Biol. Chem., Vol. 378, pp. 583 – 590, July 1997

Review

Hedgehog and Its Patched-Smoothened Receptor Complex: A Novel Signalling Mechanism at the Cell Surface

Joy Alcedo and Markus Noll*

Institut für Molekularbiologie der Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland

* Corresponding author

Pattern formation and morphogenesis depend on the careful execution of complex genetic programs, which are conserved in multicellular organisms. An important signal in some of these programs in *Drosophila* and vertebrates is the secreted Hedgehog (Hh) protein, which primarily functions as an inducer of morphogenetic signals. The Hh signal plays a decisive role in such critical developmental processes as neurulation and somite and limb formation. The Hh signalling pathway exhibits a novel mechanism of signal reception and transduction. In the absence of the Hh signal, the membrane protein Patched (Ptc) represses the constitutive signalling activity of a second membrane protein, Smoothened (Smo), by virtue of its ability to form a Ptc-Smo complex. Hence, mutations within the *ptc* gene that result in the failure of Ptc to inhibit Smo lead to constitutive activity of the Hh signalling pathway and to cancer, such as basal cell carcinoma. For activation of Hh-target genes, the N-terminal signalling domain of Hh binds to the Ptc-Smo receptor complex to activate two parallel signalling pathways. Furthermore, Hh limits its own range of action by impeding its diffusion through (i) covalent linkage of its N-terminal signalling moiety to cholesterol, mediated by the cholesterol transferase activity of its C-terminal moiety, and (ii) induction of, and sequestration by, its antagonist, Ptc.

Key words: Constitutive signalling / Ligand-regulated receptor repressor.

The Signalling Activities of the Hh Protein

The development of multicellular organisms is determined by mechanisms that specify pattern formation, of which signal transduction pathways are integral components. One of these signalling pathways, which is conserved from flies to man, involves the secreted gene products of the *hedgehog (hh)* gene family. Mutations in these genes not only disrupt *Drosophila* and mouse development but also lead to disease in humans, such as holoprosencephaly, which is characterised by poorly developed brain vesicles and severe facial abnormalities (Belloni *et al.*, 1996; Roessler *et al.*, 1996).

In *Drosophila*, Hh is crucial in the establishment of embryonic segments and their parasegmental boundaries. It is expressed in posterior compartments and exerts its function as a short-range signal by maintaining *wingless (wg)* expression in adjacent anterior stripes of cells (Hidalgo and Ingham, 1990; Ingham *et al.*, 1991; Ingham and Hidalgo, 1993). In imaginal discs, the anlagen of the adult epidermal structures, Hh also locally induces the expression of secondary signals, Wg and the TGF-β family member Decapentaplegic (Dpp) (Basler and Struhl, 1994; Tabata and Kornberg, 1994; Heberlein *et al.*, 1995; Zecca *et al.*, 1995), which function as morphogens required in the pattern formation of these structures (Nellen *et al.*, 1996; Zecca *et al.*, 1996). In addition to its short-range function, Hh is believed to have a long-range activity as observed within the dorsal epidermis of the *Drosophila* embryo (Heemskerk and DiNardo, 1994). Similarly, vertebrate Hh homologues exhibit both short-range and long-range activities in the patterning of tissues neighbouring the cells that express these signals. For example, Sonic hedgehog (Shh), a member of the vertebrate Hh protein family expressed in the chordamesodermal notochord, induces floor-plate formation within the ectodermal neural tube apparently in a contact-dependent manner (Martí *et al.*, 1995; Roelink *et al.*, 1994; Tanabe *et al.*, 1995). Yet, in other cases, induction does not require cell contact. Thus, when neural plate or presomitic mesoderm explants were grown at a distance of many cell diameters from the Shh-expressing cells (Fan and Tessier-Lavigne, 1994; Martí *et al.*, 1995; Roelink *et al.*, 1995; Tanabe *et al.*, 1995) or when contact was prevented by a nucleopore filter between explants and Shh-expressing cells (Fan and Tessier-Lavigne, 1994; Tanabe *et al.*, 1995), induction of motor neurons or the sclerotome did still occur. Although Shh and different members of the vertebrate Hh protein family have also been implicated in the patterning of other vertebrate structures, *e.g.*, the limb (Riddle *et al.*, 1993; Chang *et al.*, 1994) and the eye (Ekker *et al.*, 1995b), it remains unknown if formation of these tissues are mediated by the Hh proteins in a contact-dependent or -independent manner. However, short- and long-range activities of Hh probably do not reflect fundamental differences between contact-dependent and -independent signalling mechanisms but simply differences in the response elicited by high and low concentrations of the Hh signal (Roelink *et al.*, 1995; see below).

Hh Signalling Acts through Two Parallel Pathways

Since the correct reception or transmission of the Hh signal is a recurrent theme in animal development, molecular dissection of the Hh pathway, which has been done extensively in *Drosophila*, would allow insight into this process. Transmission of the Hh signal depends on the *smoothened (smo)* gene product, which encodes a putative membrane protein (Alcedo *et al.*, 1996). Moreover, the possibility that Smo has characteristics of G protein-coupled receptors (Alcedo *et al.*, 1996) is further supported by the observation that the cyclic AMP-dependent protein kinase (PKA) acts antagonistically downstream of the Hh signal (Jiang and Struhl, 1995; Johnson *et al.*, 1995; Lepage *et al.*, 1995; Li *et al.*, 1995; Pan and Rubin, 1995; Strutt *et al.*, 1995). The *smo* gene has also been shown to act upstream of PKA by genetic analysis (Chen and Struhl, 1996; van den Heuvel and Ingham, 1996a), a result suggesting that Smo may activate an inhibitory G protein which inhibits adenylate cyclase and consequently PKA (Figure 1; Alcedo *et al.*, 1996).

Hh function is antagonised not only by PKA but also by Patched (Ptc) (Ingham *et al.*, 1991), a multiple-spanning membrane protein (Hooper and Scott, 1989; Nakano *et al.*, 1989). Since Hh-target genes, such as *dpp*, *wg*, or *ptc*, are expressed in *pka*$^-$ cells in the absence of functional Hh despite the presence of Ptc, PKA is epistatic to Ptc and thus must act either downstream of and/or in parallel to Ptc (Jiang and Struhl, 1995; Li *et al.*, 1995). Evidence for the second possibility has been obtained in elegant experiments employing a PKA transgene whose product is constitutively active (Jiang and Struhl, 1995; Li *et al.*, 1995). If Ptc acted exclusively through PKA, the level of constitutively active PKA needed to inhibit the expression of Hh-target genes in *pka*$^-$ cells should be the same as that in *ptc*$^-$ cells. However, compared to the levels required to suppress the expression of Hh-target genes in *pka*$^-$ cells, only very high levels of constitutively active PKA can inhibit the expression of the same genes in *ptc*$^-$ cells (Li *et al.*, 1995). Therefore, Ptc can antagonise the Hh signal in a PKA-independent pathway although it may also act through PKA, as illustrated by the model in Figure 1. Ptc seems to exert its inhibitory activity on the serine/threonine kinase Fused (Fu) (Préat *et al.*, 1990; Thérond *et al.*, 1996), whose function is required to activate the expression of Hh-target genes (Limbourg-Bouchon *et al.*, 1991; Forbes *et al.*, 1993). In contrast, PKA does not appear to act on Fu (Thérond *et al.*, 1996) and, thus, the Ptc- and PKA-dependent pathways of Hh signalling seem to converge downstream of Fu. Moreover, the observation that loss of PKA in the absence of the Hh signal activates Hh-target genes regardless of Ptc concentration suggests that the Ptc inhibition through Fu is not sufficient to inhibit the Hh signalling pathway (Li *et al.*, 1995). Conversely, only unphysiologically high concentrations of constitutively active PKA are able to repress Hh-target genes in the absence of Ptc (Li *et al.*, 1995). This indicates that the inhibitory action of PKA alone is also insufficient for repression of the Hh signalling pathway. Hence, both the Ptc-dependent and PKA-dependent pathways are required to repress Hh target genes (Figure 1). Since it is unclear if Ptc action on Fu is independent of Smo, the possibility also exists that both pathways depend on Smo, i.e., Hh signalling bifurcates downstream rather than upstream of Smo (Figure 1).

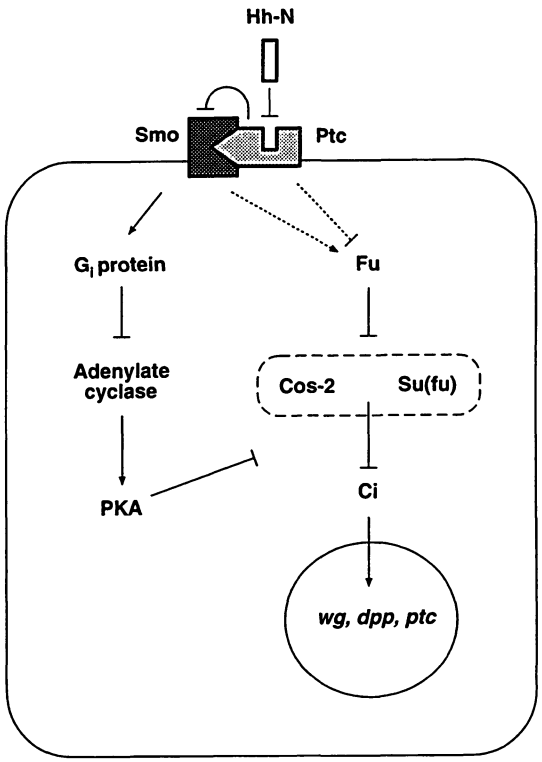


Fig. 1 A Model of the Hh Signalling Pathway.
To activate its target genes, *wg*, *dpp* and *ptc*, the N-terminal signalling domain of Hh, Hh-N, binds to its antagonist Ptc, which is complexed with Smo, to counteract the inhibition by Ptc of the constitutive signalling activity of Smo and of the Fu-branch of the signalling pathway. However, it remains unclear if the Ptc action on Fu is independent of or dependent on Smo as illustrated by the dotted lines. The left branch of the signalling pathway, namely that Smo inhibits PKA through activation of a G_i protein, which would lead to a decrease in cAMP levels, is inferred from the structural homology of Smo to G protein-coupled receptors. Thus far, no experimental evidence has been obtained for the proposed regulation of PKA activity. PKA may inhibit the Hh signal through stimulation of other antagonists of the pathway, Cos-2 or Su(fu), or through direct or indirect repression of the function of the zinc-finger protein Ci. Ptc antagonises the Hh signal not only in a PKA-dependent but also in a parallel, PKA-independent pathway through Fu. The two parallel pathways converge downstream of Fu, whose activity is required to relieve the repression mediated by Cos-2 and Su(fu) on Ci. Subsequently, Ci translocates into the nucleus to exert its action on Hh-target genes.

