Volume 325, number 1,2, 34–38 FEBS 12372 June 1993

Minireview

Location of tolerated insertions/deletions in the structure of the maltose binding protein

Jean-Michel Betton, Pierre Martineau, William Saurin and Maurice Hofnung

Unité de Programmation Moléculaire et de Toxicologie Génétique (CNRS UA1444), Département des Biotechnologies, Institut Pasteur, Paris, France

Received 15 February 1993; revised version received 14 March 1993

In a previous study [(1987) J. Mol. Biol. 194, 663–673], we isolated ten insertion/deletion mutants (indels) of the maltose binding protein for which the maltose binding constant was only a little or not at all affected. In this paper, we have localized these mutations in the recently solved three-dimensional structure. Contrary to the general expectation, most of the insertion/deletion modifications occurred within elements of secondary structure. An analysis of the inserted residues for three indels found within α helices allowed an interpretation regarding protein structure accommodation to such modifications.

Insertion/deletion; Linker insertion mutagenesis; Maltose binding protein; Protein structure

1. INTRODUCTION

There are several mutational strategies used in the analysis of protein structure and function [1]. In general, the mutational strategy depends on whether the three-dimensional structure of the protein under investigation is known. Usually, site-specific mutagenesis is applied to proteins of known structure because X-ray crystallography provides an accurate definition of protein structure and allows conventional predictions on the role of particular residues which can be tested by a simple substitution. In the absence of such structural detail, random mutagenesis coupled to classical genetic selection or screening has proven to be a useful way of studying functional regions or residues within proteins.

Genetic studies from this laboratory have focused on proteins involved in the transport of maltose and maltodextrins across the cell envelope of *Escherichia coli*. Maltose binding protein (MBP) is the periplasmic component of this specific transport system [2,3]. We have previously used a random linker insertion mutagenesis approach to identify functionally important regions within the MBP [4]. Results from this study identified several regions required for the stability and export of the protein. In addition, two regions involved in interactions with the membrane protein components of the maltose transport system were identified. At that time, due to the lack of available three-dimensional structure, these mutants were not structurally analyzed. Since then, two crystallographic structures of MBP have been determined; a liganded form with maltose and an unliganded form [5,6]. The overall structure of MBP consists of two discontinuous domains separated by a cleft in which maltose and maltodextrins bind (Fig. 1). Each domain is composed of a core of open mixed β sheets flanked on both sides with α helices.

In the light of the known structure of MBP, we can now localize these mutations in the secondary structure and start to interpret the changes in the primary structure. In this paper, we structurally identify ten linker insertion/deletions which have little or no effect on maltose binding activity of the protein. As expected, all sites of insertion/deletion lie preferentially on the surface of the protein, yet the striking observation is the identification of such sites within α helical or β strand structural elements. A possible mechanism used by the protein to accommodate insertion/deletion within helices is discussed.

2. 'INDEL' MUTANTS

We described previously a set of mutants generated by random insertion of a *Bam*HI octanucleotide linker into the gene of MBP and selection of immunoprecipitable mutant proteins [4]. Although this method generated drastic changes in the amino acid sequence, 19 mutants that accumulated a stable MBP to high levels in the cell were found. Among those, 8 mutant MBPs, and 2 additional recombinants constructed in vitro, retained

Correspondence address: J.-M. Betton, Unité de Programmation Moléculaire et de Toxicologie Génétique (CNRS UA1444), Département des Biotechnologies, Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France. Fax: (33) (1) 45 68 88 34.

 Published by Elsevier Science Publishers B.V.

nearly the wild-type binding constant for maltose (Table I).

Table I lists the amino acid sequence changes in the mutant MBPs. Depending on the reading frame and surrounding nucleotides, 2 or 3 amino acids were inserted at various sites. In most cases the linker insertions were accompanied by small deletions corresponding to the loss of 5 or 6 residues on average. The resulting modifications in the amino acid sequence are insertions with or without deletions. In the field of sequence comparison these modifications are referred to as 'indels' [8,9]. It seems that there is no amino acid preference either in the inserted residues or in the sequences that flank the indel site. The inserted residues (D, P, G, S, I and R) do not introduce a marked hydrophobic or hydrophilic behavior. However, proline has the largest structural constraint and glycine the largest conformational flexibility.

