Biochem. J. (1991) 274, 899–902 (Printed in Great Britain)

Rat tyrosine kinase inhibitor shows sequence similarity to human α_2-HS glycoprotein and bovine fetuin

Martina HAASEMANN,*‡ Peter NAWRATIL*† and Werner MÜLLER-ESTERL*§

*Institute of Physiological Chemistry and Pathobiochemistry, University of Mainz, Duesbergweg 6, D-6500 Mainz, and †Department of Clinical Chemistry and Clinical Biochemistry, University of Munich, Nussbaumstrasse 20, D-8000 Munich 2, Federal Republic of Germany

Human α_2-HS glycoprotein and bovine fetuin, abundant proteins of fetal plasma, are structural members of the fetuin family within the cystatin superfamily. They are characterized by the presence of two *N*-terminally located cystatin-like units and a unique *C*-terminal sequence segment not present in the other members of the cystatin superfamily. Search for related sequences revealed that the natural inhibitor of the insulin receptor tyrosine kinase [Auberger, Falquerho, Contreres, Pages, Le Cam, Rossi & Le Cam (1989) Cell (Cambridge, Mass.) **58**, 631–640] shows sequence similarity to the mammalian fetuins. The sequence identity between rat tyrosine kinase inhibitor, human α_2-HS glycoprotein and bovine fetuin is 56 and 60% respectively (percentage of residues in identical positions). The sequence similarity extends over the entire protein structures, except the extreme *C*-terminal portions. In particular, the number and relative positions of the cysteine residues are invariant among the proteins, suggesting that the characteristic array of linearly arranged and tandemly repeated disulphide loops of the cystatin superfamily is also present in rat tyrosine kinase inhibitor. We conclude that rat tyrosine kinase inhibitor may be classified as a novel member of the mammalian fetuin family.

INTRODUCTION

Mammalian cystatins form a superfamily of proteins characterized by their sequence similarity to hen cystatin, the principal cysteine proteinase inhibitor of the egg white [1]. By the criterion of sequence similarity, four distinct families named after their prototypic members constitute the superfamily, i.e. stefins (family 1), cystatins (family 2), kininogens (family 3) and fetuins (family 4) [2]. The presence of two cystatin-like domains in fetuins and three cystatin-like domains in kininogens respectively has led to the hypothesis that the families of fetuins and kininogens arose from the ancient cystatin family by two successive gene-duplication events [2–4].

At present, two members of the fetuin family are known, i.e. human α_2-HS glycoprotein and bovine fetuin [5,6]. Fetuins are typical secretory proteins endowed with a signal sequence of 18 amino acid residues followed by a sequence of approx. 350 residues of the mature protein. The fetuins hold three major domains, i.e. two *N*-terminally located cystatin units (D1, D2) of 116–118 residues and a single *C*-terminal domain (D3) of 106–115 residues that is seemingly unrelated to other proteins [7,8]. This latter domain harbours a proline-rich portion followed by a *C*-terminal segment that is prone to proteolytic modification in human α_2-HS glycoprotein [7,9–11].

Bovine fetuin and human α_2-HS glycoprotein are each products of a single gene [8,11]. The mature proteins are abundant in fetal and adult plasma; they are also present in the developing brain [12]. Bovine fetuin is widely used as an additive in cell-culture media and has been implicated in regulation of cell growth [13], brain development [12], lipid metabolism [14] and serine- (but not cysteine-) proteinase inhibition [15]. However, the molecular basis of the functional diversity of bovine fetuins has remained obscure. Similarly, human α_2-HS glycoprotein is thought to be involved in diverse functions such as bone metabolism [16], opsonization [17] and endocytosis [18]; however, none of the proposed roles has been substantiated so far.

In an attempt to identify sequence motifs of fetuins present in other mammalian proteins we have searched a protein database and identified a close structural relationship between fetuins and the natural inhibitor of the tyrosine kinase associated with the rat insulin receptor [19]. Two cystatin-like units are present in the *N*-terminal portion of the rat protein followed by a *C*-terminal domain holding a proline-rich sequence motif. Rat tyrosine kinase inhibitor can therefore be classified as a novel member of the mammalian fetuin family.