It is not known on which downstream component PKA acts. Through phosphorylation, it may destabilise factors essential in the activation of Hh-target genes, such as the zinc-finger transcription factor *Cubitus interruptus* (Ci) (Hidalgo and Ingham, 1990; Orenic *et al.*, 1990; Forbes

et al., 1993; Alexandre *et al.*, 1996; van den Heuvel and Ingham, 1996b). It is possible that the PKA-mediated inhibition of Ci is alleviated by the activity of a phosphatase that is activated by Smo (van den Heuvel and Ingham, 1996b). However, the mechanism proposed in Figure 1, namely inhibition of PKA by Smo through its modulation of cAMP levels via a G protein, is more economical because it is the simplest model to account for the predicted structural characteristics of Smo (Alcedo *et al.*, 1996). Recently, it has also been proposed that PKA phosphorylates the *Drosophila* homologue of the vertebrate transcription factor CREB2, which then binds to the *Drosophila* homologue of the coactivator CBP (dCBP) and thereby limits the amount of dCBP available for interaction with Ci to enhance its function (Akimaru *et al.*, 1997). On the other hand, instead of inhibiting Ci, PKA function may stimulate other antagonists of the Hh signal, such as Suppressor of fused [Su(fu)] (Préat, 1992) and Costal-2 (Cos-2) (Forbes *et al.*, 1993; Préat *et al.*, 1993).

The role of *Su(fu)* has been derived from the following observations. The absence of a functional *Su(fu)* gene product, a protein with a PEST sequence, has been shown to result in the suppression of the *fu* phenotype, *i.e.*, *Su(fu)* mutants can rescue the lethality of *fu* mutations (Préat, 1992; Pham *et al.*, 1995). Furthermore, homozygous *Su(fu)* mutants are viable and display a wild-type phenotype (Préat, 1992). Hence, the Su(fu) protein must inhibit a component downstream of Fu and the alleviation of such an inhibition must require the kinase activity of Fu, thereby allowing expression of Hh-target genes (Figure 1). Another component downstream of the Hh signal which is inhibited by Fu is the still unidentified gene product of *cos-2*, which also represses the expression of Hh-target genes (Forbes *et al.*, 1993; Préat *et al.*, 1993). Both Su(fu) and Cos-2 may block the action of Ci (Forbes *et al.*, 1993), while activation of Fu by the Hh signal might stimulate Ci function directly or indirectly (Figure 1; Forbes *et al.*, 1993; Motzny and Holmgren, 1995; Domínguez *et al.*, 1996). Whereas most of the components downstream of the Hh signal transduction pathway are found conserved in vertebrates (Concordet *et al.*, 1996; Goodrich *et al.*, 1996; Hammerschmidt *et al.*, 1996; Marigo *et al.*, 1996a, c; Stone *et al.*, 1996; Vortkamp *et al.*, 1996), vertebrate homologues of Fu or Su(fu) have not yet been identified.

The Hedgehog Signal Restricts Its Own Range of Action

The Smo protein was postulated to be a possible receptor for the Hh signal (Alcedo *et al.*, 1996) because of its homology to the *Drosophila* Frizzled (Fz) receptor protein (Vinson *et al.*, 1989) and its position in the Hh pathway. The demonstration by Bhanot *et al.* (1996) that the Fz protein and some of its close homologues bind the Wg protein and thus activate the Wg signalling pathway raises the intriguing possibility that the Hh and Wg proteins may recognise very similar receptors.

Consistent with the hypothesis that a receptor must be able to sequester its ligand, recent analysis in imaginal discs of the effect of *smo*$^-$ clones on Hh diffusion showed that its range is enhanced in the absence of Smo although this effect appears to be indirect because it is compensated by the simultaneous loss of PKA function (Chen and Struhl, 1996). This led the authors to suggest that loss of PKA would induce, even in the absence of Smo, a Hh-target gene whose product limits Hh diffusion. One of the Hh-target genes shown to be activated upon loss of PKA activity is *ptc* (Li *et al.*, 1995), whose product, a membrane protein, has been postulated previously to be the receptor of the Hh protein (Ingham *et al.*, 1991). Chen and Struhl (1996) demonstrated that the absence of Ptc causes the signal to spread further from its source, whereas upregulating Ptc expression restricts its movement. Thus, it is Ptc that sequesters the Hh signal in the absence of PKA and Smo. Therefore, it appears that Ptc and not Smo is the Hh receptor. However, the Hh signalling pathway is constitutively active in the combined absence of Ptc and Hh (Ingham *et al.*, 1991) yet is completely dependent on Smo (Hooper, 1994; Alcedo *et al.*, 1996). Considering that both Ptc and Smo are integral membrane proteins, the simplest explanation for these observations is that transduction of the Hh signal is mediated by a Ptc-Smo complex, in which the constitutive signalling activity of Smo is inhibited by its association with Ptc (Figures 1 and 2A; Alcedo *et al.*, 1996). Such a hypothesis is consistent with recent biochemical data (Stone *et al.*, 1996). In vertebrates, Ptc and Smo are coexpressed in many tissues and Ptc can form a complex with Smo (Stone *et al.*, 1996). In addition, Ptc (Marigo *et al.*, 1996b; Stone *et al.*, 1996), but not Smo (Stone *et al.*, 1996), binds the N-terminal domain of Hh (Hh-N).

The various Hh proteins, one identified in *Drosophila* (Nüsslein-Volhard and Wieschaus, 1980; Lee *et al.*, 1992; Mohler and Vani, 1992; Tabata *et al.*, 1992) and several in vertebrates (Echelard *et al.*, 1993; Krauss *et al.*, 1993; Riddle *et al.*, 1993; Chang *et al.*, 1994; Roelink *et al.*, 1994; Ekker *et al.*, 1995a, b), undergo not only signal peptide cleavage but also a further autoproteolytic cleavage into a 19 kDa amino-terminal peptide (Hh-N) and a carboxy-terminal peptide (Hh-C) that ranges in size from 26 to 28 kDa (Lee *et al.*, 1992, 1994; Chang *et al.*, 1994; Ekker *et al.*, 1995b; Lai *et al.*, 1995; Porter *et al.*, 1995), the cleavage being dependent on a conserved domain within their C-termini (Lee *et al.*, 1994; Porter *et al.*, 1995). Whereas Hh-C diffuses from the cell in which it was expressed, Hh-N remains associated with the cell surface (Lee *et al.*, 1994; Fan *et al.*, 1995; Porter *et al.*, 1995; Roelink *et al.*, 1995). Based on the spatial distribution patterns of Hh-N and Hh-C, the Hh-N fragment was postulated to mediate the short-range signalling activity of the protein, while the Hh-C fragment was thought to be the long-range signal (Lee *et al.*, 1994).

However, in *Drosophila* and vertebrates, Hh-C had no signalling activity whereas Hh-N was sufficient to exert both short- and long-range activities (Fan *et al.*, 1995; Fietz *et al.*, 1995; Martí *et al.*, 1995; Porter *et al.*, 1995;

Roelink *et al.*, 1995). Surprisingly, ubiquitous expression of a Hh-N or a full-length Hh transgene gave essentially the same dorsal and ventral cuticular defects in *Drosophila* as those produced by localised expression of a Hh-N transgene in cells that expressed the endogenous wild-type full-length protein (Porter *et al.*, 1996b). In contrast, the same localised expression of an exogenous full-length Hh protein generated dorsal and ventral cuticle similar to wild-type (Porter *et al.*, 1996b), a difference that may be explained by the observation that Hh-N, translated from a truncated Hh-coding region, is more freely diffusible than Hh-N derived from a transcript that includes the entire coding region (Roelink *et al.*, 1995; Porter *et al.*, 1996b). Hence, the Hh-C moiety is required in restricting the spatial distribution of the Hh-N signal within the extracellular matrix, a requirement decisive for the proper execution of the signal's patterning function. Indeed, Porter *et al.* (1996a, b) showed that the association of Hh-N with the cell surface requires the covalent attachment of the peptide to a cholesterol moiety in the membrane, a reaction catalysed by an intramolecular cholesterol transferase activity of Hh-C.

Furthermore, only Hh-N has signalling activity whereas mutant Hh that cannot undergo autoproteolytic cleavage is inactive (Porter *et al.*, 1996a). This implies that activation requires cleavage and, as a consequence, the simultaneous transfer of the resulting Hh-N fragment to cholesterol. There are two possibilities of how the signal could reach cells that are not in contact with the cells that display membrane-anchored Hh-N-cholesterol (Hh-N-ch) on their surface: either

(i) the Hh-N-ch fragment must diffuse from the membrane of the cell in which it is synthesised into the intercellular space (Porter *et al.*, 1996b) until captured by Ptc on the surface of adjacent and more distant cells or
(ii) Hh protein that has escaped autoproteolysis is secreted into the intercellular space, as it has been observed in S2 cells (Porter *et al.*, 1995).

In this second case, as Hh encounters its substrate cholesterol on the surface of neighbouring and more distant cells, immediate production of an active, cholesterol-linked fragment will ensue that is capable of binding to Ptc protein present on these cells. The crucial aspect is that this arrangement ensures that the range of signal is limited, which is required for its morphogenetic action, a condition that appears to be satisfied by the density of both Ptc and cholesterol molecules on the surface of the target cells.

The remaining uncertainties concern the form in which Hh is secreted, cleaved or uncleaved, the relative rates of diffusion within the intercellular space of Hh-N-ch and Hh and the affinity of Hh for cholesterol. Clearly, if the secreted form is the lipophilic Hh-N-ch fragment, binding to the membrane of target cells will be efficient and so will be binding to Ptc within the same membrane. However, if primarily hydrophilic Hh is secreted, its diffusion might not be sufficiently restricted to prevent long-range signalling, especially if the rate of reaction with cholesterol on the surface of the target cells is low. Indeed, the observation by Chen and Struhl (1996) that, in the absence of Ptc, the signal diffuses over many cell diameters, would be consistent with both of these assumptions, namely that primarily Hh is secreted and its reaction with target cell cholesterol is inefficient. The question then arises why the long-range signalling described by Chen and Struhl (1996) is prevented in the presence of Ptc despite the inefficiency of activation by cholesterol. The only way out of this dilemma would be to postulate that Hh itself binds to Ptc with an affinity that is much greater than that of its binding to the substrate cholesterol. Once retained in the membrane by Ptc, the probability of Hh colliding with membrane-bound cholesterol is greatly enhanced. This collision results in the activation of Hh that would then trigger the signal transduction.

The novel feature of this model is that binding of Hh to Ptc proceeds in two steps, a first, reversible binding that does not result in signalling, followed by the cleavage-induced allosteric transition of Hh-N-ch to a state that activates the two parallel signalling pathways. These events may be visualised to involve two sites on the extracellular surface of Ptc, one site, the Hh-binding site, that binds Hh until it reacts with cholesterol and a second site, the Hh-N-binding site, that binds the cleaved fragment Hh-N-ch to trigger the signalling pathway.

This model that could explain the observed diffusion of Hh in the absence of Ptc (Chen and Struhl, 1996) might not be compelling if the secretion product of the Hh-expressing cells is exclusively Hh-N-ch. Yet, because of the high affinity of cholesterol for membranes, diffusion over many cell diameters under these conditions seems unlikely.

While the basis for the enhanced range of Hh-N diffusion is the failure of its covalent linkage to the cell surface, the diffusion range of uncleaved Hh would be increased only when it fails to be bound by Ptc. Therefore, we would expect that mutations of Ptc exist that do not impede Hh diffusion but still repress the constitutive signalling of Smo. This contrasts with *ptc* null mutants, which fail to inhibit both Hh diffusion and constitutive Smo signalling, and with apparent null mutations of *ptc* that only impede Hh diffusion (Chen and Struhl, 1996).