Each of these mutant proteins is discussed with reference to its indel position in the secondary and tertiary structure of MBP (Figs. 1 and 2). Since the mutant proteins have maltose binding constants similar to wild-type MBP, we may assume that the indels produce only small structural perturbations. For two mutant proteins (MalE133 and MalE303), ultraviolet circular dichroism spectra were recorded (data not shown) and no detectable modifications in the periodic structures of MBP was found.

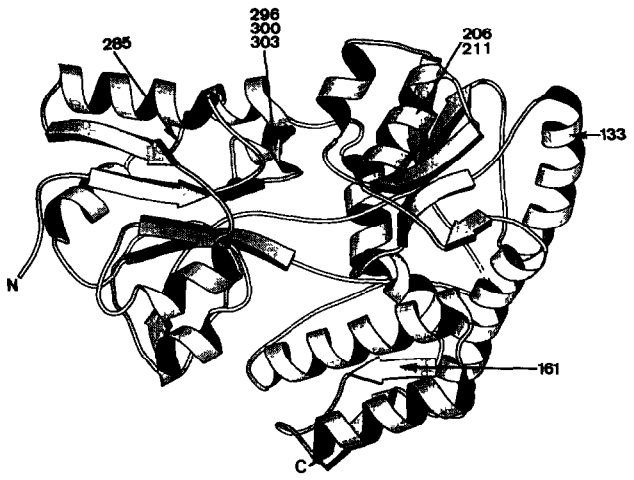

Fig. 1. Location of indel sites in the three-dimensional structure of MBP. Ribbon diagram of the X-ray structure of MBP showing the domain organization from [5]. Indel sites are indicated by arrows and identified by the residue corresponding to the N-terminus flanking position of the 8 different indels (see legend to Table I).

Table I

Indel mutant MBPs

Mutant MBPs[a]	Maltose binding[b] K_d (μM)	Secondary structure[c]	Indel length	Amino acid sequence changes[d]
MalE133 (pPD178)	4.5	αIV	−7	131EIP***DP***GKS145 EIP*ALDKELKAK*GKS
MalE161 (pPD341)	8	βG	−13	159PLI***RIR***IKD180 PLI*AADGGYAFKYENGKYD*IKD
MalE206 (pPD346)	6	αVII	−7	204MNA***RIR***FNK219 MNA*DTDYSIAEAA*FNK
MalE211 (pPD301)	2.3	αVII	−7	209DYS***GS***ETA223 DYS*IAEAAFNKG*ETA
MalE285 (pPD222)	1.6	loop	+3	283YLL***RIR***TDE288 YLLTDE
MalE296 (pPD333)	2.3	βM-αXI	−13	294NKD***TDP***KDP315 NKD*KPLGAVALKSYEEELA*KDP
MalE300 (pPD322)	8	βM	+1	298PLG***GS***VAL304 PLG*A*VAL
MalE300r (pPD322-312)	1.4	βM	−3	298PLG***GSA***YEE309 PLG*AVALKS*YEE
MalE303 (pPD65)	3.5	βM-αXI	−3	301AVA***PDP***ELA312 AVA*LKSYEE*ELA
MalE303r (pPD65-333)	1.1	βM-αXI	−6	301AVA***PDP***KDP315 AVA*LKSYEEELA*KDP

[a] Mutant MBPs are designated by a number corresponding to the N-termini flanking positions of the indel sites. Recombinants are indicated by the letter r following the number. The names of corresponding plasmids are indicated in brackets [9].

[b] Maltose binding constants of the purified mutant MBPs were previously determined by flow dialysis [4]. The maltose binding constant for the wild-type protein is 3.5 μM under these experimental conditions.

[c] Indel sites are localized to the closest corresponding secondary structure element.

[d] Sequences of indels. The numbers correspond to the amino acids of the wild-type sequence of MBP. The upper line gives the sequence in the mutant, and the lower line, in small capitals, gives the wild-type sequence in the corresponding region. Residues in bold-face and italic type correspond to amino acid changes.