EXPERIMENTAL

Comparison of protein sequences

To detect distant relationships among protein sequences we applied the FASTA program [20] on a MicroVax 3600 computer using the merged database MIPSX of the Martinsried Institute for Protein Sequences, Martinsried, Germany. Sequence alignments were done by the ALIGN program of the Protein Identification Research (PIR) Foundation, Washington, DC, U.S.A. [21], and were further optimized manually.

Hydropathy calculations

The hydropathy profiles were calculated by using the moving-segment approach which continuously determines the average hydropathy within a segment length of nine amino acid residues as it advances through the sequence [22].

Nomenclature of fetuins

Proteins from various species characterized by extensive sequence similarity with bovine fetuin are denoted as 'fetuins'. For human fetuin, the original designation of 'α_2-HS glycoprotein' [23] is used synonymously. The rat tyrosine kinase inhibitor of the insulin receptor was previously denoted 'PP^{63}' [24].

‡ Present address: Max-Planck Institute for Biochemistry, Am Klopferspitz, D-8033 Martinsried, Federal Republic of Germany.
§ To whom correspondence and reprint requests should be addressed.

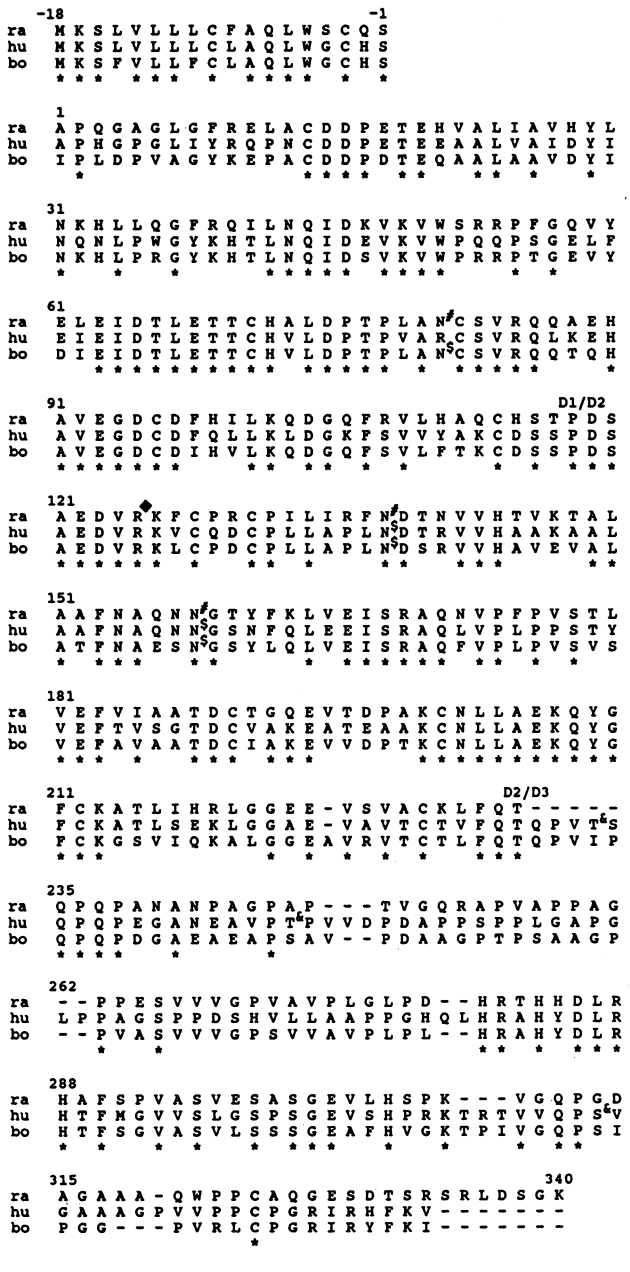

Fig. 1. Alignment of the amino acid sequences of mammalian fetuins

Sequences of rat tyrosine kinase inhibitor [19], human α_2-HS glycoprotein [11] and bovine fetuin [8] were aligned by the ALIGN algorithm [21] and further adjusted manually. A minimum number of gaps (identified by hyphens) was introduced to optimize the alignment. Numbers identify the relative positions of the residues in the rat protein. Negative numbering refers to the signal peptide portions of the proteins. Residues present in invariant positions of all three sequences are highlighted by asterisks. The # symbols indicate the potential *N*-glycosylation sites in the rat protein. *N*-Glycosylation ($) and *O*-glycosylation (&) sites previously identified in human α_2-HS glycoprotein [9,10] and bovine fetuin [29] are displayed. The dibasic cleavage site of Arg-Lys reported for human α_2-HS glycoprotein [7] and bovine fetuin [8] is marked by ◆. The proposed domain boundaries of the mature proteins (D1/D2; D2/D3) are shown above the sequences. The rat sequence starting with the second initiator methionine is given [19]. Abbreviations: ra, rat tyrosine kinase inhibitor; hu, human α_2-HS glycoprotein; bo, bovine fetuin.

Table 1. Sequence identity between rat tyrosine kinase inhibitor, human α_2-HS glycoprotein and bovine fetuin