In summary, then, we conclude that Hh restricts its own range of action via two novel mechanisms:

(i) Through the cholesterol transferase activity of Hh-C, which covalently links Hh-N to cholesterol and thus increases its affinity to the cell surface membrane (Porter *et al.*, 1996a, b); and
(ii) through the induction of its own antagonist, Ptc.

Furthermore, the antagonism by Ptc of the Hh signal features two distinct aspects:

(i) Sequestration of Hh limiting its diffusion (Chen and Struhl, 1996); and
(ii) inhibition of the constitutive signalling of Smo (Alcedo *et al.*, 1996).

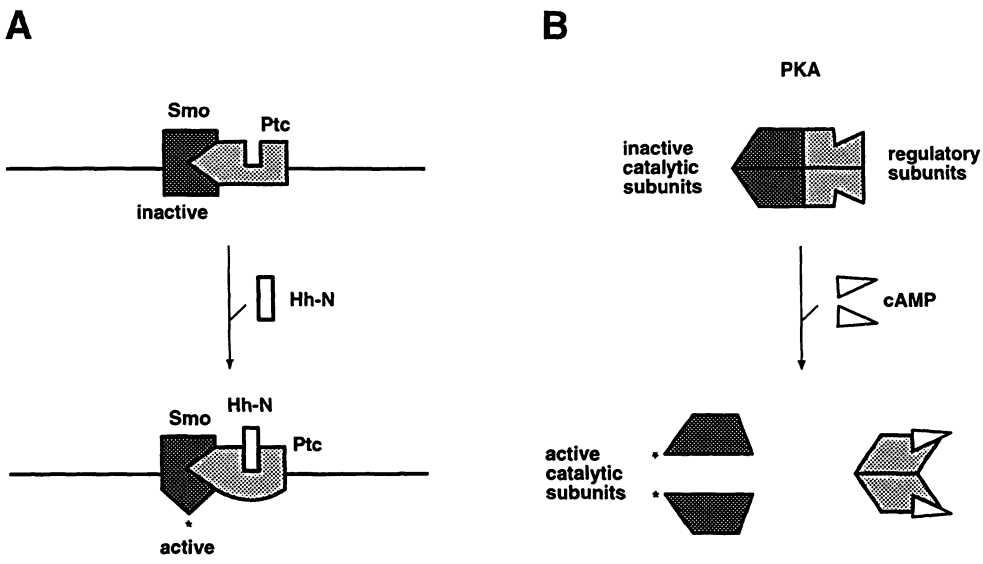


Fig. 2 Analogy between the Mode of Reception and Transmission of the Hh Signal at the Cell Surface and the Mode of cAMP-Mediated Regulation of PKA Activity within the Cytoplasm.
(A) Ptc maintains Smo in an inactive state within a Ptc-Smo complex in the absence of Hh. Upon binding of Hh-N to Ptc, Smo and Ptc undergo conformational changes that alter their state of activity. In this process, Smo may remain associated with, or dissociate from, the Ptc-Hh-N complex to assume an active state of signalling. In favor of the 'dissociation' model is the observation that constitutive signalling occurs in the absence, and hence independently, of Ptc while the model shown here is favored by the observation that Hh-N can be coimmunoprecipitated with a Ptc-Smo complex. (B) In the absence of cAMP, the regulatory subunits of PKA maintain its catalytic subunits in an inactive state. cAMP binding to its regulatory subunits releases the catalytic subunits, which become active to phosphorylate their substrates.

A Novel Mechanism of Signal Reception and Transmission at the Cell Surface

Since the signalling moiety of Hh, Hh-N, binds only to Ptc (Marigo *et al.*, 1996b; Stone *et al.*, 1996) and Ptc can bind Smo, Ptc might be a ligand-regulated repressor of a constitutively active signalling moiety, Smo, in a multi-component receptor complex (Figure 2A; Alcedo *et al.*, 1996; Stone *et al.*, 1996). While such a mechanism is new among cell surface receptors, it is analogous to long-known cytoplasmic signalling complexes (Gill and Garren, 1971; Brostrom *et al.*, 1971), of which indeed an example, the cAMP-dependent PKA, is encountered further downstream in the Hh-signalling pathway. The second messenger cAMP binds to the regulatory subunits of PKA to induce their dissociation from, and activation of, the catalytic subunits (Figure 2B). In contrast to the induction of PKA activity by cAMP, Smo may be activated by an allosteric transition without its dissociation from Ptc upon binding of Hh-N to Ptc (Figure 2A). Although not illustrated in Figure 2A, it is also conceivable that, in analogy to PKA, binding of Hh-N to Ptc shifts the equilibrium between the Ptc-Smo complex and its free components towards the dissociated state. While this second mechanism is favored by the observation that constitutive Smo signalling does not require Ptc (Hooper, 1994; Alcedo *et al.*, 1996), the first mechanism appears to be supported by the fact that Hh-N can be coimmunoprecipitated with a Ptc-Smo complex (Stone *et al.*, 1996). If Ptc does not act through Smo on Fu, a more prominent difference between the regulation of cytoplasmic PKA and that of the Smo signalling activity at the cell surface is that Ptc is more versatile than the regulatory subunit of PKA: Ptc does not act only on Smo but also regulates the signalling pathway parallel to Smo that passes through Fu (Figure 1).

However, this model of Hh signal reception and transmission appears to be inconsistent with the observation that *Drosophila* embryos mutant for both *hh* and *ptc* have a phenotype different from embryos lacking only a functional Ptc protein (Bejsovec and Wieschaus, 1993). This finding implies that Hh does not act through Ptc alone but also through another receptor whose activity it modulates to affect gene expression. For example, Hh might also bind to Smo to further stimulate the signalling function of Smo. Since Hh-N does not bind to Smo (Stone *et al.*, 1996), Hh might bind through its C-terminal portion to Smo. This C-terminal portion of Hh might have signalling properties that are undetectable in the absence of Hh-N (Fietz *et al.*, 1995; Porter *et al.*, 1995). According to this model, inhibition of Ptc by Hh-N is necessary for the upregulation of constitutive Smo activity by the C-terminal portion of Hh. However, it is not necessary to invoke such a model if the *ptc* alleles, used to compare *ptc* mutants with *ptc hh* double mutants (Bejsovec and Wieschaus, 1993), generate Ptc proteins that are not completely null in function. In this case, a strong *ptc* allele producing a weakly functional Ptc protein might still exert some inhibitory activity on both pathways in the absence, but no longer in the presence, of a functional Hh protein, and thus resemble a null allele in a wild-type but not in a *hh* mutant background.

Hh Primarily Functions as Short-Range Inducer of Secondary Long-Range Signals

As both short- and long-range signalling of Hh are attributable to Hh-N, which seems to remain predominantly associated with the cell membrane, it is possible that Hh primarily functions as a short-range inducer of secondary long-range signals, like Dpp and Wg (Nellen *et al.*, 1996; Zecca *et al.*, 1996), as had previously been proposed by Basler and Struhl (1994). The Wg and Dpp signals have been shown to have direct long-range activities because only ectopic expression of the signals, and not the ligand-independent ectopic activation of their downstream transducing components, resulted in nonautonomous repatterning (Nellen *et al.*, 1996; Zecca *et al.*, 1996). Therefore, the long-range patterning activities of both these signals are not mediated by the synthesis of secondary signals. In contrast, nonautonomous repatterning can be induced by either ectopic Hh expression or by the Hh-independent ectopic activation of its downstream transducing components because either event can induce secondary signals (Jiang and Struhl, 1995; Lepage *et al.*, 1995; Li *et al.*, 1995; Pan and Rubin, 1995; Chen and Struhl, 1996). Hence, these observations are consistent with the hypothesis that Hh mediates long-range patterning indirectly by its short-range activation of secondary signals (Zecca *et al.*, 1995; Nellen *et al.*, 1996). The long-range activity associated with Hh in the dorsal patterning of the *Drosophila* epidermis (Heemskerk and DiNardo, 1994) may also be mediated by other long-range signals, like Wg or the gene product of *lines*, which has not yet been identified, but not by Dpp (Bokor and DiNardo, 1996). In addition, the long-range activity of Hh observed in vertebrates during filter barrier experiments does not exclude the possibility that a secondary long-range signal is induced by Hh in an autocrine manner in the Hh-expressing cells.

Since the Hh-induced long-range activities of Dpp and Wg are also concentration-dependent, Dpp and Wg act as morphogens (Nellen *et al.*, 1996; Zecca *et al.*, 1996). Thus, the short-range induction properties of Hh are crucial for the generation of a morphogen source that is limited to a narrow band of cells adjacent to a region of Hh-expressing cells, which necessitated the evolution of the self-limiting properties of Hh that depend on Ptc and Smo. In other words, the strategy to specify a 'one-dimensional' morphogen source relies on a two-step mechanism. First, two adjacent and distinct areas are defined: the posterior compartment expresses a signalling protein, Hh; the anterior compartment another protein, Ptc. Subsequently, the interaction between the two proteins at the border of the two compartments limits the range of the signalling protein and thus defines a precise, 'one-dimensional' source of the morphogen (Lawrence and Struhl, 1996).

Aberrant Hh Signal Transduction in Cancer

Not only mutations in the *hh* gene but also defects in the reception or transmission of its signal can lead to disease in humans. Most notably, mutations in the *ptc* gene have been associated with the basal cell naevus syndrome (BCNS), which is characterised by developmental abnormalities and cancer, such as basal cell carcinoma (BCC; Hahn *et al.*, 1996; Johnson *et al.*, 1996). These mutations may lead to constitutive activation of Smo since a mutant Ptc protein, which is unable to bind to Smo, would fail to block its activity (Stone *et al.*, 1996). According to this view, Ptc acts as a tumour suppressor by suppressing the oncogenic activity of Smo. Furthermore, the transcription factor Ci, which is required for the activation of Hh-target genes, is the *Drosophila* homologue of the vertebrate family of Gli zinc-finger proteins, one of which has been associated with human glioblastomas due to its amplification in these tumours (Kinzler *et al.*, 1988; Hidalgo and Ingham, 1990; Orenic *et al.*, 1990; Forbes *et al.*, 1993; Alexandre *et al.*, 1996; Marigo *et al.*, 1996a; Vortkamp *et al.*, 1996). Thus, further elucidation of the Hh signalling pathway should be of benefit in counteracting diseases arising from its aberrant functions.

Acknowledgements

We thank H. Noll and K. Basler for their valuable criticism and comments on the manuscript.

References

Akimaru, H., Chen, Y., Dai, P., Hou, D.-X., Nonaka, M., Smolik, S.M., Armstrong, S., Goodman, R.H., and Ishii, S. (1997). *Drosophila* CBP is a co-activator of *cubitus interruptus* in *hedgehog* signalling. Nature *386*, 735–738.

Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M., and Hooper, J.E. (1996). The Drosophila *smoothened* gene encodes a seven-pass membrane protein, a putative receptor for the Hedgehog signal. Cell *86*, 221–232.

Alexandre, C., Jacinto, A., and Ingham, P.W. (1996). Transcriptional activation of *hedgehog* target genes in *Drosophila* is mediated directly by the Cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. Genes Dev. *10*, 2003–2013.

Basler, K., and Struhl, G. (1994). Compartment boundaries and the control of *Drosophila* limb pattern by *hedgehog* protein. Nature *368*, 208–214.

Bejsovec, A., and Wieschaus, E. (1993). Segment polarity gene interactions modulate epidermal patterning in *Drosophila* embryos. Development *119*, 501–517.

Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H.F., Donis-Keller, H., Helms, C., Hing, A.V., Heng, H.H.Q., Koop, B., Martindale, D., Rommens, J.M., Tsui, L.-C., and Scherer, S.W. (1996). Identification of *Sonic hedgehog* as a candidate gene responsible for holoprosencephaly. Nature Genetics *14*, 353–356.

Bhanot, P., Brink, M., Samos, C.H., Hsieh, J.-C., Wang, Y., Macke, J.P., Andrew, D., Nathans, J., and Nusse, R. (1996). A new member of the *frizzled* family from *Drosophila* functions as a Wingless receptor. Nature *382*, 225–230.

Bokor, P., and DiNardo, S. (1996). The roles of *hedgehog, wingless* and *lines* in patterning the dorsal epidermis in *Drosophila*. Development *122*, 1083–1092.

Brostrom, C.O., Corbin, J.D., King, C.A., and Krebs, E.G. (1971). Interaction of the subunits of adenosine 3′:5′-cyclic monophosphate-dependent kinase of muscle. Proc. Natl. Acad. Sci. USA *68*, 2444–2447.

Chang, D.T., López, A., von Kessler, D.P., Chiang, C., Simandl, B.K., Zhao, R., Seldin, M.F., Fallon, J.F., and Beachy, P.A. (1994). Products, genetic linkage and limb patterning activity of a murine *hedgehog* gene. Development *120*, 3339–3353.

Chen, Y., and Struhl, G. (1996). Dual roles for Patched in sequestering and transducing Hedgehog. Cell *87*, 553–563.

Concordet, J.-P., Lewis, K.E., Moore, J.W., Goodrich, L.V., Johnson, R.L., Scott, M.P., and Ingham, P.W. (1996). Spatial regulation of a zebrafish *patched* homologue reflects the roles of *Sonic hedgehog* and protein kinase A in neural tube and somite patterning. Development *122*, 2835–2846.

Domínguez, M., Brunner, M., Hafen, E., and Basler, K. (1996). Sending and receiving the Hedgehog signal: control by the Drosophila Gli protein Cubitus interruptus. Science *272*, 1621–1625.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., and McMahon, A.P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Cell *75*, 1417–1430.

Ekker, S.C., McGrew, L.L., Lai, C.-J., Lee, J.J., von Kessler, D.P., Moon, R.T., and Beachy, P.A. (1995a). Distinct expression and shared activities of members of the *hedgehog* gene family of *Xenopus laevis*. Development *121*, 2337–2347.

Ekker, S.C., Ungar, A.R., Greenstein, P., von Kessler, D.P., Porter, J.A., Moon, R.T., and Beachy, P.A. (1995b). Patterning activities of vertebrate *hedgehog* proteins in the developing eye and brain. Curr. Biol. *5*, 944–955.

Fan, C.-M., and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a Hedgehog homolog. Cell *79*, 1175–1186.

Fan, C.-M., Porter, J.A., Chiang, C., Chang, D.T., Beachy, P.A., and Tessier-Lavigne, M. (1995). Long-range sclerotome induction by Sonic hedgehog: direct role of the amino terminal cleavage product and modulation by the cyclic AMP signaling pathway. Cell *81*, 457–465.

Fietz, M.J., Jacinto, A., Taylor, A.M., Alexandre, C., and Ingham, P.W. (1995). Secretion of the amino-terminal fragment of the Hedgehog protein is necessary and sufficient for *hedgehog* signalling in *Drosophila*. Curr. Biol. *5*, 643–650.

Forbes, A.J., Nakano, Y., Taylor, A.M., and Ingham, P.W. (1993). Genetic analysis of *hedgehog* signalling in the *Drosophila* embryo. Development Suppl., pp. 115–124.

Gill, G.N., and Garren, L.D. (1971). Role of the receptor in the mechanism of action of adenosine 3′:5′-cyclic monophosphate. Proc. Natl. Acad. Sci. USA *68*, 786–790.

Goodrich, L.V., Johnson, R.L., Milenkovic, L., McMahon, J.A., and Scott, M.P. (1996). Conservation of the *hedgehog/patched* signaling pathway from flies to mice: induction of a mouse *patched* gene by Hedgehog. Genes Dev. *10*, 301–312.

Hahn, H., Wicking, C., Zaphiropoulos, P.G., Gailani, M.R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A.B., Gillies, S., Negus, K., Smyth, I., Pressman, C., Leffel, D.J., Gerrard, B., Goldstein, A.M., Dean, M., Toftgard, R., Chenevix-Trench, G., Wainwright, B., and Bale, A.E. (1996). Mutations of the human homolog of Drosophila *patched* in the nevoid basal cell carcinoma syndrome. Cell *85*, 841–851.

Hammerschmidt, M., Bitgood, M.J., and McMahon, A.P. (1996). Protein kinase A is a common negative regulator of Hedgehog signaling in the vertebrate embryo. Genes Dev. *10*, 647–658.

Heberlein, U., Singh, C.M., Luk, A.Y., and Donohoe, T.J. (1995). Growth and differentiation in the *Drosophila* eye coordinated by *hedgehog*. Nature *373*, 709–711.

Heemskerk, J., and DiNardo, S. (1994). Drosophila *hedgehog* acts as a morphogen in cellular patterning. Cell *76*, 449–460.

Hidalgo, A., and Ingham, P. (1990). Cell patterning in the *Drosophila* segment: spatial regulation of the segment polarity gene *patched*. Development *110*, 291–301.

Hooper, J.E. (1994). Distinct pathways for autocrine and paracrine Wingless signalling in *Drosophila* embryos. Nature *372*, 461–464.

Hooper, J.E., and Scott, M.P. (1989). The Drosophila *patched* gene encodes a putative membrane protein required for segmental patterning. Cell *59*, 751–765.

Ingham, P.W., and Hidalgo, A. (1993). Regulation of *wingless* transcription in the *Drosophila* embryo. Development *117*, 283–291.

Ingham, P.W., Taylor, A.M., and Nakano, Y. (1991). Role of the *Drosophila patched* gene in positional signalling. Nature *353*, 184–187.

Jiang, J., and Struhl, G. (1995). Protein kinase A and Hedgehog signaling in Drosophila limb development. Cell *80*, 563–572.

Johnson, R.L., Grenier, J.K., and Scott, M.P. (1995). *patched* overexpression alters wing disc size and pattern: transcriptional and post-transcriptional effects on *hedgehog* targets. Development *121*, 4161–4170.

Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bare, J.W., Bonifas, J.M., Quinn, A.G., Myers, R.M., Cox, D.R., Epstein, Jr., E.H., and Scott, M.P. (1996). Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. Science *272*, 1668–1671.

Kinzler, K.W., Ruppert, J.M., Bigner, S.H., and Vogelstein, B. (1988). The *GLI* gene is a member of the *Krüppel* family of zinc finger proteins. Nature *332*, 371–374.

Krauss, S., Concordet, J.-P., and Ingham, P.W. (1993). A functionally conserved homolog of the Drosophila segment polarity gene *hh* is expressed in tissues with polarizing activity in zebrafish embryos. Cell *75*, 1431–1444.

Lai, C.-J., Ekker, S.C., Beachy, P.A., and Moon, R.T. (1995). Patterning of the neural ectoderm of *Xenopus laevis* by the amino-terminal product of hedgehog autoproteolytic cleavage. Development *121*, 2349–2360.

Lawrence, P.A., and Struhl, G. (1996). Morphogens, compartments, and pattern: lessons from Drosophila? Cell *85*, 951–961.

Lee, J.J., von Kessler, D.P., Parks, S., and Beachy, P.A. (1992). Secretion and localized transcription suggest a role in positional signaling for products of the segmentation gene *hedgehog*. Cell *71*, 33–50.

Lee, J.J., Ekker, S.C., von Kessler, D.P., Porter, J.A., Sun, B.I., and Beachy, P.A. (1994). Autoproteolysis in *hedgehog* protein biogenesis. Science *266*, 1528–1537.

Lepage, T., Cohen, S.M., Diaz-Benjumea, F.J., and Parkhurst, S.M. (1995). Signal transduction by cAMP-dependent protein kinase A in *Drosophila* limb patterning. Nature *373*, 711–715.

Li, W., Ohlmeyer, J.T., Lane, M.E., and Kalderon, D. (1995). Function of protein kinase A in Hedgehog signal transduction and Drosophila imaginal disc development. Cell *80*, 553–562.

Limbourg-Bouchon, B., Busson, D., and Lamour-Isnard, C. (1991). Interactions between *fused*, a segment-polarity gene in *Drosophila*, and other segmentation genes. Development *112*, 417–429.

Marigo, V., Johnson, R.L., Vortkamp, A., and Tabin, C.J. (1996a). Sonic hedgehog differentially regulates expression of Gli and Gli3 during limb development. Dev. Biol. *180*, 273–283.

Marigo, V., Davey, R.A., Zuo, Y., Cunningham, J.M., and Tabin, C.J. (1996b). Biochemical evidence that Patched is the Hedgehog receptor. Nature *384*, 176–179.

Marigo, V., Scott, M.P., Johnson, R.L., Goodrich, L.V., and Tabin, C.J. (1996c). Conservation in *hedgehog* signaling: induction of a chicken *patched* homolog by *Sonic hedgehog* in the developing limb. Development *122*, 1225–1233.

Martí, E., Bumcrot, D.A., Takada, R., and McMahon, A.P. (1995). Requirement of 19 K form of Sonic hedgehog for induction of distinct ventral cell types in CNS explants. Nature *375*, 322–325.

Mohler, J., and Vani, K. (1992). Molecular organization and embryonic expression of the *hedgehog* gene involved in cell-cell communication in segmental patterning in *Drosophila*. Development *115*, 957–971.

Motzny, C.K., and Holmgren, R. (1995). The *Drosophila* cubitus interruptus protein and its role in the *wingless* and *hedgehog* signal transduction pathways. Mech. Dev. *52*, 137–150.

Nakano, Y., Guerrero, I., Hidalgo, A., Taylor, A., Whittle, J.R.S., and Ingham, P.W. (1989). A protein with several possible membrane-spanning domains encoded by the *Drosophila* segment polarity gene *patched*. Nature *341*, 508–513.

Nellen, D., Burke, R., Struhl, G., and Basler, K. (1996). Direct and long-range action of a DPP morphogen gradient. Cell *85*, 357–368.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. Nature *287*, 795–801.

Orenic, T.V., Slusarski, D.C., Kroll, K.L., and Holmgren, R.A. (1990). Cloning and characterization of the segment polarity gene *cubitus interruptus Dominant* of *Drosophila*. Genes Dev. *4*, 1053–1067.

Pan, D., and Rubin, G.M. (1995). cAMP-dependent protein kinase and *hedgehog* act antagonistically in regulating *decapentaplegic* transcription in Drosophila imaginal discs. Cell *80*, 543–552.

Pham, A., Thérond, P., Alves, G., Blanchet Tournier, F., Busson, D., Lamour-Isnard, C., Limbourg Bouchon, B., Préat, T., and Tricoire, H. (1995). The *Suppressor of fused* gene encodes a novel PEST protein involved in Drosophila segment polarity establishment. Genetics *140*, 587–598.

Porter, J.A., von Kessler, D.P., Ekker, S.C., Young, K.E., Lee, J.J., Moses, K., and.Beachy, P.A. (1995). The product of *hedgehog* autoproteolytic cleavage active in local and long-range signalling. Nature *374*, 363–366.

