JPP JOURNAL OF PROTEINS AND PROTEOMICS
5(1), January-April 2014, pp. 49-54

Research Article

SCALE FOR CONSTRUCTIVE AGGREGATION

Sujitha Mary, V. Alaguraj and S. Krishnaswamy*

School of Biotechnology, Madurai Kamaraj University, Madurai 625021, Tamil Nadu, India

Abstract: Aggregation is an inherent property of proteins. Both ordered and disordered proteins have a tendency to aggregate. Protein folding itself starts from the partially folded intermediates. The formation of native structures from these intermediates may be called as constructive aggregation. We describe the design of an intrinsic aggregation scale and its efficiency in finding hot-spots for constructive aggregation. In this paper, we are proposing a new aspect of aggregation, wherein aggregation can be considered a constructive phenomenon in proteins.

Keywords: constructive aggregation; intrinsic aggregation scale; protein folding

Introduction

Protein molecules have a specific function within the cells. They interact among themselves or with other molecules to perform their biochemical functions. It is said that except in intrinsically disordered proteins, structure determines the function. To carry out their function, a protein needs to fold into complex three dimensional structures. Biological systems have evolved various quality control mechanisms for the proper folding of the proteins and for the degradation of the unfolded proteins. The failure of these mechanisms leads to misfolding and abnormal deposition of misfolded proteins. This self association of proteins can be called as protein aggregation (Merlini *et al.*, 2001). Beta sheet structures, one of the most common structures of a folded protein, have a tendency to favour amyloid fibril formation. Amyloid fibrils consist of beta aggregates which cause diseases such as Alzheimer's, Parkinson's, type II diabetes etc. These diverse diseases are termed as conformational diseases as they are caused due to the conformational changes in proteins (Merilini *et al.*, 2001). Thus the study of protein aggregation is of fundamental medical interest as these diseases affect the brain in a devastating way. This field of study gained a wider perspective when it was found that the proteins which are not involved in diseases also aggregate. The characteristic feature of amyloidogenic proteins is structural instability which is caused due to mutations, post translational modifications or other environmental conditions such as pH, temperature etc.

Biological systems have evolved several quality control mechanisms to ensure that proteins fold correctly. Despite these controls many proteins get converted into insoluble fibrils. The study of these structural transitions revealed the existence of intermediates prior to protein misfolding. Some studies have even shown a clear kinetic lag phase before the fibril formation. This behavior shows some similarity with crystal growth and polymer gelation, where there is a need for small aggregates from which larger ones are grown (Harper and Lansbury, 1997; Horwich

Corresponding Author: **S. Krishnaswamy**
E-mail: mkukrishna@gmail.com

Received: January 25, 2014
Accepted: March 28, 2014
Published: April 30, 2014

et al., 1997).The ability to form amyloids is a generic property of proteins. The intermolecular bonds that stabilize the amyloids involve peptide backbones which are common to all proteins (Dobson, 1999). Both folding and aggregation are determined by the physico-chemical properties of amino acids such as pH, temperature, hydrophobicity, aromaticity etc. Proteins find their stable structure by trial and error. The structural properties of intermediates provide important evidence about the folding of the proteins. Normal aggregation is said to be the conversion of partially folded structure to a misfolded structure. Our suggestion is that the process of conversion of partially folded intermediates into normal folded structure itself can be called as a type of aggregation, namely, constructive aggregation. This constructive aggregation region can be a nucleation event in a folding process or it can be a site for protein-protein interactions. The formation of well defined aggregates has many similarities with the crystal growth procedure in such a way that aggregation procedure needs nuclei or seeds. The larger molecular assemblies are grown from small aggregates (Dobson, 1999). So we can consider aggregation and structure formation as same sides of a coin rather than considering them as two different sides of the same coin.

As there are limitations in getting the atomic-level structures of amyloid fibrils, computational methods play an important role in understanding protein aggregation. The prediction of aggregation-prone regions helps to prevent the uncontrolled aggregation and also to understand the process in detail. Based on the physicochemical properties of the amino acids that influence aggregation, several groups have developed aggregation prediction algorithms. However, none of them have looked at the aspect of constructive or intrinsic aggregation. Here, we discuss the design of a propensity scale for constructive aggregation which is termed as Intrinsic Aggregation Scale (IAS).