The sequence identities at the protein level between tyrosine kinase inhibitor [19], human α_2-HS glycoprotein [17] and bovine fetuin [8] are given as the percentage of identical residues normalized for the rat sequence. The calculations are based on the alignment of Fig. 1; note that the rat sequence starting with the second initiator methionine residue is used [19]. The domain organization originally proposed for human α_2-HS glycoprotein [7] was used to dissect the proteins into four major segments; the numbers of residues constituting the various domains are given for the rat protein.

Protein segment	Residue no.	Identity (%)
Overall	358	46.8
Signal peptide	18	72.7
Domain D1	118	52.5
Domain D2	116	58.6
Domain D3	106	27.4
Subdomain D3a	45	15.7
Subdomain D3b	61	36.4

RESULTS AND DISCUSSION

Sequence identity of rat tyrosine kinase inhibitor with mammalian fetuins

Screening of a protein database with the amino-acid-sequence data of mammalian fetuins revealed that the sequence of rat tyrosine kinase inhibitor predicted from the corresponding cDNA [19] shares considerable sequence coincidences with human α_2-HS glycoprotein and bovine fetuin (Fig. 1). The overall sequence identity (percentage of amino acid residues in identical positions of the sequences) of the rat inhibitor is 56.3% with human α_2-HS glycoprotein and 59.6% with bovine fetuin. Continuous stretches of 18 residues (positions 199–216) of the human protein and 14 residues (positions 74–87 and positions 200–213) of the bovine protein are invariantly present in the rat protein. Furthermore, the number ($n = 12$) and relative positions of the cysteine residues of the mature proteins are identical, thus strongly suggesting that the particular pattern of disulphide loops recently identified in human α_2-HS glycoprotein [7] is also present in the rat protein. A dibasic cleavage site of Arg–Lys previously recognized in human α_2-HS glycoprotein [7] and bovine fetuin [8], which is juxtaposed to the typical sequence motif of Cys-Xaa-Xaa-Cys [25], is present in an invariant position of the rat protein. Likewise, the relative positions of the three canonical acceptor sites for *N*-glycosylation (Asn-Xaa-Thr/Ser) are identical among the rat and the bovine proteins (note that the human protein has only two *N*-glycosidic acceptor sites). Another feature of mammalian fetuins is a proline-rich segment of 51–57 residues, where 29.8 and 25.5% of the residues are proline in human α_2-HS glycoprotein and bovine fetuin respectively. A similar region of 45 residues in length (positions 235–279) that is abundant in proline (28.8%) is found in rat tyrosine kinase inhibitor (cf. Fig. 1).

Our data indicate a striking sequence similarity among rat tyrosine kinase inhibitor, human α_2-HS glycoprotein and bovine fetuin. Furthermore, a weak, though distinct, sequence similarity is seen with human histidine-rich glycoprotein ('HRG'), another member of the cystatin superfamily containing two cystatin-like building blocks [26] (results not shown).