Porter, J.A., Young, K.E., and Beachy, P.A. (1996a). Cholesterol modification of Hedgehog signaling proteins in animal development. Science *274*, 255–259.

Porter, J.A., Ekker, S.C., Park, W.-J., von Kessler, D.P., Young, K.E., Chen, C.-H., Ma, Y., Woods, A.S., Cotter, R.J., Koonin, E.V., and Beachy, P.A. (1996b). Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. Cell *86*, 21–34.

Préat, T. (1992). Characterization of *Suppressor of fused*, a complete suppressor of the *fused* segment polarity gene of *Drosophila melanogaster*. Genetics *132*, 725–736.

Préat, T., Thérond, P., Lamour-Isnard, C., Limbourg-Bouchon, B., Tricoire, H., Erk, I., Mariol, M.-C., and Busson, D. (1990). A putative serine/threonine protein kinase encoded by the segment-polarity *fused* gene of *Drosophila*. Nature *347*, 87–89.

Préat, T., Thérond, P., Limbourg-Bouchon, B., Pham, A., Tricoire, H., Busson, D., and Lamour-Isnard, C. (1993). Segmental polarity in *Drosophila melanogaster*: genetic dissection of *fused* in a *Suppressor of fused* background reveals interaction with *costal-2*. Genetics *135*, 1047–1062.

Riddle, R.D., Johnson, R.L., Laufer, E., and Tabin, C. (1993). *Sonic hedgehog* mediates the polarizing activity of the ZPA. Cell *75*, 1401–1416.

Roelink, H., Augsburger, A., Heemskerk, J., Korzh, V., Norlin, S., Ruiz i Altaba, A., Tanabe, Y., Placzek, M., Edlund, T., Jessell, T.M., and Dodd, J. (1994). Floor plate and motor neuron induction by *vhh-1*, a vertebrate homolog of *hedgehog* expressed by the notochord. Cell *76*, 761–775.

Roelink, H., Porter, J.A., Chiang, C., Tanabe, Y., Chang, D.T., Beachy, P.A., and Jessell, T.M. (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of Sonic hedgehog autoproteolysis. Cell *81*, 445–455.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S.W., Tsui, L.-C., and Muenke, M. (1996). Mutations in the human *Sonic hedgehog* gene cause holoprosencephaly. Nature Genetics *14*, 357–360.

Stone, D.M., Hynes, M., Armanini, M., Swanson, T.A., Gu, Q., Johnson, R.L., Scott, M.P., Pennica, D., Goddard, A., Phillips, H., Noll, M., Hooper, J.E., de Sauvage, F., and Rosenthal, A. (1996). The tumour-suppressor gene *patched* encodes a candidate receptor for Sonic hedgehog. Nature *384*, 129–134.

Strutt, D.I., Wiersdorff, V., and Mlodzik, M. (1995). Regulation of furrow progression in the *Drosophila* eye by cAMP-dependent protein kinase A. Nature *373*, 705–709.

Tabata, T., and Kornberg, T.B. (1994). Hedgehog is a signaling protein with a key role in patterning Drosophila imaginal discs. Cell *76*, 89–102.

Tabata, T., Eaton, S., and Kornberg, T.B. (1992). The *Drosophila hedgehog* gene is expressed specifically in posterior compartment cells and is a target of *engrailed* regulation. Genes Dev. *6*, 2635–2645.

Tanabe, Y., Roelink, H., and Jessell, T.M. (1995). Induction of motor neurons by Sonic hedgehog is independent of floor plate differentiation. Curr. Biol. *5*, 651–658.

Thérond, P.P., Knight, J.D., Kornberg, T.B., and Bishop, J.M. (1996). Phosphorylation of the fused protein kinase in response to signaling from hedgehog. Proc. Natl. Acad. Sci. USA *93*, 4224–4228.

van den Heuvel, M., and Ingham, P.W. (1996a). *smoothened* encodes a receptor-like serpentine protein required for *hedgehog* signalling. Nature *382*, 547–551.

van den Heuvel, M., and Ingham, P.W. (1996b). 'Smoothening' the path for hedgehogs. Trends Cell Biol. *6*, 451–453.

Vinson, C.R., Conover, S., and Adler, P.N. (1989). A *Drosophila* tissue polarity locus encodes a protein containing seven potential transmembrane domains. Nature *338*, 263–264.

Vortkamp, A., Lee, K., Lanske, B., Segre, G.V., Kronenberg, H.M., and Tabin, C.J. (1996). Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. Science *273*, 613–622.

Zecca, M., Basler, K., and Struhl, G. (1995). Sequential organizing activities of engrailed, hedgehog, and decapentaplegic in the *Drosophila* wing. Development *121*, 2265–2278.

Zecca, M., Basler, K., and Struhl, G. (1996). Direct and long-range action of a Wingless morphogen gradient. Cell *87*, 833–844.

Biol. Chem., Vol. 378, pp. 591 – 597, July 1997

Review

Telomere Length Regulation: Getting the Measure of Chromosome Ends

David Shore

Department of Molecular Biology, University of Geneva, Sciences II, 30, quai Ernest-Ansermet, CH-1211 Geneva 4, Switzerland

Telomeres, the protein-DNA complexes that comprise the ends of linear eukaryotic chromosomes, serve to protect the chromosome ends and allow their complete replication. Telomeres also appear to play an essential role in chromosome segregation. In most organisms telomeric DNA consists of a series of short repeats that are variable in length, but regulated at a fixed average value in the germline. The possible involvement of telomere repeat shortening in aging and carcinogenesis has recently attracted attention to the more basic question of how telomere length is sensed and regulated by the cell. Telomere length in the budding yeast *Saccharomyces cerevisiae* has been known for over a decade now to be under complex genetic control, and this organism has provided a useful model system to address basic mechanistic questions. This review focuses on recent studies in yeast which indicate that the double-strand telomere-repeat binding protein Rap1 may play an important role in a negative-feedback mechanism that senses and controls the length of the telomere repeats. Although the same carboxy-terminal domain of Rap1p is involved in both telomere length regulation and telomeric silencing (telomere position effect), it appears that these two functions are mediated by separate sets of Rap1p-interacting proteins. Results from other systems suggest that negative regulation of telomere elongation by a double-stranded telomere-repeat binding protein may be a highly conserved strategy for telomere length control.

Key words: DNA-binding / Length regulation / Rap1p / Rif proteins / Sir proteins / Telomerase / Telomere / Transcriptional silencing / Taz1 / TRF1.

Introduction

The partitioning of eukaryotic genomes into multiple linear chromosomes has clearly proven to be a successful strategy for the faithful replication and segregation of DNA required for cell division. However, the ends of chromosomes create at least two problems for the cell. First, DNA polymerases are unable to completely replicate ends due to the requirement for an RNA primer on the lagging strand (Watson, 1972). In addition, free DNA ends could be susceptible to exonucleolytic degradation or to recombination reactions that would create unstable di-centric chromosomes. Telomeres, the specialized protein-DNA complexes that comprise the ends of linear chromosomes have thus evolved to deal with these problems [reviewed recently in (Lingner *et al.*, 1995; Zakian, 1996)]. In addition, it is now becoming clear that telomeres serve other functions important for chromosome segregation. In yeast telomeres have been shown to be essential for chromosome stability (Sandell and Zakian, 1993), and recent studies in the ciliate *Tetrahymena thermophila* indicate a specific role for telomeres in mitosis (Kirk *et al.*, 1997).

In most eukaryotes, chromosomes terminate in many copies of a simple repeat sequence [$(T_2AG_3)_n$ in mammals]. The total length of these telomeric repeat tracts can vary in different cells or tissues of an organism, but is centered about a fixed, species-specific length in the germline. Telomeric repeat sequences are generated by a special reverse transcriptase, called telomerase, which carries its own RNA template. Two types of observations have recently created a considerable amount of interest in telomeres and the telomerase enzyme [reviewed in (De Lange, 1995; Harley and Villeponteau, 1995)]. First, as many organisms age the telomeres in their somatic cells gradually shorten, apparently due to the absence of telomerase. This process, which could eventually lead to cell death, has been proposed to be a molecular clock that plays an important role in determining an organism's life span. Many cancer cells, however, contain active telomerase and appear to have stabilized telomere repeats. It has been proposed, therefore, that telomerase re-activation is an obligatory step in tumorigenesis that allows cancer cells to divide without limit. Now that the gene encoding the RNA template of a mouse telomerase enzyme has been cloned (Blasco *et al.*, 1995), it should be possible to test these ideas rigorously. In the meantime, however, these observations have focused attention on the importance of understanding the more basic question of how cells sense and regulate the lengths of their telomeres, which will be the subject of this short review. I will emphasize very recent results, primarily from work on yeast; the reader is referred to an excellent review for a more detailed discussion of previous work on this problem (Greider, 1996).

The Yeast Telomere DNA-Binding Protein Rap1: A Negative Regulator of Telomere Elongation

The budding yeast *Saccharomyces cerevisiae* has proven a useful model system to study telomere length regulation. Telomeres in this yeast terminate in an irregular repeat (usually abbreviated TG_{1-3}) whose generation and maintenance requires a telomerase enzyme (Singer and Gottschling, 1994). Although telomere repeat tracts vary in length within a culture of yeast cells, they do so about an average value that is characteristic of a given strain (~ 300 – 400 bp for most laboratory strains) and genetically determined (Lustig and Petes, 1986). Interestingly, a number of genes have been identified which when mutated result in either a decrease or an increase in this average tract length [reviewed in (Zakian 1996)]. These studies have suggested that regulation of telomere repeat length might have important consequences for the cell. For example, mutations that lead to elongated telomere tracts are often associated with decreased fidelity of chromosome segregation (Conrad *et al.*, 1990; Kyrion *et al.*, 1992; Wotton and Shore, 1997). Furthermore, it has been shown that alterations in telomere tract length can influence the expression of genes elsewhere in the chromosome (Buck and Shore, 1995), by a mechanism that probably involves sequestration of silencing factors [the Sir proteins (Lustig *et al.*, 1996; Maillet *et al.*, 1996; Marcand *et al.*, 1996)]. The modulation of silencing throughout the genome by telomeres might have a direct effect on cell life span in yeast, which is known to be controlled, at least in part, by the Sir silencing factors (Kennedy *et al.*, 1995).

The components of a yeast telomerase protein-RNA complex, which are likely to play a central role in telomere length regulation, are just now being identified and characterized (Lendvay *et al.*, 1996; Lingner *et al.*, 1997; Singer and Gottschling, 1994). Despite this fact, significant molecular insights into telomere length regulation have resulted from the identification of a double-strand telomere DNA-binding protein (Berman *et al.*, 1986; Buchman *et al.*, 1988a, b), the product of the essential *RAP1* gene (Shore and Nasmyth, 1987). In addition to its very high affinity TG_{1-3} binding sites [which occur at a frequency of roughly once per 18 bp in telomere repeat tracts; (Gilson *et al.*, 1993)], Rap1p also binds to a large number of promoter elements and to mating-type gene silencers [reviewed in (Shore, 1994)]. Genetic analysis has shown that Rap1p can function either as an activator or repressor of transcription, depending upon the context of its bindig site (Kurtz and Shore, 1991). The telomeric role of Rap1p was initially somewhat puzzling, until it was shown that telomeres themselves can exert a position-effect on transcription with many of the same features as mating-type gene silencing (Aparicio *et al.*, 1991; Gottschling *et al.*, 1990).

