

Anomalous scattering is a resonant effect and occurs when the energy of X-rays equals or exceeds the excitation energy of electrons in some atoms, so that X-rays are absorbed and then emitted by these anomalous scatterers in the form of fluorescence. In MAD, the phases are deduced from differences in the signals of anomalous scatterers at several X-ray wavelengths in the vicinity of the X-ray absorption edge of these atoms. MAD experiments can be done only at synchrotron beam lines, where the X-ray wavelength can be tuned to precise values at the selected absorption edge.

An alternative method that is gaining in popularity is single-wavelength anomalous diffraction (SAD). In SAD, it is enough to measure diffraction data with only one wavelength, not necessarily tuned to the absorption edge of the anomalous scatterers. Synchrotron radiation may be useful but is not a prerequisite for SAD phasing, and diffraction data can be collected at copper-anode X-ray sources found in university laboratories. SAD phasing requires fewer data than MAD and is methodologically simpler.

Iodine has been used as a heavy atom in the past, either by soaking crystals in KI<sub>3</sub><sup>4</sup> or N-iodosuccinimide<sup>5</sup>, which iodinate tyrosines, or by cryo-soaking crystals in iodide salts<sup>6</sup>. With these techniques, however, the level of iodine inclusion is often low. In addition, the iodine can perturb the protein structure or cause the crystals' diffraction properties to deteriorate.

The ingenious method of Xie *et al.*—a compelling application of the unnatural-amino-acid technology developed by this group<sup>7,8</sup>—circumvents these problems and allows convenient production of iodine-containing proteins for SAD phasing (Fig. 1). The unique pair tRNA<sup>Tyr</sup><sub>CUA</sub>-tyrosyl tRNA synthetase was imported from *Methanococcus jannaschii* to *Escherichia coli* and made specific for iodophenylalanine by iterative mutations and selection<sup>7</sup>. Xie *et al.* used this approach to generate a mutant bacteriophage T4 lysozyme in which a single phenylalanine residue was replaced by iodophenylalanine. The fidelity of iodophenylalanine incorporation was almost 100%.

The mutant lysozyme was crystallized, diffraction data were collected in-house and the structure was solved using SAD. One iodine atom among 164 residues resulted in a strong anomalous signal, about 3% of the total diffraction intensities, comparing favorably with the level routinely achieved with selenomethionine-containing crystals at synchrotron beam lines. The presence of iodophenylalanine in the hydrophobic core of lysozyme did not perturb the molecular structure.

Looking to the future, it is possible that the method of Xie *et al.* could be exploited in conjunction with the recently introduced method, radiation-damage induced phasing<sup>9</sup>. Irradiation of macromolecular crystals by intense X-ray beams causes certain specific chemical changes in crystalline samples, including radiolysis of relatively weak bonds, such as in disulfide bridges and in halide-substituted aromatic rings. Gradually vanishing heavy atoms (in analogy to the famous Cheshire cat) create isomorphous changes that can be used to evaluate protein phases. It may be expected that iodophenylalanine will behave analogously to bromouracil<sup>9</sup> and iodotyrosine<sup>10</sup>, so that the disappearance of

iodines would provide a useful additional phasing signal.

1. Xie, J., Wang, L., Brock, A., Spraggon, G. & Schultz, P.G. *Nat. Biotechnol.* **22**, 1297–1301 (2004).
2. Blundell, T.L. & Johnson, L.N. *Protein Crystallography* (Academic Press, New York, 1976).
3. Hendrickson, W.A. *J. Synchrotron Rad.* **6**, 845–851 (1997).
4. Sigler, P.B. *Biochemistry* **9**, 3609–3617 (1970).
5. Brzozowski, A.M. *et al. Nature* **351**, 491–494 (1991).
6. Dauter, Z., Dauter, M. & Rajashankar, K.R. *Acta Crystallogr.* **D56**, 232–237 (2000).
7. Wang, L., Brock, A., Herberich, B. & Schultz, P.G. *Science* **292**, 498–500 (2001).
8. Chin, J.W. *et al. Science* **301**, 964–967 (2003).
9. Ravelli, R.B.G. & McSweeney, S.M. *Structure* **8**, 315–328 (2000).
10. Zwart, P.H., Banumathi, S., Dauter, M. & Dauter, Z. *Acta Crystallogr.*, in press (2004).