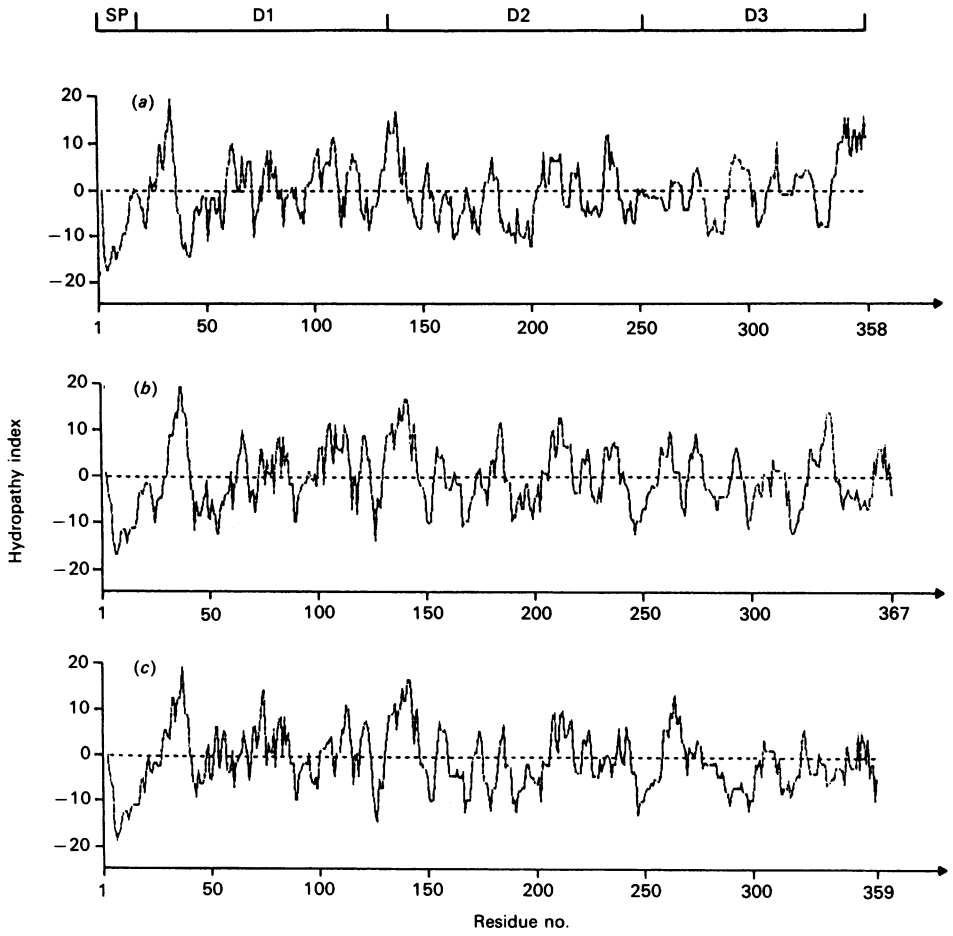

Fig. 2. Hydropathy profiles of rat tyrosine kinase inhibitor, human α_2-HS glycoprotein and bovine fetuin

The hydropathy profiles were calculated by using the moving-segment approach [22]. The hydropathy index (ordinate) identifies hydrophilic (positive numbering) and hydrophobic segments (negative numbering) of rat tyrosine kinase inhibitor (*a*), human α_2-HS glycoprotein (*b*), and bovine fetuin (*c*). The relative positions of the residues in the amino acid sequences are given (abscissa). No gaps have been introduced into the sequences. The domain organization of the proteins [SP (signal peptide); D1, D2 and D3 (domains 1, 2 and 3)] is shown on top of the diagram.

Domain structure of tyrosine kinase inhibitor and mammalian fetuins

To evaluate further the sequence similarity of tyrosine kinase inhibitor to mammalian fetuins, we have compared the relative lengths and sequence identities of the various domains originally proposed for α_2-HS glycoprotein [7] (Table 1). The lengths of the putative signal peptides ($n = 18$ residues), the first cystatin-like domains (D1; $n = 118$ residues) and the second cystatin-like domains (D2; $n = 116$–117) are almost the same for the three proteins, whereas the length of the terminal domains (D3) is variable ($n = 106$–115). This finding correlates with the observation that the pattern of sequence coincidence is unevenly distributed over the domains of the proteins (cf. Table 1). The sequence identity is highest in the signal-peptide portions (72.7%), moderate in the two cystatin-like domains (52.5 and 58.1% for D1 and D2 respectively), but decreases sharply in domain D3 (27.4%). This latter domain holds two subdomains, i.e. a proline-rich segment of 45 residues (D3a; positions 235–279 of the rat protein) and a *C*-terminal region of 61 residues (D3b; positions 280–340). The sequence similarity is lowest in the proline-rich portions (15.7%), whereas it approaches that of the overall molecule (46.8%) in the *C*-terminal regions (36.4%). The clusters of sequence identity dissipated over the molecules are reflected by the corresponding hydropathy profiles, which are similar for the various proteins, except for the extreme *C*-termini (Fig. 2) (cf. below). In particular, the hydropathy profiles of the *N*-terminal portions (signal peptide and domains D1 and D2) of the rat and the human proteins are strikingly similar. Note that the overall hydrophobicity of the proline-rich region is well preserved among the various proteins, despite considerable sequence divergence in this region.