Materials and Methods

Design of Intrinsic Aggregation Scale

The protein structure coordinate files of monomers and oligomers were downloaded from the structures deposited in PDB as of July 2013 (*www.rcsb.org*). The search condition used was X-ray structures with resolution greater than 2Å, containing only proteins with less than 30% identity. The proteins with missing residues were removed from the dataset. The PDB ids of the final dataset and the fold are given in Table 3 and Table 4. The solvent accessibility values were found using DSSP program (Kabsch and Sander, 1983). In order to find the interacting residues the spatial distance between the atoms were calculated using python scripts. The residues for which Calpha – Calpha distance is less than 8Å were identified. The core residues were defined as the residues that are buried in monomeric structures and are at a distance less than 8Å. The residues having relative solvent accessibility less than 20% are considered as buried. The interface residues were identified employing the method used by Tuncbag *et al.* (2009).

ie. $RASA_i$=((ASA_imonomer-ASA_icomplex)/ ASA_imax)*100

if $RASA_i$ > 20% then that residue is considered as interface residue (Tuncbag *et al.*, 2009).

The probability of interface residues and core residues for the proteins in the dataset were calculated and their products were taken. This is defined as IAindex. The frequency of interface residue of type *i*,

$$\text{Finter}_i = N_i / T_i \qquad \text{eq.1}$$

where N_i is the total number of residue *i* in the interface and T_i is the total number of residue *i* in the protein.

Ftotal = Total no: of interface residues/Total no: of residues in the Protein

The probability of residues in interface region Pint is

$$\text{Pint}_i = \text{Finter}_i / \text{Ftotal}_i \qquad \text{eq.2}$$

The probability of residues in core region $Pcore_i$ is also calculated using eq. 1 and eq. 2

$$\text{IAindex}_i = \text{Pint}_i * \text{Pcore}_i \qquad 3$$

The values for IAindex for 20 residues are given in Table 1.

According to Manavalan and Ponnusamy (1978) bulk hydrophobic character obtained from

Table 1
IAindex of 20 amino acids

Residue	*IAindex*
A	0.348
C	0.663
D	0.309
E	0.098
F	1.082
G	0.477
H	0.718
I	0.790
K	0.055
L	0.810
M	0.880
N	0.290
P	0.290
Q	0.188
R	0.417
S	0.256
T	0.295
V	0.802
W	1.499
Y	0.651

the surrounding hydrophobicity reflects the hydrophobic index of a protein better than all other hydrophobic indices.

Surrounding hydrophobicity is sum of hydrophobic indices assigned to all the residues within 8Å distance (Manavalan and Ponnusamy, 1978). Similarly, surrounding aggregation is defined which will better reflect the aggregation environment of a protein. Surrounding aggregation of a residue is the sum of IAindex of residues within 8Å distance.

$$\mathrm{Sagg_j} = \Sigma N_{ij} I_i$$

where N_{ij} is the total number of surrounding residues of ith type associated with jth residue. I_i is the IAindex of ith residue. All Sagg values for the same residue are calculated for each protein and their average is taken as surrounding aggregation for that residue.

When the surrounding aggregation was plotted against average number of contacts (Figure1), it showed that when the number of contacts increases surrounding aggregation value also increases. The Sagg values were normalized as Intrinsic Aggregation Scale (IAS). The values are shown in Table 2.

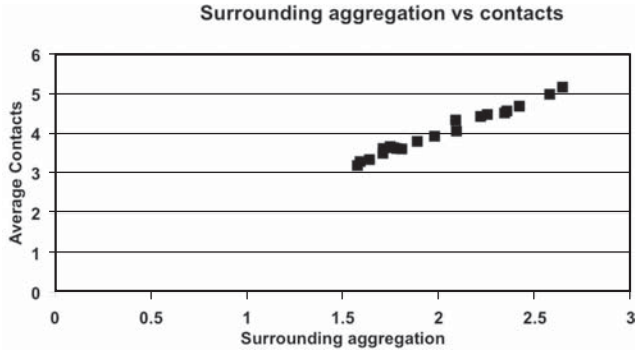


Figure 1 : **Surrounding aggregation plotted against the average number of contacts**