Although it was also clear from other genetic studies that Rap1p is involved in telomere length regulation, its precise role in this process was difficult to infer from the phenotypes of the first point mutations that were identified. For example, ts-lethal mutations in *RAP1* cause reversible telomere shortening when grown at semi-permissive temperatures (Conrad *et al.*, 1990; Lustig *et al.*, 1990), whereas silencing-defective (*rap1^s*) mutants result in telomere repeat tract elongation (Sussel and Shore, 1991). The isolation of a set of intra-genic suppressors of a *rap1^{ts}* mutant, the so-called *rap1^t* mutants, helped to clarify this situation (Kyrion *et al.*, 1992). All of the *rap1^t* mutants lead to carboxy-terminal truncations and result in essentially unregulated telomere repeat tract elongation, as well as decreased chromosome stability and a complete loss of telomere position effect (TPE) (Kyrion *et al.*, 1993). The simplest interpretation of these results is that the carboxy-terminal domain of Rap1p functions as both a negative regulator of telomere elongation and of transcription. This conclusion is consistent with earlier studies which showed that the introduction of extra telomere repeat sequences into yeast cells on plasmids causes the chromosomal telomere repeats to elongate (Runge and Zakian, 1989), implying that a limiting factor which binds to the TG_{1-3} repeats (either directly or indirectly) negatively regulates their elongation.

A different approach, which utilized mutations in the telomerase template RNA of the related yeast *K. lactis*, has also lead to the conclusion that Rap1p is a negative regulator of telomere elongation (Krauskopf and Blackburn, 1996; McEachern and Blackburn, 1995). In these experiments, telomerase RNA mutations are introduced into cells, which results in the gradual replacement of distal telomere repeats with mutant repeats that are known from *in vitro* experiments to weaken or abolish Rap1p binding. These mutant template RNAs result in either a delayed or rapid expansion of the telomere repeat tracts (depending upon their Rap1p binding defect), indicating that strong Rap1p binding is essential to prevent unregulated telomere growth.

The two different lines of evidence indicating that Rap1p is a negative regulator of telomere elongation are summarized diagrammatically in Figure 1.

Rap1-Interacting Factors: Separate Roles in Silencing and Telomere Length Regulation?

Subsequent studies pointed to the existence of proteins which interact with the Rap1p carboxyl terminus to control both telomere length and telomeric silencing. Overexpression of this domain of Rap1p in the absence of its (centrally-located) DNA-binding domain leads to telomere elongation (Conrad *et al.*, 1990; Hardy, 1991), as well as a loss of silencing (Hardy *et al.*, 1992a). The first candidate for such a protein, called Rif1p (**R**ap1p-**i**nteracting **f**actor 1) was identified in a two-hybrid screen employing the Rap1p carboxyl terminus fused to the Gal4p DNA-binding domain (Hardy *et al.*, 1992b). Mutation of the *RIF1* gene results in modest telomere elongation (from 300 to 450 – 600 bp), as would be predicted for a protein that helps to mediate the telomere length regulation function of Rap1p.

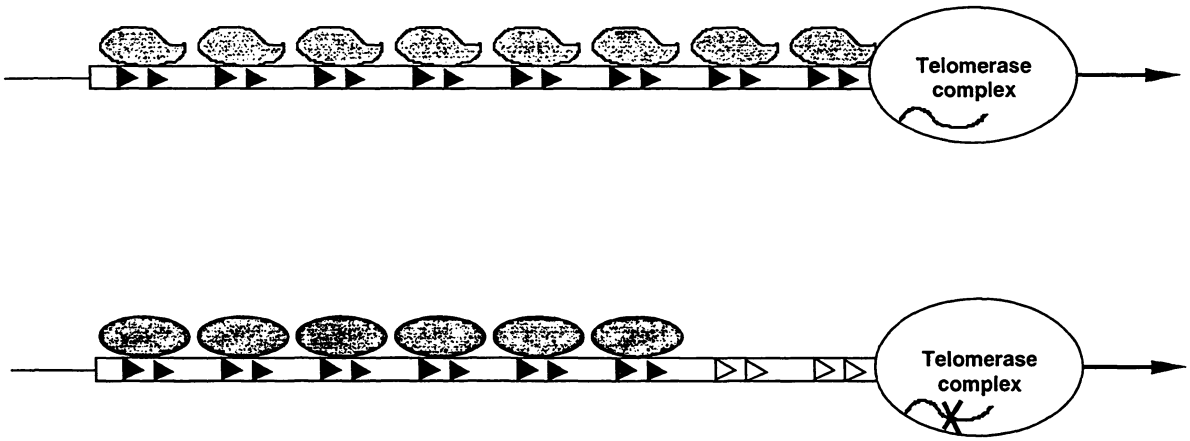


Fig. 1 Rap1p Is a Negative Regulator of Telomere Elongation.
Mutations that truncate the Rap1p carboxyl terminus (top) result in essentially uncontrolled telomere elongation (Kyrion *et al.*, 1992). Conversely, mutations in the telomerase RNA template that result in the incorporation of telomere repeat tracts to which Rap1p cannot bind (bottom) also result in uncontrolled telomere growth (Krauskopf and Blackburn, 1996; McEachern and Blackburn, 1995). Wild-type telomere-repeat Rap1p binding sites are indicated by pairs of filled arrowheads, mutant repeats by open arrowheads.

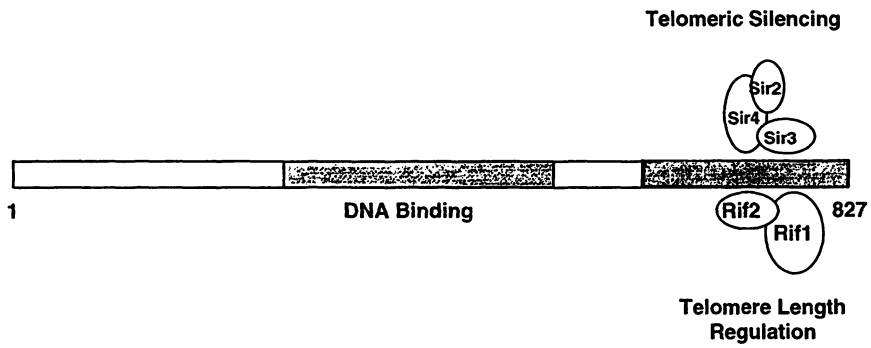


Fig. 2 Protein-Protein Interactions at the Rap1p Carboxyl Terminus Implicated in Telomeric Silencing and Telomere Length Regulation.
A linear representation of the Rap1 protein (827 amino acids) is shown, indicating the centrally-located Myb-like DNA-binding domain (Konig *et al.*, 1996) and the carboxy-terminal domain involved in TPE and telomere length regulation (approximately amino acids 667–827 and 653–827, respectively; Buck and Shore, 1995 and S. Marcand and D. Shore, unpublished results).

Strikingly, *rap1^s* mutants, which display a similar telomere elongation phenotype to *rif1* mutants, encode a protein that is defective in the Rif1p two-hybrid interaction. More recently, a second such factor (Rif2p) has been identified (Wotton and Shore, 1997). Significantly, cells lacking both Rif1p and Rif2p have extremely elongated telomeres (~1–3 kb) which are indistinguishable from those in mutant cells (e.g. the *rap1^t* mutant *rap1–17*) lacking the Rap1p carboxyl terminus (Wotton and Shore, 1997). Taken together, these data are consistent with a model in which Rif1p and Rif2p (which also interact with each other in the two-hybrid system) form a complex that is capable of regulating telomere length when recruited to telomeres by the Rap1p carboxyl terminus.

Although the Rap1p carboxyl terminus carries out both telomeric silencing and telomere length regulatory functions, it appears that these are mediated by different, and in fact opposing sets of protein-protein interactions. The Rap1p carboxyl terminus interacts with both Sir3p and Sir4p (Cockell *et al.*, 1995; Moretti *et al.*, 1994), both of which are required for TPE (Aparicio *et al.*, 1991). However, *sir3* or *sir4* mutations, unlike *rif* mutations, actually result in (slight) telomere shortening (Palladino *et al.*, 1993), suggesting that the wild-type function of these genes somehow promotes telomere elongation. In addition, *sir* and *rif* mutations have opposite effects on TPE: *sir3* and *sir4* mutants are completely defective in telomeric silencing (Aparicio *et al.*, 1991), whereas *rif1* and *rif2* mutations lead to a synergistic increase in TPE (Kyrion *et al.*, 1993; Wotton and Shore, 1997). Furthermore, targeting of either Sir3p or Sir4p to a telomere (by fusing the proteins to a DNA binding domain) can restore silencing in cells carying mutations in the Rap1p carboxyl terminus (Lustig *et al.*, 1996; Marcand *et al.*, 1996), but fails to restore telomere length regulation (Marcand *et al.*, 1997). Finally, mutant alleles of *RAP1*, the *rap1^s* mutants, appear to separate the silencing and telomere length regulatory functions of the carboxy-terminal domain. *Rap1^s* mutants actually display improved telomeric silencing compared to wild-type (Buck and Shore, 1995), but, as pointed out above, these mutations lead to a partial loss of telomere length regulation (Sussel and Shore, 1991). The dual function of the Rap1p carboxyl

terminus and the separate roles of the interacting complexes of Sir and Rif proteins is indicated schematically in Figure 2.

It should be pointed out that although there is now ample direct evidence that Sir proteins are bound at telomeres (Hecht *et al.*, 1995, 1996) through interactions with both Rap1p and histone H3 and H4 amino termini, this issue has not yet been addressed with regard to the Rif proteins. It is still possible that their action at telomeres is indirect. In addition, although there is evidence that Sir2p interacts with Sir4p (Moazed and Johnson, 1996; Moazed *et al.*, 1997) and is localized at telomeres (Strahl-Bolsinger *et al.*, 1997) where it is required for TPE, its role in telomere length regulation in *S. cerevisiae* has not yet been explored. Interestingly, mutation of a *SIR2* homologue in *K. lactis* has been reported to cause telomere elongation in combination with a carboxy-terminal *RAP1* mutation (Krauskopf and Blackburn, 1996).

A Role for Rap1p in Sensing Telomere Length

The fact that telomeres in yeast (and many other organisms) have a characteristic average length implies the existence of a mechanism that can sense telomere length and use that information to regulate either the rate of elongation, shortening, or both. Although this fundamental aspect of telomere length regulation is far from being clearly understood, recent studies suggest that the Rap1 protein might play a critical role in a mechanism that senses telomere length (Marcand *et al.*, 1997). Two lines of evidence are consistent with the idea that Rap1p (and specifically its carboxyl terminus) functions as part of a simple negative-feedback mechanism to control telomere length. To begin with, experiments in which telomeres are generated *de novo* during yeast transformation and homologous recombination indicate that cells regulate TG_{1-3} tract length irrespective of repeat orientation. This result suggests that it is the amount of protein binding to the repeats (presumably Rap1p) rather than some other feature of these sequences that is 'counted' and regulated to center about a fixed value. A more direct demonstration of this point comes from experiments in which additional Rap1p carboxy termini are added near the proximal end of a telomere (using Gal4 DNA-binding domain/Rap1p hybrid proteins). These artificially added Rap1p molecules lead to a loss of TG_{1-3} sequences from the distal end of the targeted telomere, such that the total number of Rap1p carboxy termini at that telomere is kept constant.