We conclude that the natural inhibitor of the insulin-receptor tyrosine kinase exhibits the prominent structural features of mammalian fetuins, thus representing a novel member of the fetuin family (family 4) of the cystatin superfamily.

Structural divergence of tyrosine kinase inhibitor and mammalian fetuins

Apart from the overall sequence similarity among the various proteins, two major differences are seen between rat tyrosine kinase inhibitor and mammalian fetuins. First, the cDNA sequence of the tyrosine kinase inhibitor predicts an initiation site at a triplet of AUG 51 nucleotides upstream of a second AUG codon (not shown). Auberger *et al.* [19] noted that the first AUG codon lies in an unfavourable context (U is position −3), whereas the second AUG codon is preceded by the canon-

ical sequence of GCC flanking translation initiation sites in eukaryotic mRNAs [27]. Therefore the authors [19] suggested that the second initiator codon is preferentially used for translation initiation. Their notion is supported by the fact that the first methionine residue is followed by a hydrophilic stretch of amino acid residues, whereas the second precedes a hydrophobic sequence conforming to the characteristics of mammalian signal peptides [19]. Our sequence alignments demonstrate that the second translation initiation site coincides with the single initiator methionine residue of mammalian fetuins (cf. Fig. 1, position −18), thereby supporting the concept that the second, rather than the first, AUG is the dominant translation initiation site of the tyrosine-kinase-inhibitor mRNA.

Further structural differences between mammalian fetuins and tyrosine kinase inhibitor are seen in the extreme *C*-terminal sequences of domain D3, which diverge considerably between the rat protein and mammalian fetuins, whereas the sequences of human α_2-HS glycoprotein and bovine fetuin are similar in this region (cf. Fig. 1). Also, the *C*-terminal sequence of rat fetuin extends the sequences of the other mammalian fetuins by seven residues. In this context it is worth mentioning that a second type of tyrosine-kinase-inhibitor cDNA was found by Auberger *et al.* [19] which contained a slightly shorter coding region within the same reading frame. Unfortunately the nucleotide sequence of this extra cDNA was not specified, thus precluding the sequence comparison with human and bovine fetuins at the present time.

The question might arise as to whether rat tyrosine kinase inhibitor represents a structurally related, but functionally divergent, protein of the rat, or the rat homologue of human α_2-HS glycoprotein and bovine fetuin, i.e. rat fetuin [28]. Elucidation of the primary structure of rat fetuin will allow the unequivocal evaluation of the structural relationship of tyrosine kinase inhibitor and fetuin from the rat.

Functional diversity of the cystatin superfamily

We arrive at the conclusion that the cystatin superfamily holds two sets of structurally related, but functionally divergent, enzyme inhibitors, i.e. the well-established cysteine-proteinase inhibitors of the stefin, cystatin and kininogen type, and a family of fetuin-like proteins operative in tyrosine kinase inhibition, thereby serving functions in fundamental biological processes such as growth, development and differentiation.

We are indebted to our colleagues Dr. J. Sgouros, Martinsried Institute for Protein Sequences, for help with the database searches, and Dr. W. Jahnen-Dechent, Department of Pathobiochemistry, University of Mainz, for critical reading of the manuscript. The contributions of S. Holz and E. Haas in the preparation of the manuscript are gratefully acknowledged. This work was supported by grants from the Deutsche Forschungsgemeinschaft (Mu 598/3-1) and the Sonder forschungsbereich 207 (F-1) of the University of Munich (to W.M.E.)