Table 2
IAS values for 20 amino acids

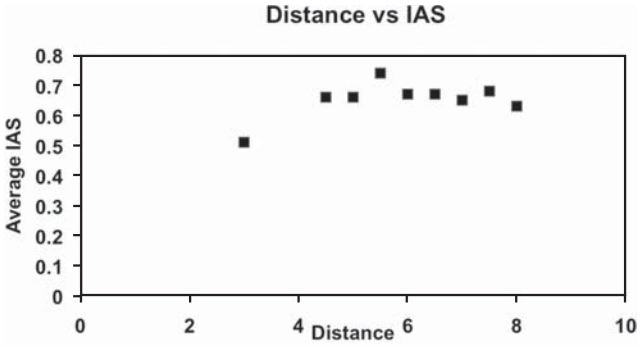
Residue	*IAS*
A	0.111
C	0.272
D	-0.199
E	-0.342
F	0.237
G	-0.234
H	-0.093
I	0.439
K	-0.391
L	0.597
M	0.362
N	-0.271
P	-0.003
Q	-0.405
R	-0.174
S	-0.272
T	-0.216
V	0.374
W	0.664
Y	0.107

Results and Discussion

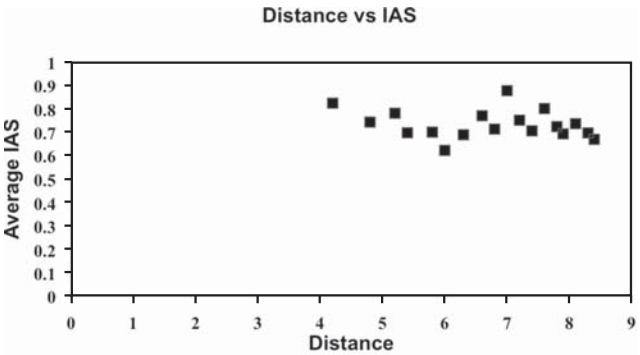
Analyzing protein structures using IAS

The protein structures taken from the PDB were analysed using IAS to find the hot-spots for constructive aggregation. The sum of intrinsic aggregation value of interacting residues was

calculated and was compared with their spatial distance to know whether there is any change in the IAS with change in distance. Two residues whose distance is less than 8Å and located either in interface region or in core region were considered as interacting pairs. When the sum of IAS values of the interacting residue pair is greater than 0.6, there is a change in the trend is evident from the graphical representation of the comparison between spatial distance and sum IAS shown in Figure 2. So 0.6 was considered as a threshold for constructive aggregation. If the sum of IAS values of a interacting residue pair in the core or interface region of a protein is greater than 0.6 then they are said to be a hot spot for constructive aggregation. The secondary structure elements in these regions were found using DSSP program. The frequency of strand and helical structures were high in these regions (Figure 3). So these hot spots for constructive aggregation may have nucleation sites for protein folding. We



(a)



(b)

Figure 2 : **Plot showing spatial distance of residue pairs and average IAS sum for (a) monomeric proteins and (b) oligomeric proteins.**

The IAS sum cut-off for constructive aggregation is taken as 0.6 as there is a shift in the nature of relation between distance and IASsum at 0.6

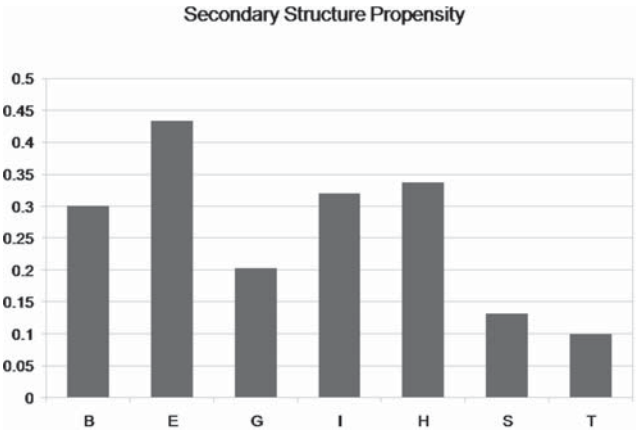


Figure 3: **Secondary structure propensity in constructive aggregating regions (B- Bridge; E- Beta ladder & Beta Sheet; G- 3/10 Helix; I-pi Helix; H-Alpha Helix; S-Bend; T-Turn)**

selected the proteins chymotrypsin inhibitor II, ubiquitin, lysozyme as their nucleation sites were already known. Their constructive aggregation regions were determined using intrinsic aggregation scale.

