponding GO terms are given in **Tables S3** and **S4**, respectively. Most of the proteins containing reverse octapeptides are enzymes. Interestingly, 12 hydrolase classes of enzymes contain identical reverse octapeptides. Reverse

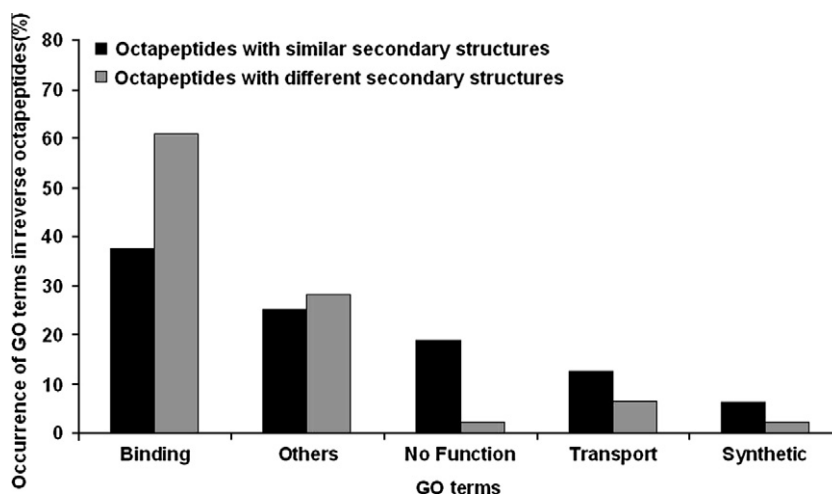
octapeptides are also observed in transferases, isomerases and toxins. From the GO terms extracted, in the case of octapeptide pairs with similar secondary structure, 37.5% of the proteins are involved in binding (DNA binding, metal ion binding, GTP and ATP binding). 18.75% of the proteins are hypothetical (**Figure 3**). One synthetic designed protein called GCN4 leucine zipper protein is observed, whose biophysical properties have been extensively studied [24].

In the case of octapeptide pairs with different secondary structures, the GO terms extracted contain 60.86% of proteins with the term binding, which include zinc ion binding, manganese binding, DNA/RNA binding, cation binding, metal ion binding, carbohydrate binding, cellulose binding, magnesium ion binding, protein binding, ATP binding and copper ion binding. 6.52% of the proteins have transferase phosphate containing groups. Also, they have two hypothetical and one synthetic protein (**Figure 3**). Among 46 proteins, 11 belong to the family of hydrolases and 6 belong to the family of transferases. These suggest that reverse octapeptides are mostly found in enzymes particularly in the hydrolase family.

### Discussion

In recent years, several methods based on similarity and dissimilarity comparison have advanced rapidly to understand the structure, function and property of proteins [25–28] paving the way to the design of novel proteins. For example, Alexander et al [29] designed two proteins with identical sequence apart from one residue mismatch and found that these two proteins adopt completely different folds and functions. Despite the progress, it still remains whether information gained from such studies is strong enough to predict the structure with 100% accuracy. Therefore, we believe that the inverse sequence similarity comparisons of protein or peptide sequences are also important to gain valuable insights into protein sequence-structure relationship.

Since an identical match of up to eight amino acids may not imply structural similarity [30], we performed an identical reverse octapeptide search to know whether inverse octapeptide sequences were present in unrelated proteins. Although these kinds of data analyses on oligopeptides have been actively pursued previously [31,32], no previous reports were found in the



**Figure 3** Distribution of GO terms of protein containing identical reverse octapeptides with similar and different secondary structures

literature regarding reverse peptide sequence search. This is the first analysis on polypeptides. Our analysis examined recently-increased datasets of experimentally-determined high resolution PDB sequences and structures (3124 unrelated proteins) and reveals the following observations. We find very low occurrences of octapeptides and their reverse in many PDB sequences. Due to the directionality of a protein molecule along the sequence, a protein with the reversed sequence is not something with opposite nature. Thus, it is not surprising that the pairs of certain octapeptides and their reverse octapeptides sometimes have similar secondary structure.