2.1. *MalE133*

The site of this mutation is helix IV which is at the surface of the C-domain. As helix IV connects two antiparallel strands (βE and βF), it does not correspond to a β-α-β motif (like helix VII, see below). Its packing against the central β sheet of the C-domain is weak [5]. The indel occurs after a proline residue in the first helical turn, where proline residues are often found in α helices. This mutation corresponds to a 'permissive' site, as we previously defined [10]. Several amino acid sequences have been successfully inserted at this site without major effect on the activity of MBP [11].

2.2. *MalE161*

This indel introduces a complete deletion of strand βG belonging to the two-stranded antiparallel sheet of the small C2 domain. The PD341 corresponding mutant strain retains a wild-type phenotype on dextrins. Thus the two helices (XIII and XIV) which interact with the small β sheet, and surround this indel must have maintained their original orientation because they form a binding subsite for maltodextrins [5].

2.3. *MalE206, MalE211*

These two indels are located in helix VII which faces the substrate binding cleft in the C-domain. In

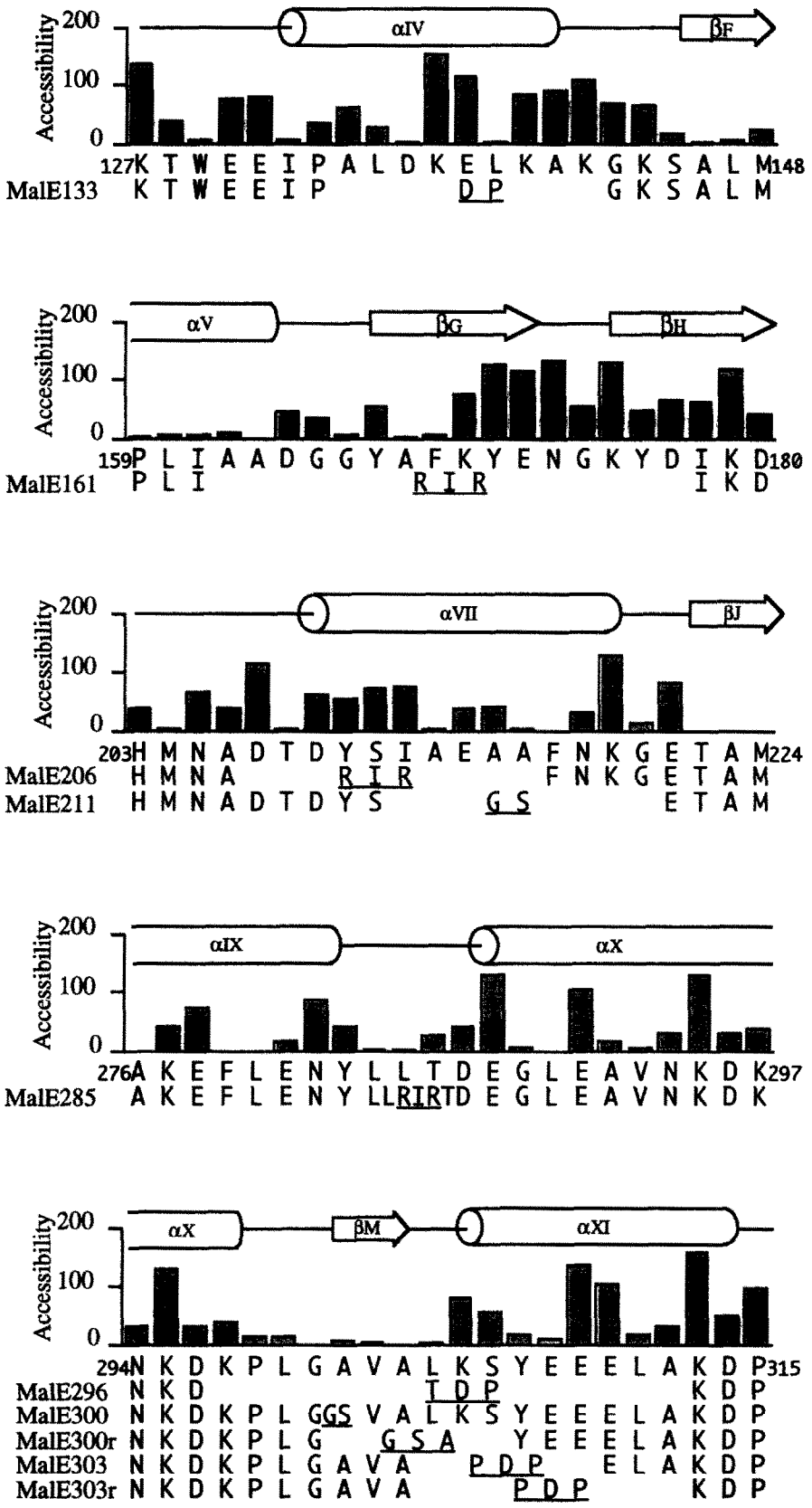

Fig. 2. Location and accessibility of indel sites in the secondary structures of MBP. The secondary structure elements were characterized by the criteria of Kabsch and Sander [7] and labeled according to the numbering of Spurlino et al. [5]. Solvent accessibility was calculated by using the program DSSP [7]. Deletions are indicated by a blank, and inserted amino acid sequences are underlined.

	130										140							
MalE wt	E	E	I	P	A	L	D	K	E	L	K	A	K	G	K	S	A	L	M
X-ray data	t	t	h	h	h	h	h	h	h	h	h	t	t	t	c	b	b	b	c
CF predictions	h	h	h	i	h	h	h	h	h	h	h	h	i	t	h	h	h	h	h
hydrophobic moment	50	36	52	44	43	40	18	25	24	20	17	25	17	25	20	23	42	33	55
MalE133	E	E	I	P	-	-	D	-	-	-	P	-	-	G	K	S	A	L	M
CF predictions	h	h	h	i			t				i			t	h	h	h	h	h