## A universal TANGO?

Vijay S Pande

**A new algorithm can predict the propensity of proteins to aggregate.**

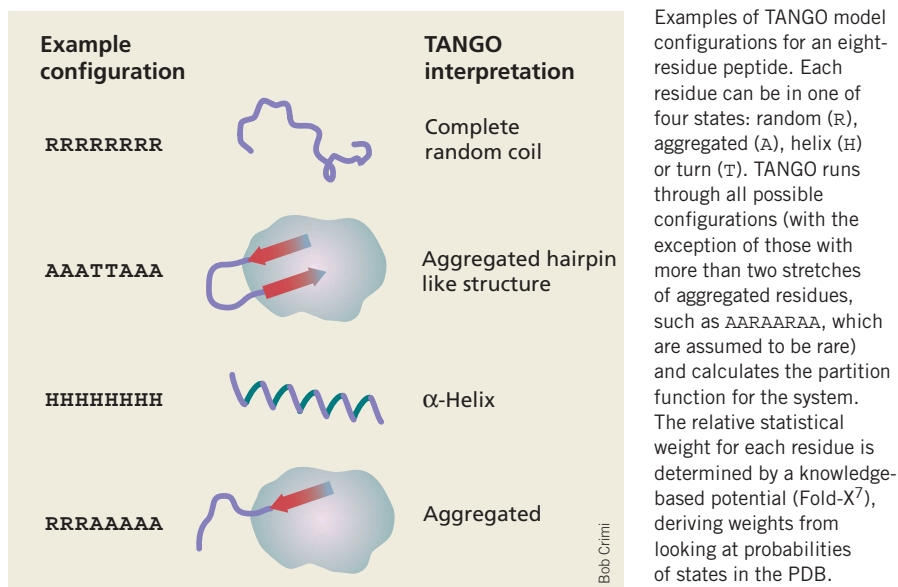
Many diseases have been linked to protein aggregation, including Creutzfeldt-Jakob, Alzheimer, Huntington and Parkinson diseases<sup>1</sup>. With the growing list of aggregation-related diseases, it is tempting to ask whether protein aggregation is a universal phenomenon. In other words, are there common aggregation-related aspects to these diseases? The strongest way to test a hypothesis is to use its predictive aspects; this approach also has the important and imminently practical implication of creating a technology with numerous diagnostic applications. The work of Serrano and coworkers<sup>2</sup> in this issue is a beautiful test of the hypothesis of the existence of universal aspects in protein aggregation-related diseases. By developing TANGO, a novel means for identifying  $\beta$ -sheet aggregation propensity, the authors have found that this propensity can be predicted. Additionally, they show that this propensity can predict which peptides and their mutants will aggregate in various amyloidogenic diseases.

Several other recent studies have provided tantalizing evidence that common aggregation-related characteristics exist among diseases. For example, Dobson and coworkers<sup>1</sup> have shown that many different proteins that are normally well-behaved can be induced to aggregate. Moreover, Glabe and coworkers<sup>3</sup>

have reported that an antibody raised against aggregates involved in Alzheimer disease (A $\beta$  oligomers) can also bind oligomers of peptides believed to be involved in several diseases, including Parkinson ( $\alpha$ -synuclein) and Huntington (poly-glutamine) diseases and type II diabetes (islet amyloid polypeptide), but does not bind to the monomers or fibrils of any of these peptides<sup>3</sup>. If an antibody recognizes this broad range of protein aggregates, it may indicate that protein aggregate structures are similar in some way.