These data suggest the outlines of a simple negative-feedback model for telomere length regulation by Rap1p (see Figure 3). The central feature of this model is the notion that the Rap1p molecules bound at the chromosome end can create a number-dependent molecular switch to control telomere elongation. The basic proposition of this model is that a threshold number of telomere repeats, and

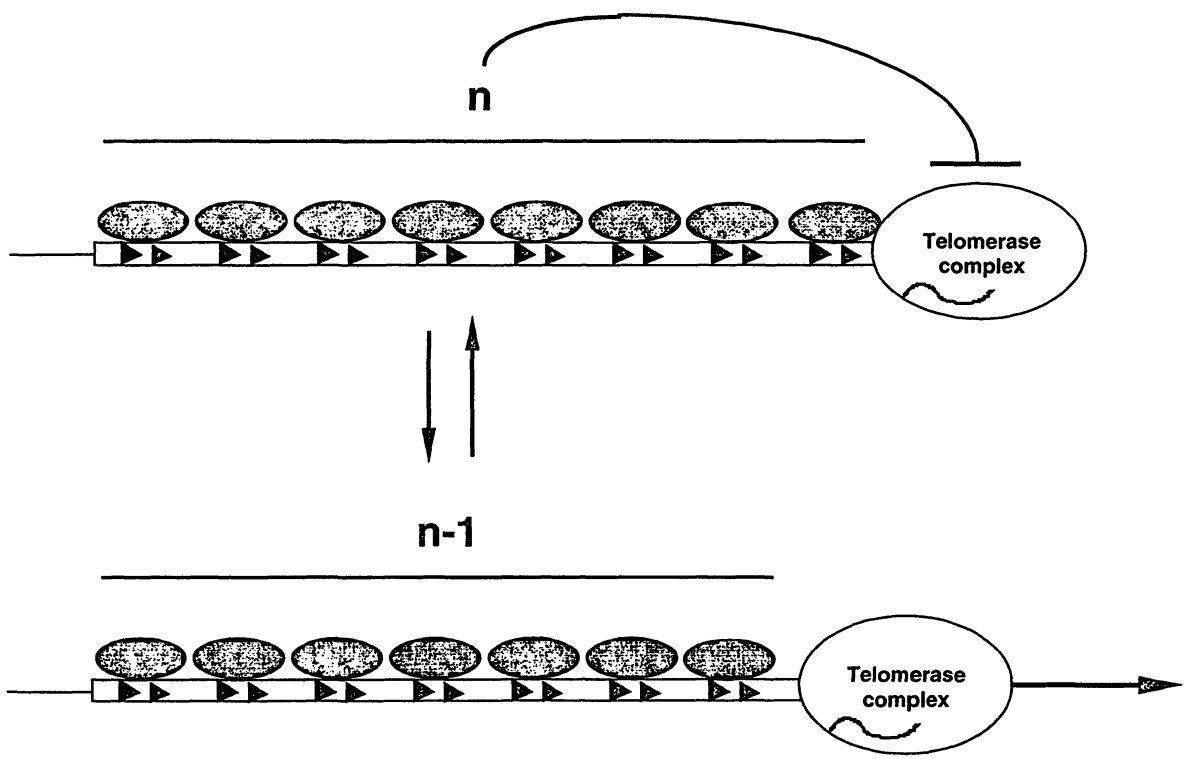


Fig. 3 Outline of a Negative-Feedback Model for Telomere Length Regulation by Rap1p in Yeast.
See text for discussion. The number ('n') of Rap1 molecules 'counted' may correspond to those proteins specifically bound by the Rif1p/Rif2p complex (Marcand *et al.*, 1997; Wotton and Shore, 1997). Conversely, those molecules interacting with Sir complexes, which are implicated in the initiation of telomeric silencing over adjacent nucleosomes (to the left, not indicated in the figure) may be excluded from the length regulation mechanism.

hence bound Rap1p molecules, will generate a signal that blocks telomere elongation (presumably by directly or indirectly regulating the telomerase enzyme). When this number drops below the threshold value, due to incomplete replication in the absence of telomerase activity (or to recombination or nucleolytic activities), the telomere will then switch to an 'active' state in which the telomerase can add TG_{1-3} repeats. This would then lead to increased Rap1p binding and a switch back to the repressed state, explaining the ability of the cell to regulate telomere length within a relatively narrow window centered about a fixed average value. This model might help to explain the explosive telomere elongation seen when mutant telomere repeats are incorporated at telomere ends in *K. lactis* (Krauskopf and Blackburn, 1996; McEachern and Blackburn, 1995). When a sufficient number of mutant repeats are added such that Rap1p binding cannot reach the proposed threshold level for telomerase repression, the resulting absence of negative feedback would be predicted to lead to uncontrolled telomere elongation.

Presumably the Rif proteins play some role in the 'counting' process described above. One possibility, which has yet to be tested, is that the Rif proteins directly regulate telomerase, either through protein-protein interactions with the catalytic subunit [Est2p; (Lingner *et al.*, 1997)] or associated factor(s). Several telomere end-binding proteins have recently been identified that might serve as regulators of telomerase: Est1p (Virta-Pearlman *et al.*, 1996), Est4p/Cdc13p (Lin and Zakian, 1996; Nugent *et al.*, 1996), and the Est4p/Cdc13p-interacting protein Stn1 (Grandin *et al.*, 1997). Although both *EST4/CDC13* and *STN1* are essential for viability and appear to be required to prevent exonucleolytic degradation of telomeres, certain alleles of these genes lead to a telomere elongation phenotype reminiscent of that seen in *rif* mutants (Grandin *et al.*, 1997). Both proteins are therefore excellent candidates for factors that might negatively regulate telomerase, either by controlling its access to the telomere or its activity when bound there. It will be important, then, to determine first whether the Rif proteins are in fact localized at telomeres, and if so whether they interact directly with any of the known telomere end-binding proteins. Although the negative feedback model described above provides a satisfying explanation of the basic features of telomere length regulation in yeast, it should be clear that the molecular details remain obscure. It may prove difficult to test this model further until the various components of the system (particularly the Rif and telomerase complexes) are better characterized and can be studied in *in vitro* systems.

Finally, it should also be pointed out that telomere length may be regulated at many different levels, and alternative mechanisms of regulation to those outlined above are not excluded. In fact, recent studies have shown that extremely elongated telomeres (inherited from a *rap1^t* mutant strain in a genetic cross) can be restored to a normal length through a (*RAD52*-dependent) recombinational mechanism (Li and Lustig, 1996). In addition, the *PIF1* gene, whose product is a DNA helicase that inhibits telomere elongation, appears to regulate telomere length by a mechanism that does not involve the Rap1p carboxyl terminus (Schulz and Zakian, 1994).

A Conserved Role for Double-Strand Telomere-Repeat Binding Proteins in Length Regulation

Quite strikingly, recent studies with the fission yeast *S. pombe* and with human cells suggest that negative regulation of telomere elongation by a double-strand telomere-repeat binding protein may be a highly conserved feature of telomere biology. Several years ago de Lange and colleagues reported the cloning and characterization of a human telomere repeat [$(T_2AG_3)_n$] binding protein which they called TRF1 (Chong *et al.*, 1995). This group has recently reported two lines of evidence which suggest that the amount of TRF1 protein bound at a telomere determines the repeat length (van Steensel and de Lange, 1997). Expression of a dominant-negative mutant of TRF1 that reduces telomere binding *in vivo* results in telomere elongation. Conversely, overexpression of the wild-type protein leads to telomere shortening. These observations are consistent with the idea that TRF1 is limiting in the cell and that the amount of protein bound to a telomere can serve as a length-dependent regulator of telomerase. Therefore, although the details may differ considerably (to begin with there is no evidence that Rap1p is limiting for binding to telomeres in yeast) both systems may employ a negative feedback mechanism that can sense the amount of a particular protein bound to the telomeric repeats. It should be noted that a cDNA encoding a possible second TRF1-like human protein has been identified recently (Bilaud *et al.*, 1996). An interesting possibility, then, is that this second protein competes with TRF1 for binding to telomeres and thus contributes to a length regulation mechanism acting at the level of DNA binding. Characterization of this putative telomere-binding protein may thus help to clarify some of the details of telomere length regulation in human cells.

In *Schizosaccharomyces pombe*, the *Taz1$^+$* gene encodes a double-stranded telomere-repeat protein, which oddly enough is more similar in primary amino acid sequence to the human TRF1 than it is to its analog (Rap1p) in budding yeast (Cooper *et al.*, 1997). Nonetheless, deletion of the *Taz1$^+$* gene results in both a loss of TPE and massive telomere repeat elongation, two phenotypes strikingly similar to *rap1^t* mutants in *S. cerevisiae*. It would appear, then, that both the silencing and telomere length regulatory functions of telomere repeat binding proteins are conserved between fission and budding yeasts, despite the absence of any strong similarity between these proteins at the primary sequence level. It remains to be seen whether the detailed mechanisms are conserved, since to date there are no known homologues of either Sir or Rif proteins in *S. pombe* [with the possible exception of Sir2p, which appears to be conserved from bacteria to humans (Brachmann *et al.*, 1995)].

Perspectives

This brief review has focused on recent results, primarily from the budding yeast *S. cerevisiae*, that provide some new insights into the questions of how cells sense telomere length, and how they might use this information to control telomere replication. Although important principles and several new proteins implicated in these processes have been identified in the past few years, our picture of the molecular mechanisms that control telomere replication is still very fragmentary. This situation is almost certain to change dramatically in the coming years, since it is now clear that genetic studies (primarily, but not solely in budding yeast) and biochemical advances in several different systems (Harrington *et al.*, 1997; Lingner *et al.*, 1997; Nakayama *et al.*, 1997) are beginning to converge towards a picture of the telomerase enzyme and a probable set of auxiliary or regulatory factors. It seems likely that these proteins will constitute a target for other regulatory molecules, including those (such as Rap1p and the Rif proteins) that are involved in telomere length control. The identification of protein-protein interactions and a clear demonstration of their *in vivo* relevance may provide important insights into the possible connection between Rap1p (and related telomere-repeat binding proteins in more complex eukaryotes) and telomerase. The development of novel genetic screens to identify genes specifically involved in telomere length sensing and regulation may also be critical. We may still be unaware of some important players in this process for the simple reason that much of the work to date has been driven by studies of either Rap1p or the telomerase enzyme. As the picture of telomere-binding proteins and the telomerase enzyme complex comes into focus in the next few years, it should become possible to explore the mechanisms of telomere replication and telomere length regulation biochemically.

Acknowledgements

I would like to thank several colleagues, particularly past and present co-workers Stéphane Marcand, Steve Buck, David Wotton, Krishnaveni Mishra, and Irina Serdobova for discussions. Work in my laboratory has been supported by the National Institutes of Health (GM40094) and the American Cancer Society (VM-62), and more recently by the Swiss National Fund and the State of Geneva.

References

Aparicio, O.M., Billington, B.L., and Gottschling, D.E. (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. Cell *66*, 1279 – 1287.

Berman, J., Tachibana, C.Y., and Tye, B.K. (1986). Identification of a telomere-binding activity from yeast. Proc. Natl. Acad. Sci. USA *83*, 3713 – 3717.