hydrophobic moment	41	36	32	29			29				16			25	20	23	42	33	55

	205										215								
MalE wt	N	A	D	T	D	Y	S	I	A	E	A	A	F	N	K	G	E	T	A
X-ray data	c	t	t	c	c	h	h	h	h	h	h	h	h	h	t	t	c	b	b
CF predictions	h	h	t	t	i	i	h	h	h	h	h	h	h	h	h	i	i	b	b
hydrophobic moment	12	6	19	13	10	32	43	41	32	26	28	27	40	19	11	18	23	9	16
MalE206	N	A	-	-	R	-	-	I	-	-	R	-	F	N	K	G	E	T	A
CF predictions	h	h			h			h			h		i	t	t	t	t	b	b
hydrophobic moment	26	16			6			11			27		40	19	11	18	23	9	16
MalE211	N	A	D	T	D	Y	S	-	-	G	-	-	-	S	-	-	E	T	A
CF predictions	h	h	t	t	t	t	t			t				i			i	b	b
hydrophobic moment	29	26	18	23	25	35	16			12				18			23	9	16

Fig. 3. Secondary structure predictions for the three indels within helices. The letters h, b, c and t correspond to α helix, β strand, coil, and turn, respectively. Chou and Fasman predictions were performed by using a program kindly provided by P. Prevelige [14]. The letter i, which means indetermination, reflects the difficulty of predicting by this conventional method because of overlapping β and α structure predictions [15]. Helical hydrophobic moments were calculated from the hydrophobicity scale of Eisenberg, by using an 8-residue window and vectors projecting radially every 100° [16].

MalE206, the first two turns of the helix are deleted, whereas the last two turns are deleted in MalE211. Contrary to helix IV, helix VII is well packed on the central hydrophobic β sheet. This helix, which appears to play no significant role in maltose binding, has been identified as one of the two regions involved in the interactions between MBP and the membrane components of the maltose transport system [4].

2.4. *MalE285*

This indel corresponds to the longest net insertion (3 residues). It occurs in an unstructured region connecting two helices (IX and X) within the N-domain. Preliminary results concerning the identification of epitopes of MBP have emphasized the importance of this region (Betton et al., unpublished results).

