Protein folding

A case of mistaken identity

Roger H. Pain

THE structure of molecular biology is built on several basic hypotheses, including the belief that one gene encodes one polypeptide chain and that the product of expression of a single gene is associated with a single biochemical function or activity. It has been realized for some time that the relation between DNA and protein-mediated functions is more complex, but the recent report¹ by Pihlajaniemi *et al.*, concluding that the product of a single gene seems to possess two entirely different enzymatic functions, one associated with the structure of organisms and the other with protein folding, challenges some of these intuitively held assumptions.

Elephants are largely held together by collagen, and the biosynthesis and assembly of this deceptively simple protein require the involvement of a dozen or more enzymes. In particular, the stability of the three-stranded helix of collagen depends on a proportion of the proline residues in the characteristic repeating sequence Gly-X-Pro being hydroxylated to hydroxyproline after synthesis of the collagen polypeptide. This reaction is catalysed by prolyl 4-hydroxylase, an enzyme comprising two α -subunits and two β -subunits. The two β -subunits are encoded by different genes and assemble from a large pool of β which is present in excess. The β -subunit is intimately involved in prolyl hydroxylase activity, because the α -subunit is inactive alone, because the whole enzyme is inactivated by an antibody specific for the β -subunit, and because both

subunits contribute to the ascorbate-binding site. As part of an extensive research programme on the biosynthesis of collagen, Pihlajaniemi and colleagues cloned and sequenced the gene encoding the β -subunit of prolyl 4-hydroxylase¹, thereby deriving the amino-acid sequence of the protein itself.

Apart from providing essential information for mechanistic and structural investigations, protein sequences are the basis for determining family relationships between proteins, and the existence of computerized databanks allows rapid comparison of new sequences with the many proteins whose sequences are already known. Operation of this now routine procedure on the sequence of the β -subunit threw up an entirely unexpected result: 94 per cent of the amino acid residues comprising the β -subunit of human prolyl 4-hydroxylase are identical to those in the rat enzyme, protein disulphide isomerase (PDI). Assaying the β -subunit shows that it has the same level of PDI activity as the enzyme previously isolated under this name. Further, the appropriate DNA probe shows the existence of only one gene encoding the common sequence in the human genome. In other words it seems highly likely that the small differences in sequence reflect species differences, and that the β -subunit of prolyl 4-hydroxylase is the same protein as PDI, both enzymic activities being encoded by a single gene.

Protein disulphide isomerase is an enzyme thought to catalyse the rate of reformation of correct disulphide bridges, hence allowing refolding of globular proteins whose three-dimensional conformation has been unfolded by urea and whose disulphide bridges have been broken by a reducing agent². Its tissue distribution has been studied extensively and its highest specific activities are in collagen-synthesizing cells. Its distribution correlates closely with that of prolyl 4-hydroxylase, although the two enzymes were at that time presumed to be different³. The sequence reveals two pairs of cysteine residues, each set in a sequence of 16 residues of nearly identical sequence and whose function is still uncertain but which, as in thioredoxin, could function as an oxidation-reduction couple interacting with cysteine residues in refolding proteins⁴. Although located in the cell and the endoplasmic reticulum, where protein folding takes place, the hypothesis that PDI is physiologically important in the folding of proteins *in vivo* is still unconfirmed.

What is the significance of this in-

triguing finding? Are there two distinct and unconnected enzyme activities, with two separate active sites on the same polypeptide chain together with binding sites for the α -subunit? If so, do these activities reside on separate — and separable — domains, suggesting an earlier gene-fusion event, using the possibly redundant PDI as a convenient piece of protein to stabilize the interaction of the complementary α - and β -catalytic units? Or has PDI been viewed in the wrong light all along? PDI in collagen-synthesizing tissue can now be seen to be prolyl 4-hydroxylase, whose function is well established, obviating the previously held view that PDI is required in collagen biosynthesis.

Could it be that those seeking the connection between the two enzyme activities should focus on their ability to bind proline? The rate-limiting step of protein folding is frequently the requirement for proline residues to take up one of two possible isomeric configurations, *cis* or *trans*, before the polypeptide chain can condense and pack optimally. This requirement applies to collagen as well as to many globular proteins⁵. Could the enzyme that binds proline in collagen in order to hydroxylate it also bind proline in a refolding polypeptide and thereby catalyse the *cis-trans* isomerization reaction?

The case would be argued first on the grounds that, for PDI to increase the rate of refolding, some reducing agent must be present to allow disulphide interchange to occur — only more quickly now that the accompanying non-covalent reshuffling is not being delayed by unaided proline isomerization. The hypothesis involves the interaction of 'PDI' with a proline residue at the surface of the collapsed and folding protein. This scheme is simpler to envisage than a model⁶ in which a disulphide bond on PDI has to interact in an analogous fashion with two separate sulphydryls. This alternative view would involve, however, a rather strange shifting of the *cis-trans* proline equilibrium on binding rather than classical catalysis, and to what extent this would result in a higher proportion of proline configurations that fit into the native protein conformation is not immediately obvious.

It is to be hoped that this latest discovery¹, serendipitous though it may be, will lead to further studies on this interesting molecule to clarify the details of how it performs these two seemingly unconnected functions. □

1. Pihlajaniemi, T. *et al.* *EMBO J.* **6**, 643–649 (1987).
2. Freedman, R.B. *Trends biochem. Sci.* **9**, 438–441 (1986).
3. Brockway, B.E., Forster, S.J. & Freedman R.B. *Biochem. J.* **191**, 873–876 (1980).
4. Edman, J.C. *et al.* *Nature* **317**, 267–270 (1985).
5. Freedman, R.B. *Nature* **279**, 756 (1979).
6. Creighton, T.E., Hillson, D.A. & Freedman, R.B. *J. molec. Biol.* **142**, 43–62 (1980).

Roger H. Pain is Professor of Physical Biochemistry at the University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.

100 years ago

GEOGRAPHICAL NOTES

THE rumour as to the death of Mr Stanley is universally discredited in geographical circles. The rumour seems quite inconsistent with the news as to Mr Stanley's having left the Aruwimi River on June 3 for Wadelai. Had he been shot, as reported, it must have been after this date, and during the land journey, whereas one version of the rumour gives out that he was killed on the Congo. He may meet with Emin Pasha sooner than he expected. Emin, it seems, is at present exploring on the south of the Albert Nyanza so that he and Mr Stanley may meet half way. Letters from Mr Stanley are expected in this country early in August. From *Nature* **36**, 309; 28 July 1887.