BIOCHEMICAL LETTERS

Factors that govern the specificity of transglutaminase-catalysed modification of proteins and peptides

In an earlier discussion of this topic (Coussons *et al.*, 1991), we had concluded that there were no obvious regular patterns of sequence around the sites of transglutaminase-catalysed modification in peptides and proteins, and we suggested that conformational factors might be predominant. We have now taken into account additional sequences from ACTH (Pucci *et al.*, 1988a), substance P (Pucci *et al.*, 1988b), melittin (Perez-Paya *et al.*, 1991), fibrin peptide (Gorman & Folk, 1980), crystallins β Bp, β B3 and β A3 (Berbers *et al.*, 1983), seminal vesicle protein SV-IV (Porta *et al.*, 1989) and phospholipase A₂ (Cordella-Miele *et al.*, 1990). As far as we are aware, there are no other established sites of modification by the liver enzyme reported in the literature.

In this Letter, we propose that the minimal requirements for modification of a particular glutamine side chain in a protein are: (i) the glutamine must first satisfy an 'accessibility' criterion, either by being in a highly flexible region of polypeptide chain or by being clearly exposed to solvent in a more structured region of polypeptide chain, and (ii) in the sequence around the glutamine there must be an *absence* of features which appear to *discourage* correct interaction with the enzyme. We shall briefly discuss each of these factors in turn and then illustrate the application of the proposed requirements.

There is no convenient, reliable method for predicting the accessibility of a side chain from sequence data. However, by applying the BIPED procedure (Akrigg *et al.*, 1988) to the Brookhaven protein structure database, it is found that some 8.2% (229 out of 2778) glutamine side chains have relative accessibilities (Chothia, 1975) of 0.66 or greater. Using this figure we can calculate that some 22000 (8.2% of 266873) glutamines in the SWISSPROT (Bairoch & Boeckman, 1991) database (release no. 17) could act as putative substrates in the transglutaminase-catalysed reaction. The choice of 0.66 or greater accessibility is somewhat arbitrary, but can be justified by noting that no glutamine side chains in ribonuclease (which does not act as a substrate) have accessibilities in this range, and that in the case of glucagon, the two glutamines (Gln-3 and Gln-20) which are modified are in this range, whereas the glutamine which is not (Gln-24) has a lower accessibility (0.49).

As mentioned above, there is no obvious consensus sequence around the sites of modification in proteins. However, it is possible to discern regularities if we look for the types of side chains which are *absent* from such sequences. Initially we have chosen to concentrate on the pattern of neighbouring charged side chains, though undoubtedly the distribution of other types of side chain could be important. (For instance, glycine, valine and leucine side chains occur reasonably frequently on the C-terminal side of modified glutamines.) In our analysis, we have used a 'window' of five amino acids on either side of the site of modification, implying that transglutaminase has a type of

extended active site reminiscent of that found, for example, in papain (Berger & Schechter, 1970). If, for example, a positively charged side chain appears at a certain position (say four residues to the N-terminal side) in any substrate, then we conclude that such a residue in this position is compatible with modification. In drawing up a table of these features, we have divided the substrates into two types (A and B). Type A substrates possess little ordered structure on the basis of X-ray, n.m.r. or c.d. measurements, whereas Type B substrates possess ordered structures. The various discouraging features are summarized in Table 1. Overall it appears that positively charged side chains to the C-terminal side of the glutamine side chain play an important role in determining specificity.

We could use the discouraging features listed in Table 1 to refine our estimate of the number of potential sites of modification among the glutamines in the database. A negative charge (Asp/Glu) has a frequency of about 11.5%, whilst a positive charge (Lys/Arg) has a frequency of about 11.1%. Assuming that the occurrence of the charged side chains is random, the probability that at least one of the six definite, discouraging features shown in Table 1 for Type A substrates will *not* occur in a glutamine containing peptide is about 50% (0.89⁶). This would still leave about 4.1% (or 11000) of the glutamines in a position to be modified in the transglutaminase-catalysed reaction.

Given the size of this estimate, it is easier at this stage to apply the rules to proteins whose three-dimensional structure and/or whose behaviour in the reaction is known. We have successfully applied the rules to account for the non-reactivity of at least 15 glutamines of high accessibility in a number of different proteins. Three examples of the application of the rules are given below.

From the X-ray structure of β -lactoglobulin (Papiz *et al.*, 1986), it is clear that the side chains of Gln-13, Gln-35, Gln-68 and Gln-120 are not significantly exposed to solvent. Application of the charge rules shows that none of Gln-5 (Type A; lysine at +3), Gln-59 (Type B; lysine at +1) or Gln-115 (Type B; glutamate at -3) would act as substrates. Gln-155 and Gln-159 would act as substrates if they were sufficiently accessible. This is in agreement with the experimental findings (Coussons *et al.*, 1992) that only Gln-155 and Gln-159 are modified in the presence of dithiothreitol (which presumably breaks the disulphide bond between Cys-66 and Cys-160 to render the C-terminal loop flexible). In the absence of dithiothreitol, no modification is observed.

Of the three glutamine side chains in lysozyme, Gln-41 and Gln-121 are relatively highly accessible, whereas Gln-57 has low accessibility (Blake *et al.*, 1967). All three glutamines have a positively charged side chain (arginine) at the +4 position, which acts as a discouraging feature; thus the lack of modification (Brenner & Wold, 1978) can be understood.