2.5. *MalE296, MalE300, MalE303 and recombinants*

Since all these mutations are localized in the same region of the protein, they can be grouped in one class. The modifications vary from the longest deletion (MalE296), including β strand M and helix XI, to a single residue insertion (MalE300) within β strand M. This region corresponds to the edge of the α/β structure of the N-domain, and just precedes the third inter domain connecting segment, called cross-over 3 by Spurlino et al. [5]. From the unliganded structure [6], recently resolved, a particular role of the helix XI has been attributed in the hinge bending motion which leads to the closed liganded conformation of MBP. Curiously, the maltose binding constants are not seriously affected in these mutant MBPs. This suggests that helix XI might have a less significant role in the domain motion than was attributed by Sharff et al. [6]. The accommodation of various stretches of amino acids underlines the weak tertiary interactions existing in this region. Consistent with this suggestion is the existence of the second identified permissive site (MalE303) among these indels [11].

3. SECONDARY STRUCTURE PREDICTIONS

Indels, as a class of protein modifications, are still relatively unexplored. In a set of evolutionarily related proteins, surface loops are generally the sites of insertions and deletions [12]. These structures can better accommodate structural perturbations than α helices or β sheets. Insertion mutants in alkaline phosphatase have shown that enzymatic activity was fully or partially retained when the insertions were targeted to loops, but was abolished when an α helix or a β strand was interrupted [13]. So, in this study, an unforseen result is that most indels occur within elements of the secondary structure. It would be hazardous to predict the resulting structure accommodation for all indels. Nevertheless, we can attempt to interpret those occur-

ring within α helices (MalE133, 206 and 211). Although all the three indels have the same length, the amino acid sequence inserted corresponds to the three possible coding phases of the linker. We can imagine that deletion accommodation within helices occurs by shifting the packing and bonding of both inserted and flanking residues. In comparing the value of the pitch per residue for an α helix (1.5 Å) and for a β strand (3.5 Å), and the hydrophobic periodicity, one mechanism used by the polypeptide chain to compensate for the deletion could be to create amphiphilic strand instead of amphiphilic helix. Secondary structure predictions on the three helical indels are compatible with this idea. Fig. 3 shows that all indels introduce breaks inside the deleted helices and decrease the helical hydrophobic moment. In MalE206, the Chou and Fasman prediction also suggests a possible shift of helix to residues preceding the indel. Such a structural response to insertion was crystallographically identified in T4 lysosyme [17] (see below).

In the past years, linker mutagenesis has been used to investigate the structure–function relationships in several proteins [18,21]. Unfortunately, the absence of X-ray structures of these proteins precluded the possibility of locating, in the secondary structures, the tolerated indels. However, the frequency of active indel sites found in these proteins denotes that they may exhibit a high degree of adaptibility to indel mutations. Shortle and co-workers [22,23] have investigated the structural accomodation of staphyloccocal nuclease to single and double amino acid insertions. These authors found ten tolerant insertion sites within loops and secondary structure elements. In each case the inserted residues reduce the free energy of denaturation by an amount comparable to an alanine or glycine substitution at the same site. Although we did not measure the energetic effects of modifications, it is predicted that these indel mutant MBPs have lower stabilities. However, all ten indel mutants accumulate within the periplasm to roughly comparable levels to the wild-type protein [4]. Since correlations between thermal stability and rate of intracellular turnover [24] or proteolytic susceptibility [25] have been reported, these mutants MBPs might have lower but comparable stability to the wild-type protein. Very recently, Heinz et al. [17] showed that 1–3 amino acids insertions in an α helix of T4 lysosyme can be accepted either by the translocation of residues from the helix to the preceding loop or by a 'looping-out' in the first or last turn of the helix. In both cases, the length of the helix is conserved and appears to be determined by the extent of tightly packed interface with the rest of the protein. These observations suggest that inserted residues in indel mutant MBPs might have packing interactions to compensate for deleted interfaces.