How does TANGO work? It is informative to first look at a precursor of TANGO, AGADIR<sup>4</sup>, which addresses the helical propensities of peptides through a statistical mechanics perspective. AGADIR uses statistical mechanics to calculate the relative probability of finding a helical versus coiled state for a given peptide sequence, with just a few empirical parameters. The sequence-specific elements are critical, as different side chains will lead to different helical propensities. In a sense, TANGO can be thought of as an AGADIR for  $\beta$ -sheet aggregation. Using a strategy similar to AGADIR, TANGO models proteins using a set of four discrete states (unfolded, helix, turn, aggregated) per residue and calculates the partition function for this system based on a few empirically derived parameters (see Fig. 1). With four states, the possible number of configurations becomes computationally intractable with even relatively short proteins (a 20-mer peptide would have trillions of possible states). Therefore, a double stretch

Vijay S. Pande is in the Departments of Chemistry and Structural Biology, Stanford University, Stanford, California 94305, USA. e-mail: pande@stanford.edu



Examples of TANGO model configurations for an eight-residue peptide. Each residue can be in one of four states: random (R), aggregated (A), helix (H) or turn (T). TANGO runs through all possible configurations (with the exception of those with more than two stretches of aggregated residues, such as AARAARA, which are assumed to be rare) and calculates the partition function for the system. The relative statistical weight for each residue is determined by a knowledge-based potential (Fold-X<sup>7</sup>), deriving weights from looking at probabilities of states in the PDB.

approximation (see Fig. 1) is made, assuming that states with more than two ordered stretches would be rare. The key to TANGO is the choice of these four representative states and the ability to empirically obtain parameters from the Protein Data Bank (PDB).

Although TANGO uses a simplified representation, its caveats are not onerous. As the extent of aggregation will depend both on the concentration of the peptide and on its association constant, Serrano and coworkers report that TANGO can only give relative probabilities; that is, it allows quantitative comparison inside the same polypeptide chain, or with mutants of the polypeptide chain, but only qualitative comparison between different polypeptide chains. Furthermore, the authors acknowledge and present data that suggest the possibility that the double-stretch approximation may break down with relatively long (e.g., 50 residue) peptides. A natural follow-up to TANGO could relax the double-stretch assumption, potentially allowing one to address longer chains.

It is important to stress that TANGO is not solely a  $\beta$ -sheet version of AGADIR, but is rather a model for aggregation (not just  $\beta$ -sheet formation). The parameters used in the aggregated state incorporate burial of chains in addition to formation of  $\beta$ -sheet. It is for this reason that TANGO can predict a peptide's role in aggregation-related disease, whereas  $\beta$ -sheet propensities cannot. However, it should also be emphasized that although TANGO predicts aggregation, it is not a predictor of amyloid formation. Although aggregation may be related to the formation of amyloids, TANGO does not address the amyloid state.

Nevertheless, TANGO has been remarkably successful in predicting aggregation properties of peptides, especially those related to disease. Serrano and coworkers show that TANGO predicts the aggregation of a data set of 179 peptides from 21 different proteins as well as of a new set of 71 peptides derived from disease-related human proteins measured in their laboratory. TANGO also correctly predicts pathogenic as well as protective mutations of the Alzheimer  $\beta$ -peptide, human lysozyme and transthyretin.

The ability of TANGO to predict protein aggregation related to disease, but not amyloid properties, suggests a potentially important implication of this work. Although amyloids have been proposed to be causative in diseases, such as Alzheimer, recent work<sup>1,5,6</sup> suggests that the relevant states for neurotoxicity in this case are the oligomeric states and not amyloid fibrils. TANGO's ability to predict mutations relevant to Alzheimer disease strongly supports this interpretation. Finally, considering the implications of antibodies that appear to recognize a wide range of peptide aggregates<sup>3</sup>, perhaps the issues related to aggregation (captured by TANGO's elegant model) are indeed a universal property.

1. Bucciantini, M. *et al.* *Nature* **416**, 507–511 (2002).
2. Fernandez-Escamilla, A. *et al.* *Nat. Biotechnol.* **22**, 1302–1306 (2004).
3. Kaye, R. *et al.* *Science* **300**, 486–489 (2003).
4. Muñoz, V. & Serrano, L. *Nat. Struct. Biol.* **1**, 399–409 (1994).
5. Lambert, M.P. *et al.* *Proc. Natl. Acad. Sci. USA* **95**, 6448–6453 (1998).
6. Lambert, M.P. *et al.* *J. Neurochem.* **79**, 595–605 (2001).
7. Guerois, R., Nielsen, J.E. & Serrano, L. *J. Mol. Biol.* **320**, 369–387 (2002).