In an earlier paper (Coussons *et al.*, 1991) we showed that wild type yeast phosphoglycerate kinase is not a substrate for transglutaminase. There are eight glutamine side chains in the enzyme; of these Gln-135 and Gln-145 are in the 'nose' region of the N-terminal domain which is exposed to solvent and highly flexible

Table 1. Charged side chains that act as discouraging features in the transglutaminase-catalysed reaction

In this Table the presence of an asterisk at a particular position relative to the glutamine (at position 0) indicates that the given charged side chain would act as a discouraging feature. Such side chains are absent from the known substrates for transglutaminase. The parentheses indicate that these designations are less definite, since they refer to relatively poor substrates (succinylated β -casein and the heat-denatured α -chain of haemoglobin). The distinction between Type A and Type B substrates is explained in the text. Since the majority of the modification reactions are carried out at a pH of 7.5 or above, histidine side chains are not included in the charged side chain category.

	Position ...	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Positive charges												
Type A substrates		*					Q	(*)	(*)	*	*	*
Type B substrates							Q	*			*	
Negative charges												
Type A substrates			*				Q				*	*
Type B substrates				*			Q					

on the basis of X-ray studies (Watson *et al.*, 1982). Applying the rules for Type A substrates, neither Gln-135 (lysine at +1 and arginine at +3) or Gln-145 (lysines at +1 and +3) would be allowed as substrates. Of the remaining glutamines, Gln-37 has the highest accessibility (0.63). This Type B substrate (α -helical region of polypeptide chain) would not be allowed because of an arginine at the +1 position. Hence, the lack of reactivity can be accounted for.

These examples serve to illustrate the application of the rules. We believe that they can serve as a means of understanding some of the features of the transglutaminase-catalysed reaction and should act as a basis for future experimentation and discussion.

We thank the Science and Engineering Research Council for the award of a CASE (Cooperative Awards in Science and Engineering) studentship to P.J.C.

Peter J. COUSSONS*, Nicholas C. PRICE*, Sharon M. KELLY*, Bryan SMITH† and Lindsay SAWYER‡

*Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, Scotland, U.K., †Department of Protein Chemistry, Celltech Ltd., 216 Bath Road, Slough SL1 4EN, Berks., U.K., and ‡Department of Biochemistry, University of Edinburgh, George Square, Edinburgh EH8 9XD, Scotland, U.K.

Akrigg, D., Bleasby, A. J., Dix, N. I. M., Findlay, J. B. C., North, A. T. C., Parry-Smith, D., Wootton, J. C., Blundell, T. L., Gardner,

S. P., Hayes, F., Islam, S., Sternberg, M. J. E., Thornton, J. M., Tickle, I. J. & Murray-Rust, P. (1988) *Nature* (London) **335**, 745–746
 Bairoch, A. & Boeckman, B. (1991) *Nucleic Acids Res.* **19**, 2247–2249
 Berbers, G. A. M., Bentlage, H. C. M., Brans, A. M. M., Bloemendal, H. & de Jong, W. W. (1983) *Eur. J. Biochem.* **135**, 315–320
 Berger, A. & Schechter, I. (1970) *Philos. Trans. R. Soc. London Ser. B* **257**, 249–264
 Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967) *Proc. R. Soc. London Ser. B* **167**, 365–377
 Brenner, S. C. & Wold, F. (1978) *Biochim. Biophys. Acta* **522**, 74–83
 Chothia, C. (1975) *Nature* (London) **254**, 304–308
 Cordella-Miele, E., Miele, L. & Mukherjee, A. B. (1990) *J. Biol. Chem.* **265**, 17180–17188
 Coussons, P. J., Kelly, S. M., Price, N. C., Johnson, C. M., Smith, B. & Sawyer, L. (1991) *Biochem. J.* **273**, 73–78
 Coussons, P. J., Price, N. C., Kelly, S. M., Smith, B. & Sawyer, L. (1992) *Biochem. J.*, in the press
 Gorman, J. J. & Folk, J. E. (1980) *J. Biol. Chem.* **255**, 419–427
 Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E. & Krauli, P. J. (1986) *Nature* (London) **324**, 383–385
 Perez-Paya, E., Thiaudiere, E., Abad, C. & Dufourcq, J. (1991) *FEBS Lett.* **278**, 51–54
 Porta, R., Esposito, C., Metafora, S., Malorni, A., Pucci, P. & Marino, G. (1989) *Adv. Exp. Med. Biol.* **250**, 403–409
 Pucci, P., Malorni, A., Marino, G., Metafora, S., Esposito, C. & Porta, R. (1988a) *Biochem. Biophys. Res. Commun.* **154**, 735–740
 Pucci, P., Malorni, A., Marino, G., Metafora, S., Esposito, C. & Porta, R. (1988b) *Adv. Exp. Med. Biol.* **231**, 73–78
 Watson, H. C., Walker, N. P. C., Shaw, P. J., Bryant, T. N., Wendell, P. L., Fothergill, L. A., Perkins, R. E., Conroy, S. C., Dobson, M. J., Tuite, M. F., Kingsman, A. J. & Kingsman, S. M. (1982) *EMBO J.* **1**, 1635–1640

Received 14 November 1991