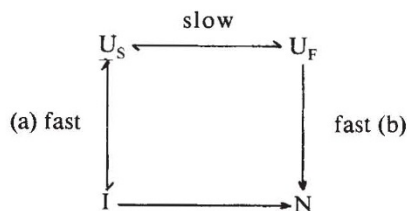


50-fold. The more general basic model for refolding is now



where I represents one, or a sequence, of intermediate states with improper prolines. Levitt has drawn attention to the consequence inherent in this model that the refolding route taken will be a function of the difference in conformational energy between the native state N and the similar though perturbed conformation containing the wrong proline isomer (*J. molec. Biol.* **144**; 1981). The higher the strain associated with the incorporation of incorrect proline isomers into the native conformation, the greater the relative importance of pathway (b) and the more the system will approximate to Brandts' original model. At the other extreme, if the difference in energy is low, pathway (a) will predominate and isomerization only be detectable insofar as the folded state I differs spectroscopically or in stability from state N. State I will be expected to be significantly populated under conditions favouring the 'native' state.

Levitt, on the basis of conformational enthalpy calculations, postulates three levels of strain energy resulting from the insertion of incorrect proline isomers, types I, II and III corresponding to energy differences of approximately 1, 10 and 30 kcal mol⁻¹. Types III and I correspond, respectively, to the two cases cited above, while in type II the folded state containing an incorrect proline would be relatively stable but significantly less so than the native state.

Jullien and Baldwin have taken up this classification and have suggested kinetic and environmental criteria whereby the three types may in principle be recognized experimentally. On the basis of spectroscopic and urea gel electrophoresis (Creighton *J. molec. Biol.* **137**, 61; 1980) investigations on about ten proteins, they find definite evidence for proteins with type II prolines only, although the existence of type I prolines cannot be ruled out. Thus, in any system some molecules refold fast, others slowly. Generalizations about globular proteins are notoriously prone to contradiction, however, and investigation of a wider range of proteins may reveal different behaviour.

Levitt's main contribution is to calculate the strain energies for each of the four prolines placed, as the incorrect isomer, in bovine pancreatic trypsin inhibitor (BPTI). The effect on the protein is shown to involve mainly local and relatively small shifts from the native conformation. He finds that the destabilization of the native

structure is 1 kcal mol⁻¹ for Pro 13 (therefore type I), 11 kcal mol⁻¹ for Pro 2 and Pro 9 (type II) and 33 kcal mol⁻¹ for Pro 8 (type III). On the basis of these assignments, one can calculate the proportion of BPTI molecules which will refold rapidly and Levitt obtains a value of 0.77. Jullien and Baldwin, who carry out a detailed experimental investigation of the refolding of a slightly modified BPTI, find that the relative amplitude of the fast folding U_F → N reaction is 75 per cent, in good agreement with the prediction. They also find one slow folding reaction which satisfies the criteria for a type II residue together with preliminary evidence strongly suggestive of one or two type I residues. Type III behaviour has not been found, although this may be due to greater conformational mobility in the modified BPTI which they study. The 'catalysis' of isomerization found experimentally has been studied by Levitt who calculates that states intermediate between *trans* and *cis*, with the peptide bond angle $\omega \sim 90^\circ$, have lower energies than the strain energies in

the *cis* form. The protein seems to be able to accelerate the rate of isomerization by applying a couple to the ω angle.

Taken together, these two different approaches provide more circumstantial evidence for proline isomerization as the explanation for the biphasic refolding of proteins. If the fast and slow phases do, in fact, correspond to the sequential U_S → I and I → N reactions where I is a compact configuration, and if most proteins turn out to fold as type II or type I, the proline quirk affects the exquisite efficiency of the folding process less than might have been initially feared. The polypeptide chain still collapses rapidly, as suggested by Creighton, to a compact form within which the prolines can sort themselves out at leisure. The collapse may be the most important feature in that it reduces the possibilities for entanglement between elements of different domains. Presumably the conformational requirements for proline residues in the folded protein outweigh the necessity for an overall rapid folding process. □

Defining a lectin

from Jan Kocourek and Václav Hořejší

Recent correspondence (see *Nature* **285**, 66; 1980) demonstrates that there is an urgent need to formulate a definition of lectins which clearly delineates the meaning of the term. However, Goldstein *et al.*'s proposed definition of a lectin as "a sugar-binding protein or glycoprotein of non-immune origin which agglutinates cells and/or precipitates glycoconjugates" seems to disregard some important points. In particular, because the role of lectins in living objects is unclear, it does not seem appropriate to define lectins by their *in vitro* biological activities. The agglutination of cells — a complex and poorly understood phenomenon — may, in the majority of cases, be a fortuitous effect of lectins. The definition of lectins should be based primarily on their most important physicochemical properties, that is, on their interaction with carbohydrates.

The inclusion into the definition of some *in vitro* activities of lectins, like agglutination of cells and precipitation of glycoconjugates, would exclude those lectins which have only one binding site and/or show some toxic or hormone-like activities. This is especially true in the case of the toxins of *Ricinus communis*

and *Abrus precatorius* which are structurally related to the 'agglutinins' and occur in the same sources so that their common evolutionary origin is obvious. Recent research shows that many of these agglutinins (including that of *R. communis*) may exert very similar properties to their toxic counterparts, if acting in cell-free systems (Barbieri *et al. Biochem. J.* **182**, 633; 1979).

As long as we consider lectins as recognition molecules and assume that they mediate information in various (but still largely unknown) physiological reactions, their definition should be based on only three distinct features which they all have in common. Lectins

- (1) have at least one binding site which can bind free or glycosidically linked carbohydrates: this interaction can be inhibited by low-molecular-weight sugars;
- (2) are not synthesized due to an immune response and do not have immunoglobulin structure;
- (3) do not show enzymatic activities towards carbohydrates to which they bind.

We thus suggest the following definition: lectins are sugar-binding proteins or glycoproteins of non-immune origin which are devoid of enzymatic activity towards sugars to which they bind and do not require free glycosidic hydroxyl groups on these sugars for their binding.

Jan Kocourek is in the Department of Biochemistry, Charles University, Prague and Václav Hořejší is in the Institute of Molecular Genetics, Czechoslovakia Academy of Sciences, Prague.