The amino acid composition of octapeptides and their reverse octapeptides reveals a relatively high proportion of L and E and null occurrence of W and Y. Also, the high proportion of hydrophilic residues in octapeptide pairs implies that these residues may form few contacts and contribute to binding processes rather than contributing to the stability of the protein. Therefore, our analysis indicates that the low abundance of identical reverse octapeptides in unrelated proteins is due to amino acid composition.

From the volume and surface area computations, the identical reverse octapeptide pairs have similar volume and surface area whereas unrelated octapeptide pairs do not. Also, the higher continuous symmetry measure of identical reverse octapeptides indicates a high distortion of the symmetry of molecules. From the analysis of GO terms, we found that the proteins containing octapeptide pairs are mostly enzymes and involved mainly in binding and transport processes. The physical properties such as dipole moment and charge are important factors, which can be related to the protein folding laws and its function. Differences in surfaces of protein fragments and differences in dipole and quadrupole moments are the major reasons for the octapeptide pairs adopting different structures and functions.

Although, a systematic search for the reverse form of the overlapping octapeptide segments is an interesting topic, only 8 octapeptide pairs (octapeptide and reverse octapeptide) with similar secondary structures and 23 octapeptide pairs with different secondary structures were identified in 3124 PDB sequences. Although the number of octapeptide pairs seems too low to provide insight into the sequence-structure relationship of proteins, the current study indicates that the proteins containing identical reverse octapeptides are not very abundant and they do not display specific biophysical and geometric properties as compared to unrelated octapeptides.

## Materials and methods

### Dataset

In our previous work [15], we compiled 3124 non-redundant protein sequences and their corresponding three state secondary structure assignments from PDB [16] to perform various computational studies. The protein sequences in the dataset were reversed by using a string reverse algorithm with an in house PERL program. Overlapping octapeptide segments of each of the reversed protein sequences were searched against 3124 non-redundant PDB sequences. The search process results in a reversed form of the octapeptide which occurs in unrelated proteins. The octapeptide pairs (octapeptide and its reverse octapeptide) thus obtained were grouped as octapeptide pairs with similar secondary structures and octapeptide pairs with

different secondary structures. The percentage of occurrences of amino acid residues in each of the groups was also computed. Additionally, in order to know whether reverse octapeptides having different secondary structures display specific properties compared to unrelated octapeptides, we created a control group of unrelated octapeptides by randomly shuffling 23 pairs of identical reverse octapeptides with different conformations for a comparative study.

### Computation of percentage of hydrophobic, hydrophilic and charged residues

The percentage of hydrophobic (F, I, L, M, V, W, A and P), hydrophilic (D, E, R, K and H) and other residues (G, S, T, C, N, Q and Y) was computed for the octapeptide pairs adopting similar and different secondary structures. We have also computed net charge and attributes (acidic, basic and neutral) of octapeptide pairs. To perform the aforementioned computations, we used a peptide property calculator web server, which is freely available at [https://www.genscript.com/ssl-bin/site2/peptide\\_calculation.cgi](https://www.genscript.com/ssl-bin/site2/peptide_calculation.cgi). The structures of identical reverse octapeptide pairs and unrelated octapeptides were superimposed on each other to obtain the RMSD value using the superimpose server [33].

### Computation of volume and surface area of identical reverse octapeptides

To find the difference in physical properties of octapeptide pairs (octapeptide and its reverse), we computed volume and surface area using the structure analysis web servers of the Helix Systems at Scientific Super Computing at the National Institute of Health accessible at <http://helixweb.nih.gov/structbio/basic.html>.

### Computation of continuous symmetry measure

To understand how a change of symmetry affects the chemical and physical properties of a protein fragment, we computed the CSM by using a web server accessible at <http://www.csm.huji.ac.il/new/?cmd=symmetry>. The measure is based on the distance between the investigated molecule and the nearest structure that has the desired symmetry. The value of this measure ranges from 0 to 100 and higher CSM values indicate the change in the physical shape of octapeptide fragments [34].

