Supporting Information

The Burial of Solvent-Accessible Surface Area Is a Predictor of Polypeptide Folding and Misfolding as a Function of Chain Elongation

Neşe Kurt and Silvia Cavagnero*

Department of Chemistry, University of Wisconsin - Madison, Madison, Wisconsin 53706

* To whom correspondence should be addressed

E-mail: cavagnero@chem.wisc.edu

Methods

Calculation of fraction of nonpolar solvent accessible surface area (NSASA): The SurfaceRacer[1] program was used to calculate the solvent accessible surface area values for Figure 2. The data set of Richmond and Richards[2] was used for the van der Waals radii of the protein atoms. The probe water molecule was assigned a radius of 1.4 Å. Surface Racer provides values of polar, nonpolar and total surface area in $Å^2$ for any given protein conformation in PDB format. The fractional NSASA values of Figure 2 were calculated as the ratios of nonpolar surface to the total solvent accessible surface area. The 1ey0, 1bnr, 2ci2 and 1vxf PDB files were used for the *staphylococcal* nuclease (SNase), barnase (Bar), chymotrypsin inhibitor 2 (CI2) and apomyoglobin (apoMb) structures, respectively.[3] The coordinates for the first five residues of SNase are missing in the PDB file, as no electron density can be detected by x-ray crystallography for these residues. In the case of CI2, the first residue for which coordinates are available is amino acid 19. We have renumbered the CI2 residues starting from 1, in accordance with conventions widely accepted in the literature for this protein.[4] PDB files for extended chain conformations were generated with the InsightII software (Accelrys, Inc., San Diego, CA).

Calculation of fractional buried area per residue (FBA): The hydrophobicity scale of Rose[5] was used to calculate the overall nonpolar nature of the polypeptides at each given chain length. The scale is based on the fraction of surface area that becomes buried upon folding, relative to a standard area, defined as the total area in an extended chain conformation. The scores were derived from known protein structures in the database, and averaged for each amino acid type. We utilized these scores to calculate the average predicted fractional area buried upon folding for all polypeptides at each chain length. Calculations were carried out with the Excel software (Microsoft Corp., Redmond, WA). The results were then divided by the total number of amino acids at each

chain length, in order to enable a direct comparison between different chain lengths and different proteins. The resulting parameter, i.e., the fractional area buried upon folding per residue, was defined as FBA. The local maxima in the curves of Figure 3 are due to the presence of clustered hydrophobic residues along the amino acid sequence of each polypeptide. The large fluctuations observed at very short chain lengths are expected, and are caused by the large relative weight of each added amino acid for a small total number of residues.

The peptides denoted by stars in Figure 3 are known to aggregate heavily and form amyloid fibrils. These peptides are derived from human proteins and are associated with amyloidogenic diseases. They are Amyloid-β(1-40) and (1-42), prion protein PrP (106-126) and acetylcholinesterase AChE (586-599) (plotted in Figure 3 from right to left). The amino acid sequences of the peptides are listed below.

Amyloid-β(1-42): DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA

PrP peptide: TNMKHMAGAAAAGAVVGGLG

AChE peptide: AEFHRWSSYMVHWK.

We also calculated FBA scores for the following two amyloidogenic peptides that form disulfide bridges under native conditions. These peptides are derived from the human BRI protein (244-277) and human amylin (1-37).

ABri peptide: EASNCFAIRHFENKFAVETLICSRTVKKNIIEEN

Amylin peptide: KCNTATCATQRLANFLVRSSNNLGAILSSTNVGSNTY.

The FBA scores for all the above peptides are plotted in Figure S1 below as a function of chain length.

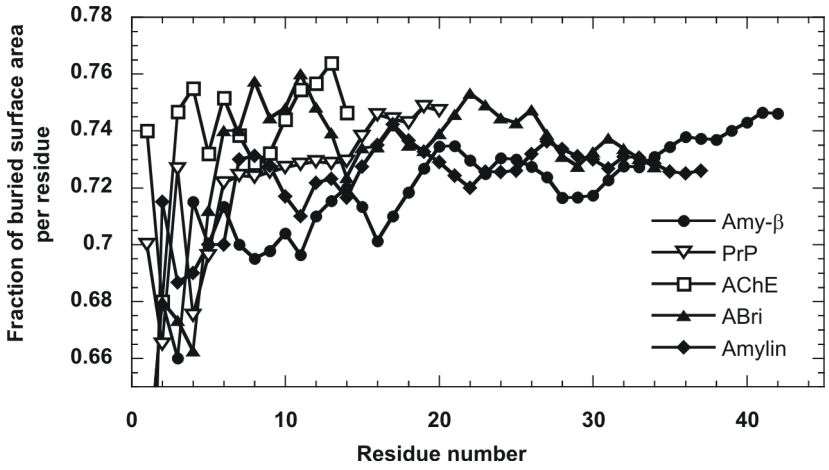

Figure S1. Fractional buried surface area per residue for peptides that heavily aggregate and form amyloid-like fibrils.

Additional Comments

Fraction of NSASA as a Function of Chain Length. One of the main findings of this communication is that the burial of nonpolar surface area is most effective at longer chain lengths, for a native-like topology. This finding is distinct from the natural expectation that nonpolar surface area would become exposed upon removal of surface residues in a folded protein. Rather than a progressive monotonic increase in the exposure of nonpolar area as residues are removed, we find a sharp variation in the slope of the fractional NSASA curves of Figure 2, at long chain lengths. This correlates with the experimentally observed behavior of truncated polypeptide chains, and it is relevant to the intrinsic conformational trends to be considered upon cotranslational protein folding.

Significance of FBA values. The comparative analysis of experimental findings and calculations presented in Figure 3 reveals that, at chain lengths bearing no significant driving forces for native-like structure formation, FBA values greater than c.a. 0.73 support misfolding and aggregation. This specific value should be regarded as a semi-quantitative predictor of chain behavior, rather

than a definitive limiting value. It is not possible to provide any statistically significant error bars at this stage given that, as mentioned in the communication, other forces than those encompassed by FBA calculations may affect polypeptide aggregation to some degree. In summary, the FBA results reported here provide a solid semi-quantitative prediction on the role of buried nonpolar surface area as a function of chain elongation, with no quotable associated statistical parameters.

Acknowledgement

We thank Dr. Qiang Cui for the use of the InsightII software.

References

(1) Tsodikov, O. V.; Record, M. T.; Sergeev, Y. V. *J. Comput. Chem.* 2002, *23*, 600-609.
(2) Richmond, T.; Richards, F. M. *J. Mol. Biol.* 1978, *119*, 537-555.
(3) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* 2000, *28*, 235-242.
(4) de Prat Gay, G.; Fersht, A. R. *Biochemistry* 1994, *33*, 7957-7963.
(5) Rose, G. D.; Geselowitz, A. R.; Lesser, G. J.; Lee, R. H.; Zehfus, M. H. *Science* 1985, *229*, 834-838.