Received 5 December 1990; accepted 17 January 1991

REFERENCES

1. Barrett, A. J. (1987) Trends Biochem. Sci. **12**, 193–196
2. Rawlings, N. D. & Barrett, A. J. (1990) J. Mol. Evol. **30**, 60–71
3. Müller-Esterl, W., Fritz, H., Kellermann, J., Lottspeich, F., Machleidt, W. & Turk, V. (1985) FEBS Lett. **191**, 221–226
4. Kellermann, J., Thelen, C., Lottspeich, F., Henschen, A., Vogel, R. & Müller-Esterl, W. (1987) Biochem. J. **247**, 15–21
5. Christie, D. L., Dziegielewska, K. M., Hill, R. M. & Saunders, N. R. (1987) FEBS Lett. **214**, 45–49
6. Elzanowski, A., Barker, W. C., Hunt, L. T. & Seibel-Ross, E. (1988) FEBS Lett. **227**, 167–170
7. Kellermann, J., Haupt, H., Auerswald, E. A. & Müller-Esterl, W. (1989) J. Biol. Chem. **264**, 14121–14128
8. Dziegielewska, K. M., Brown, W. M., Casey, S. J., Christie, D. L., Foreman, R. C., Hill, R. M. & Saunders, N. R. (1990) J. Biol. Chem. **265**, 4354–4357
9. Gejyo, F., Chang, J. L., Bürgi, W., Schmid, K., Offner, G. D., Troxler, R. F., Van Halbeek, H., Dorland, L., Gerwig, G. J. & Vliegenthart, J. F. G. (1983) J. Biol. Chem. **258**, 4966–4971
10. Yoshioka, Y., Gejyo, F., Marti, T., Rickli, E. E., Bürgi, W., Offner, G. D., Troxler, R. F. & Schmid, K. (1986) J. Biol. Chem. **261**, 1665–1676
11. Lee, C. C., Bowman, B. H. & Yang, F. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 4403–4407
12. Dziegielewska, K. M., Møllgard, K., Reynolds, M. L. & Saunders, N. R. (1987) Cell Tissue Res. **248**, 33–41
13. Puck, T. T., Waldren, C. A. & Jones, C. (1968) Proc. Natl. Acad. Sci. U.S.A. **59**, 192–199
14. Cayatte, A. J., Kumbla, L. & Subbiah, M. T. R. (1990) J. Biol. Chem. **265**, 5883–5888
15. Rohrlich, S. T. & Rifkin, D. B. (1981) J. Cell. Physiol. **109**, 1–15
16. Colclasure, G. C., Lloyd, W. S., Lamkin, M., Gonnerman, W., Troxler, R. F., Offner, G. D., Bürgi, W., Schmid, K. & Nimberg, R. B. (1988) J. Clin. Endocrinol. Metab. **66**, 187–192
17. van Oss, C. J., Gillman, C. F., Bronson, P. M. & Border, J. R. (1974) Immunol. Commun. **3**, 329–335
18. Lewis, J. G. & André, C. M. (1980) Immunology **39**, 317–322
19. Auberger, P., Falquerho, L., Contreres, J. O., Pages, G., Le Cam, G., Rossi, B. & Le Cam, A. (1989) Cell (Cambridge, Mass.) **58**, 631–640
20. Pearson, W. R. & Lipman, D. J. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 2444–2448
21. George, D. G., Orcutt, B. C., Dayhoff, M. O. & Barker, W. C. (1986) Protein Identification Resource Report, REL-0286, National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC
22. Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
23. Schultze, H. E., Heide, K. & Haupt, H. (1962) Naturwissenshaften **49**, 15
24. Le Cam, A., Magnaldo, I., Le Cam, G. & Auberger, P. (1985) J. Biol. Chem. **260**, 15965–15971
25. Vogel, R., Assfalg-Machleidt, I., Esterl, A., Machleidt, W. & Müller-Esterl, W. (1988) J. Biol. Chem. **263**, 12661–12668
26. Koide, T. & Odani, S. (1987) FEBS Lett. **216**, 17–21
27. Kozak, M. (1989) J. Cell Biol. **108**, 229–241
28. Dziegielewska, K. M., Bock, E., Cornelis, M. E. P., Møllgard, K., New, H. & Saunders, N. R. (1983) Comp. Biochem. Physiol. **76A**, 241–245
29. Yet, M. G., Chin, C. C. Q. & Wold, F. (1988) J. Biol. Chem. **263**, 111–117