### Computation of dipole and quadrupole moment

Dipole and quadrupole moments of molecules might play a significant role in affecting the properties and activities of proteins. Hence, we have computed dipole and quadrupole moments of octapeptide fragment pairs by a web server located at the Weizmann institute, Israel [21].

### Prediction of stabilizing residues in identical reverse octapeptides

In order to identify whether the residues in an octapeptide and its reverse octapeptide play a vital role in stabilizing the protein molecule, we predicted the stabilizing residues using the SRide server [22]. This server predicts stabilizing residues using several structure-based parameters such as surrounding hydrophobicity [35], long-range order [36], conservation score, *etc.*

## Extraction of GO terms

In order to find what type of proteins or enzymes possesses identical reverse octapeptide, GO terms of the proteins containing octapeptide pairs (octapeptide and its reverse) were extracted from PDB annotations and stored in appropriate format. The extraction process was performed in a completely-automatic manner using in house java scripts on SUN ULTRA 40M2 workstation.

## Authors' contributions

SS and KMS designed the research. KMS performed computation. SS and KMS analyzed data. KMS wrote the manuscript. Both authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Acknowledgements

KMS thanks the Council of Scientific and Industrial Research (CSIR) for the award of Senior Research Fellowship (Grant No. 09/475(0157)/2010-EMR-I).

## Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.gpb.2012.11.005>.