In an analysis performed on homologous proteins, Pascarella and Argos [9] have shown a similar size for indel length (around 5 residues), but they observed that interruptions in helices and strands were very rare events. From our results, a possible reason for the high frequency of indels found within secondary structure elements could be the complex insertion/deletion modifications introduced by linker insertion mutagenesis. Attempts to identify the detailed structural changes in two of these indel mutant MBPs (MalE133 and MalE303) by X-ray crystallography are currently underway (F.A. Quiocho et al., personnal communication).

Acknowledgements: We thank John Spurlino and Florante Quiocho for kindly providing the full set of coordinates of MBP refined at 1.7 Å used in this paper. We are also grateful to Pary Guilford for carefully reading the manuscript. This work was supported by grants from Ligue Nationale Française contre le Cancer, Fondation pour la Recherche Médicale and Association pour le développement de la Recherche sur le Cancer.

REFERENCES

[1] Zoller, M.J. (1992) Curr. Opin. Biotechnol. 3, 348–354.
[2] Nikaido, H. and Saier, M.H. (1992) Science 258, 936–942.
[3] Kellermann, O. and Szmelcman, S. (1974) Eur. J. Biochem. 47, 139–149.
[4] Duplay, P., Szmelcman, S., Bedouelle, H. and Hofnung, M. (1987) J. Mol. Biol. 194, 663–673.
[5] Spurlino, J.C., Lu, G.-Y. and Quiocho, F.A. (1991) J. Biol. Chem. 266, 5202–5219.
[6] Sharff, A.J., Rodseth, L.E., Spurlino, J.C. and Quiocho, F.A. (1992) Biochemistry 31, 10657–10663.
[7] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.
[8] Kruskal, J.B. (1983) in: Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison (Sankoff, D. and Kruskal, J.B. eds.) pp. 1–44, Addison-Wesley, Reading, MA.
[9] Pascarella, S. and Argos, P. (1992) J. Mol. Biol. 224, 461–471.
[10] Hofnung, M., Bedouelle, H., Boulain, J.C., Clément, J.C., Charbit, A., Duplay, P., Gehring, K., Martineau, P., Saurin, W. and Szmelcman, S. (1988) Bull. Inst. Pasteur 86, 95–101.
[11] Martineau, P., Guillet, J.-G., Leclerc, C. and Hofnung, M. (1992) Gene, 113, 35–46.
[12] Leszczynski, J.F. and Rose, G.D. (1986) Science, 234, 849–855.
[13] Freimuth, P.I., Taylor, J.W. and Kaiser, E.T (1990) J. Biol. Chem. 265, 896–901.
[14] Prevelige, P. and Fasman, G.D. (1989) in: Prediction of Protein Structure and the Principle of Protein Conformation (Fasman, G.D. ed.) pp. 391–416, Plenum Press, New York.
[15] Busseta, B. and Hospital, M. (1982) Biochim. Biophys. Acta 701, 111–118.
[16] Eisenberg, D., Weiss, R.M. and Terwilliger, T.C. (1984) Proc. Natl. Acad. Sci. 81, 140–144.
[17] Heinz, D.W., Baase, W.A., Dahlquist, F.W. and Matthews, B.W. (1993) Nature 361, 561–564.
[18] Barany, F. (1985) Proc. Natl. Acad. Sci. USA 82, 4202–4206.
[19] Tanese, N. and Goff, S.P. (1989) Proc. Natl. Acad. Sci. USA 85, 1777–1781.
[20] Chen, M. and Horwitz, M.S. (1989) Proc. Natl. Acad. Sci. USA 86, 6116–6120.
[21] Breul, A., Kuchinke, W., von Wilcken-Bergmann, B. and Müller-Hill, B. (1991) Eur. J. Biochem. 195, 191–194.
[22] Sondek, J. and Shortle, D. (1990) Proteins 7, 299–305.
[23] Sondek, J. and Shortle, D. (1992) Proteins 13, 132–140.
[24] Parsell, D.A. and Sauer, R.T. (1989) J. Biol. Chem. 264, 7590–7595.
[25] Reidhaar-Olson, J.F., Parsell, D.A. and Sauer, R.T. (1990) Biochemistry 29, 7563–7571.