The folding nucleus of chymotrypsin inhibitor II was 35A, 68L, 39I, 70V, 76I. Lysozyme had nucleation sites at 8L, 29V, 32A, 123T. Ubiquitin protein had its folding nucleus at 3I, 5V, 15L, 17V, 30I, 44I (Poupon and Mornon, 1999; Shaknowich *et al.*, 1996). The experimentally proven folding nucleus in these three proteins was observed in the constructive aggregation hotspots. When the constructive aggregating sites of these proteins were examined it was found that among the 50% of aggregating pairs at least one of the residues in the pair was a nucleating residue. The residues which are both intrinsically aggregating residues as well as nucleating residues in chymotrypsin inhibitor II, lysozyme and ubiquitin is shown in Figure 4.The statistics of the predicted hotspots for constructive aggregation are shown in Figure 5. The presence of nucleating residues in the constructive aggregating regions tells that these hotspots may initiate proteins folding. The constructive aggregation hotspots in these three structures are shown in Figure 6. These residues may also be involved in protein-protein interactions or inter protein interactions.

The oligomeric interfaces were examined for the hotspots for constructive aggregation. The pairing frequency of hotspots residues in the interface region is shown in Figure 6. In interface

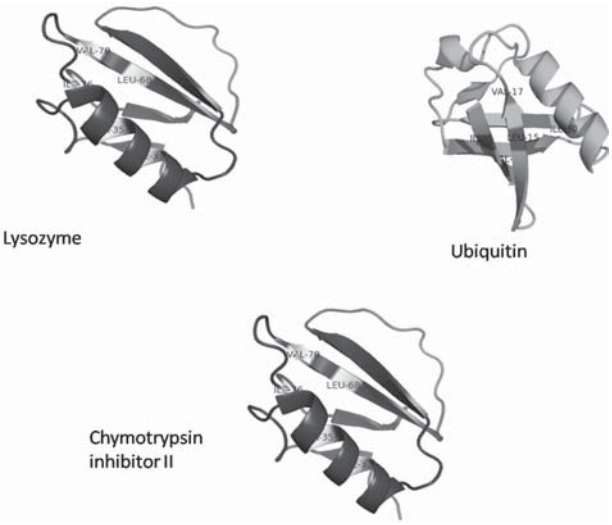


Figure 4: **The hot spots for constructive aggregation with nucleation sites. The predicted IA hot spots which are experimentally proven folding nucleus in lysozyme, ubiquitin and chymotrypsin inhibitor II.**

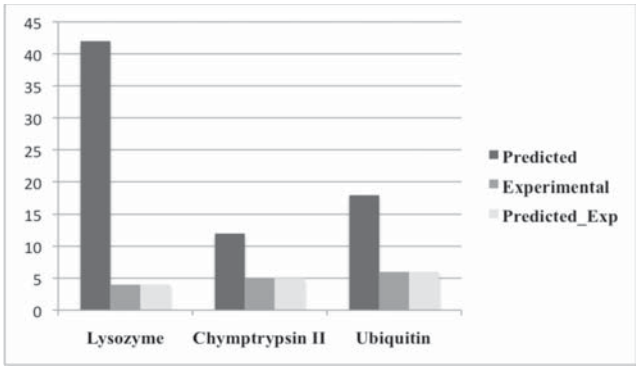


Figure 5: **Statistics of hotspot residues (Predicted-The no. of residues which are predicted as hot spots of constructive aggregation; Experimental- The no. of folding nucleus residues; Predicted_Exp – The no: of folding nucleus residues which are observed in the constructive aggregattion hotspots)**