## References

- [1] Lorenzen S, Gille C, Preissner R, Frommel C. Inverse sequence similarity of proteins does not imply structural similarity. *FEBS Lett* 2003;545:105–9.
- [2] Olszewski KA, Kolinski A, Skolnick J. Does a backwardly read protein sequence have a unique native state. *Protein Eng* 1996;9:5–14.
- [3] Chou KC, Zhang CT, Elrod DW. Do “antisense proteins” exist? *J Protein Chem* 1996;15:59–61.
- [4] Preibner R, Goede A, Michalski E, Frommel C. Inverse sequence similarity in proteins and its relation to the three dimensional fold. *FEBS Lett* 1997;414:425–9.
- [5] Nebel JC, Godfrey WC. Why inverse proteins are relatively abundant. *Protein Pept Lett* 2010;17:854–60.
- [6] Chorev M, Goodman M. Recent developments in retro peptides and proteins-an ongoing topochemical exploration. *Trends Biotechnol* 1995;13:438–45.
- [7] Rai J. Retroinverso mimetics of S-peptide. *Chem Biol Drug Des* 2007;70:552–6.
- [8] Gupta Sharma P. Reversal of peptide backbone direction may result in the mirror imaging of protein structure. *FEBS Lett* 1992;310:205–10.
- [9] Milton RC, Milton SC, Kent SB. Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* 1992;256:1445–8.
- [10] McDonnell JM, Fushman D, Cahill SM, Sutton BJ, Cowburn D. Solution structures of FceRI  $\alpha$ -chain mimics: a  $\beta$ -hairpin peptide and its retro-enantiomers. *J Am Chem Soc* 1997;119:5321–8.
- [11] Pan PK, Zheng ZF, Lyu PC, Huang PC. Why reversing the sequence of the  $\alpha$ -domain of human metallothionein-2 does not change its metal-binding and folding characteristics. *Eur J Biochem* 1999;266:33–9.
- [12] Saravanan KM, Selvaraj S. Analysis and visualization of long range contact networks in homologous family of proteins. *Open Struct Biol J* 2009;3:104–25.
- [13] Saravanan KM, Harihar B, Saranya N, Selvaraj S. Sequence and structural analysis of two designed proteins adopting 88% sequence identity. *Protein Eng Des Sel* 2010;23:911–8.
- [14] Itzhaki LS, Neira JL, Ruiz-Sanz J, De Prat Gay G, Fersht AR. Search for nucleation sites in smaller fragments of chymotrypsin inhibitor-2. *J Mol Biol* 1995;254:289–304.
- [15] Saravanan KM, Selvaraj S. Search for identical octapeptides in unrelated proteins: structural plasticity revisited. *Biopolymers* 2012;98:11–26.
- [16] Henrick K, Feng Z, Bluhm WF, Dimitropoulos D, Doreleijers JF, Dutta S, et al. Remediation of the protein data bank archive. *Nucleic Acids Res* 2008;36:D426–33.
- [17] Sapolsky R. *Biology and human behavior: the neurological origins of individuality*. 2nd ed. The Teaching Company; 2005.
- [18] Pallaghy PK, Melnikova AP, Jimenez EC, Olivera BM, Norton RS. Solution structure of contryphan-R, a naturally occurring disulphide bridged octapeptide containing d-tryptophan: comparison with protein loops. *Biochemistry* 1999;38:11553–9.
- [19] Bogan AA, Thorn KS. Analysis of hot spots in protein interfaces. *J Mol Biol* 1998;280:1–9.
- [20] Clackson T, Wells JA. A hot spot of binding energy in a hormone receptor interface. *Science* 1995;267:383–6.
- [21] Felder CE, Prilusky J, Silman I, Sussman JL. A server and database for dipole moments of proteins. *Nucleic Acids Res* 2007;35:W512–21.
- [22] Magyar C, Gromiha MM, Pujadas G, Tusnady GE, Simon I. Sride: a server for identifying stabilizing residues in proteins. *Nucleic Acids Res* 2005;33:W303–5.
- [23] Saraboji K, Gromiha MM, Ponnuswamy MN. Relative importance of secondary structure and solvent accessibility to the stability of protein mutants: a case study with amino acid properties and energetics on T4 and human lysozymes. *Comput Biol Chem* 2005;29:25–36.
- [24] Mittl PR, Deillon C, Sargent D, Liu N, Klauser S, Thomas RM, et al. The retro-GCN4 leucine zipper sequence forms a stable three dimensional structure. *Proc Natl Acad Sci U S A* 2000;97:2562–6.
- [25] Uthayakumar M, Sanjeev P, Nagarajan R, Sekar K. Sequence-structure similarity: do sequentially identical peptide fragments have similar three dimensional structures? *Curr Bioinformatics* 2012;7:111–5.
- [26] Saravanan KM, Selvaraj S. Performance of secondary structure prediction methods on proteins containing conformationally ambivalent sequence fragments. *Biopolymers Pept Sci* 2012. <http://dx.doi.org/10.1002/bip.22178>.
- [27] Ravella D, Kumar MU, Sherlin D, Shankar M, Vaishnavi MK, Sekar K. SMS-2.0: an updated database to study the structural plasticity of short peptide fragments in non redundant proteins. *Genomics Proteomics Bioinformatics* 2012;10:44–50.
- [28] Uthayakumar M, Benazir B, Sanjeev P, Kirti Vaishnavi M, Gurusaran M, Sureka K, et al. Homopeptide repeats: implications in protein structure, function and evolution. *Genomics Proteomics Bioinformatics* 2012;10:217–25.
- [29] Alexander PA, He Y, Chen Y, Orban J, Bryan PN. A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci U S A* 2009;106:21149–54.

- [30] Sudarsanam S. Structural diversity of sequentially identical subsequences of proteins: identical octapeptides can have different conformations. *Proteins* 1998;15:228–31.
- [31] McCaldon P, Argos P. Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* 1988;4:99–122.
- [32] Argos P. An Investigation of Oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J Mol Biol* 1990;211:943–8.
- [33] Bauer RA, Bourne PE, Formella A, Frommel C, Gille C, Goede A, et al. Superimpose: a 3D structural superposition server. *Nucleic Acids Res* 2008;36:W47–54.
- [34] Zabrodsky H, Peleg S, Avnir D. Continuous symmetry measures. *J Am Chem Soc* 1992;114:7843–51.
- [35] Manavalan P, Ponnuswamy PK. Hydrophobic character of amino acid residues in globular proteins. *Nature* 1978;275:673–4.
- [36] Gromiha MM, Selvaraj S. Comparison between long range interactions and contact order in determining the folding rates of two state proteins: application of long range order to folding rate prediction. *J Mol Biol* 2001;310:27–32.