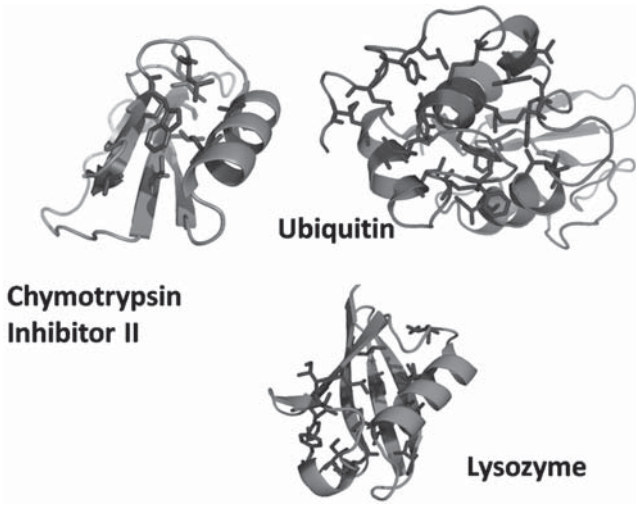


Figure 6: **Constructive aggregating hotspots shown in stick model**

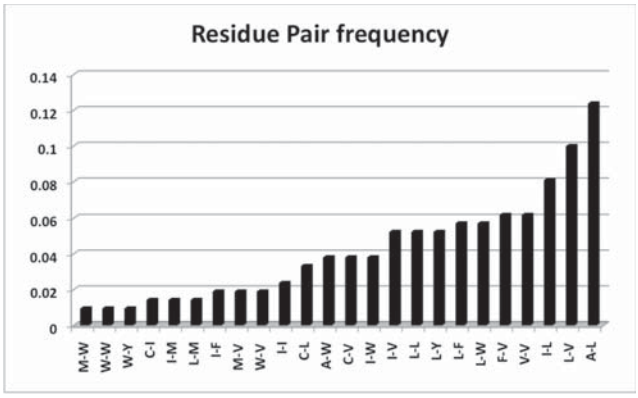


Figure 7: **Residue frequency of IA pairs in interface (x-axis denotes the residue pair- names of the residues are given as single letter amino acid code. y-axis is the frequency of occurrence of that particular residue type)**

Table 3
PDB ids of structures in the dataset

monomers
1BXO, 1C75, 1G4I, 1GVT, 1J0P, 1KWF, 1L9L, 1LUG, 1OK0, 1PJX, 1PQ5, 1PQ7, 1TQG, 1UCS, 1X8P, 1X8Q, 1XMK, 1XVO, 1YWA, 1YWB, 1ZLB, 2AYW, 2FOU, 2OV0, 3C78
oligomers
1CPC, 1DOW, 1EPT, 1FCC, 1IRD, 1J34, 1JLT, 1KTP, 1PBY, 1PHN, 1UW4, 2BL0, 2F4M, 2PU9

Table 4
List of SCOP domains in the dataset

Cytochrome_C
Acid proteases
Phospholipase A2
Multiheme cytochrome_C
Alpha/alpha toroid
Saposin-like
Carbonic anhydrase
Alpha-amylase inhibitor tendamistat
6-bladed beta propeller
7-bladed beta propeller
Trypsin like serine proteases
Four helical up & down bundle
Beta clip
Lipocalins
DNA/RNA binding 3-helical bundle
Cupredoxin-like
Globin-lie
GLA-domain
c-Type lectin like
Streptavadin
Immunoglobulin-like beta sandwich
Alpha-alpha super helix
Ferrodoxin-like
XPC binding domain
Cysteine proteinases

region the pairing of hydrophobic residues were more in the constructive aggregating hotspot. These residues can take part in hydrophobic interactions which plays an important role in defining homo-oligomeric interfaces (Ali and Imperiali, 2005).

Conclusion

Protein aggregation is always considered as a hindrance to protein folding. The partially folded structures can either get folded to native structure or get misfolded which leads to aggregation. Our suggestion is that the process of structure formation itself involves an aspect of aggregation, which we term as constructive or intrinsic aggregation. The Intrinsic Aggregation Scale which we designed will help in finding the hotspots for constructive aggregation in other proteins.

Acknowledgement

We thank Dr. Daniel Otzen, Aarhus University, Denmark for discussions. We acknowledge Department of Biotechnology for Centre of Excellence in Bioinformatics facilities and funding.

Abbreviations

IAindex, Intrinsic Aggregation index; IAS, Intrinsic Aggregation Scale

References

Ali, M.H., and Imperiali, B. (2005). Protein oligomerization: How and Why? Bioorg. Med.Chem 13, 5013-5020.

Dobson C.M. (1999). Protein misfolding, evolution and disease. Trends Biochem. Sci. 24, 329-332.

Harper, J.D., and Lansbury, P.T. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time dependent solubility of amyloid proteins. Annu. Rev. Biochem. 66, 385-407.

Horwich, A. L., and Weissman, J. S. (1997). Deadly conformations - Protein misfolding in prion disease. Cell 89, 499-510.

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577-2637.

Manavalan, P., and Ponnuswamy, P.K. (1978). Hydrophobic character of amino acid residues in globular proteins. Nature 275, 673–674.

Merlini, G., Bellotti, V., Andreola, A., Palladini, G., Obici, L., Casarini, S., and Perfetti, V. (2001). Protein aggregation. Clin. Chem. Lab Med 9, 1065.

Poupon, A., and Mornon, J.P. (1999) Predicting the protein folding nucleus from sequences (correction of a sequence). FEBS Lett. 452, 283-289 [Erratum in 457, 525].

Shakhnovich, E., Abkevich, V., and Ptitsyn, O. (1996). Conserved residues and the mechanism of protein folding. Nature 379, 96–98.

Tuncbag, N., Gursoy, A. and Keskin, O. (2009). Identification of computational hot spots in protein interfaces: Combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics 25, 1513-1520.