Bilaud, T., Koering, C.E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S.M., and Gilson, E. (1996). The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. Nucleic Acids Res. *24*, 1294 – 1303.

Blasco, M.A., Funk, W., Villeponteau, B., and Greider, C.W. (1995). Functional characterization and developmental regulation of mouse telomerase RNA. Science. *269*, 1267 – 1270.

Brachmann, C.B., Sherman, J.M., Devine, S.E., Cameron, E.E., Pillus, L., and Boeke, J.D. (1995). The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. Genes Dev. *9*, 2888 – 2902.

Buchman, A.R., Kimmerly, W.J., Rine, J., and Kornberg, R.D. (1988a). Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in *Saccharomyces cerevisiae*. Mol. Cell Biol. *8*, 210 – 225.

Buchman, A.R., Lue, N.F., and Kornberg, R.D. (1988b). Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. Mol. Cell. Biol. *8*, 5086 – 5099.

Buck, S.W., and Shore, D. (1995). Action of a RAP1 C-terminal silencing domain reveals an underlying competition between *HMR* and telomeres in yeast. Genes and Development. *9*, 370 – 384.

Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P., and de Lange, T. (1995). A human telomeric protein. Science. *270*, 1663 – 1667.

Cockell, M., Palladino, F., Laroche, T., Kyrion, G., Liu, C., Lustig, A.J., and Gasser, S.M. (1995). The carboxy termini of Sir4 and Rap1 affect Sir3 localization: evidence for a multicomponent complex required for yeast telomeric silencing. J. Cell Biol. *129*, 909 – 924.

Conrad, M.N., Wright, J.H., Wolf, A.J., and Zakian, V.A. (1990). RAP1 protein interacts with yeast telomeres *in vivo*: overproduction alters telomere structure and decreases chromosome stability. Cell. *63*, 739 – 750.

Cooper, J.P., Nimmo, E.R., Allshire, R.C., and Cech, T.R. (1997). Regulation of telomere length and function by a Myb-domain protein in fission yeast. Nature. *385*, 744 – 747.

De Lange, T. (1995). Telomere dynamics and genome instability in human cancer. Telomeres. E.H. Blackburn and C.W. Greider. (Plainview, N.Y.: Cold Spring Harbor Press), pp. 265 – 293.

Gilson, E., Roberge, M., Giraldo, R., Rhodes, D., and Gasser, S.M. (1993). Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. J. Mol. Biol. *231*, 293 – 310.

Gottschling, D.E., Aparicio, O.M., Billington, B.L., and Zakian, V.A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of pol II transcription. Cell. *63*, 751 – 762.

Grandin, N., Reed, S.I., and Charbonneau, M. (1997). Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13. Genes Dev. *11*, 512 – 527.

Greider, C.W. (1996). Telomere length regulation. Annu. Rev. Biochem. *65*, 337 – 365.

Hardy, C.F.J. (1991). Studies on RAP1 and the RAP1 interacting factor, RIF1. Ph.D. thesis, Columbia University.

Hardy, C.F.J., Balderes, D., and Shore, D. (1992a). Dissection of a carboxy-terminal region of the yeast regulatory protein RAP1 with effects on both transcriptional activation and silencing. Mol. Cell. Biol. *12*, 1209 – 1217.

Hardy, C.F.J., Sussel, L., and Shore, D. (1992b). A RAP1-interacting protein involved in silencing and telomere length regulation. Genes Dev. 6, 801 – 814.

Harley, C.G., and Villeponteau, B. (1995). Telomeres and telomerase in aging and cancer. Current Opin. Genet. and Dev. *5*, 249 – 255.

Harrington, L., McPhail, T., Mar, V., Zhou, W., Oulton, R., Bass, M.B., Arruda, I., and Robinson, M.O. (1997). A mammalian telomerase-associated protein. Science. *275*, 973 – 977.

Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S.M., and Grunstein, M. (1995). Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell. *80*, 583 – 592.

Hecht, A., Strahl-Bolsinger, S., and Grunstein, M. (1996). Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. Nature *383*, 92 – 96.

Kennedy, B.K., Austriaco, N.R., Jr., Zhang, J., and Guarente, L. (1995). Mutation in the silencing gene SIR4 can delay aging in *S. cerevisiae*. Cell. *80*, 485 – 496.

Kirk, K.E., Harmon, B.P., Reichardt, I.K., Sedat, J.W., and Blackburn, E.H. (1997). Block in anaphase chromosome separation caused by a telomerase template mutation. Science *275*, 1478 – 1481.

Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996). The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. Cell *85*, 125 – 136.

Krauskopf, A., and Blackburn, E.H. (1996). Control of telomere growth by interactions of RAP1 with the most distal telomeric repeats. Nature *383*, 354 – 357.

Kurtz, S., and Shore, D. (1991). The RAP1 protein activates and silences transcription of mating-type genes in yeast. Genes Dev. *5*, 616 – 628.

Kyrion, G., Boakye, K.A., and Lustig, A.J. (1992). C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *12*, 5159 – 5173.

Kyrion, G., Liu, K., Liu, C., and Lustig, A.J. (1993). RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. Genes and Development *7*, 1146 – 1159.

Lendvay, T.S., Morris, D.K., Sah, J., Balasubramanian, B., and Lundblad, V. (1996). Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional *EST* genes. Genetics *144*, 1399 – 1412.

Li, B., and Lustig, A.J. (1996). A novel mechanism for telomere size control in *Saccharomyces cerevisiae*. Genes Dev. *10*, 1310 – 1326.

Lin, J.J., and Zakian, V.A. (1996). The Saccharomyces CDC13 protein is a single-strand TG_{1-3} telomeric DNA-binding protein *in vitro* that affects telomere behavior *in vivo*. Proc. Natl. Acad. Sci. USA *93*, 13760 – 13765.

Lingner, J., Cooper, J.P., and Cech, T.R. (1995). Telomerase and DNA: no longer a lagging strand problem? Science *269*, 1533 – 1534.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. Science *276*, 561 – 567.

Lustig, A.J., Kurtz, S., and Shore, D. (1990). Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. Science. *250*, 549 – 553.

Lustig, A.J., Liu, C., Zhang, C., and Hanish, J.P. (1996). Tethered Sir3p nucleates silencing at telomeres and internal loci in *Saccharomyces cerevisiae*. Mol. Cell. Biol. *16*, 2483 – 2495.

Lustig, A.J., and Petes, T.D. (1986). Identification of yeast mutants with altered telomere structure. Proc. Natl. Acad. Sci. *83*, 1398 – 1402.

Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E., and Gasser, S.M. (1996). Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. Genes Dev. *10*, 1796 – 1811.

Marcand, S., Buck, S.W., Moretti, P., Gilson, E., and Shore, D. (1996). Silencing of genes at nontelomeric sites in yeast is controlled by sequestration of silencing factors at telomeres by Rap1 protein. Genes Dev. *10*, 1297 – 1309.

Marcand, S., Gilson, E., and Shore, D. (1997). A protein-counting mechanism for telomere length regulation in yeast. Science *275*, 986 – 990.

McEachern, M.J., and Blackburn, E.H. (1995). Runaway telomere elongation caused by telomerase RNA gene mutations. Nature *376*, 403 – 409.

Moazed, D., and Johnson, D. (1996). A deubiquitinating enzyme interacts with SIR4 and regulates silencing in *S. cerevisiae*. Cell *86*, 667 – 677.

Moazed, D., Kistler, A., Axelrod, A., Rine, J., and Johnson, A.D. (1997). Silent information regulator protein complexes in *Saccharomyces cerevisiae*: a SIR2/SIR4 complex and evidence for a regulatory domain in SIR4 that inhibits its interaction with SIR3. Proc. Natl. Acad. Sci. USA *94*, 2186 – 2191.

Moretti, P., Freeman, K., Coodley, L., and Shore, D. (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. Genes and Development *8*, 2257 – 2269.

Nakayama, J., Saito, M., Nakamura, H., Matsuura, A., and Ishikawa, F. (1997). TLP1: a gene encoding a protein component of mammalian telomerase is a novel member of WD repeats family. Cell *88*, 875 – 884.

Nugent, C.I., Hughes, T.R., Lue, N.F., and Lundblad, V. (1996). Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. Science *274*, 249 – 252.

Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., and Gasser, S.M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. Cell *75*, 543 – 555.

Runge, K.W., and Zakian, V.A. (1989). Introduction of extra telomeric DNA sequences into *Saccharomyces cerevisiae* results in telomere elongation. Mol. Cell Biol. *9*, 1488 – 1497.

Sandell, L.L., and Zakian, V.A. (1993). Loss of a yeast telomere: arrest, recovery, and chromosome loss. Cell *75*, 729 – 739.

Schulz, V.P., and Zakian, V.A. (1994). The Saccharomyces PIF1 DNA helicase inhibits telomere elongation and de novo telomere f ormation. Cell *76*, 145 – 155.

Shore, D. (1994). RAP1: a protean regulator in yeast. Trends in Genetics *10*, 408 – 412.

Shore, D., and Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell *51*, 721 – 732.

Singer, M.S., and Gottschling, D.E. (1994). *TLC1*: template RNA component of *Saccharomyces cerevisiae* telomerase. Science *266*, 404 – 409.

Strahl-Bolsinger, S., Hecht, A., Luo, K., and Grunstein, M. (1997). SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. Genes Dev. *11*, 83 – 93.

Sussel, L., and Shore, D. (1991). Separation of transcriptional activation and silencing functions of the *RAP1*-encoded repressor/activator protein 1: Isolation of viable mutants affecting both silencing and telomere length. Proc. Natl. Acad. Sci. USA *88*, 7749 – 7753.

van Steensel, B., and de Lange, T. (1997). Control of telomere length by the human telomeric protein TRF1. Nature *385*, 740 – 743.

Virta-Pearlman, V., Morris, D.K., and Lundblad, V. (1996). Est1 has the properties of a single-stranded telomere end-binding protein. Genes Dev. *10*, 3094 – 3104.

Watson, J.D. (1972). Origin of concatemeric T7 DNA. Nature New Biology *239*, 197 – 201.

Wotton, D., and Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *S. cerevisiae*. Genes Dev. *11*, 748 – 760.

Zakian, V.A. (1996). Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. Annu. Rev. Genet. *30*, 141 – 172.

Biol. Chem., Vol. 378, pp. 599 – 607, July 1997

Review

The Role of DNA Conformation in Transcriptional Initiation and Activation in *Escherichia coli*

Johannes Geiselmann

Department of Molecular Biology, University of Geneva, 30, Quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland

Transcription activation in prokaryotes relies on a multitude of molecular mechanisms. Many of them use DNA bending or other deformations of the DNA to control the rate of transcription initiation. All steps of initiation involve a particular transconformation reaction of the DNA and can be controlled individually by activators. This review discusses the thermodynamic and kinetic basis of transcription initiation and illustrates the strategies used by activators. Particular emphasis is given to mechanisms that necessitate DNA-bending by the activator.

Introduction

An increasing number of transcriptional activators have been studied in detail and they reveal a large spectrum of different mechanisms used to control transcription. While several recent reviews treat detailed structural and mechanistic aspects of transcriptional initiation and activation in prokaryotes, I will focus here on the role of the conformation of DNA in these processes. DNA is dynamically manipulated during transcription and changes its conformation at every step of the process. It is therefore not surprising that proteins that affect DNA conformation can have a profound influence on transcription. A survey of prokaryotic transcriptional activators demonstrates that most of them bend DNA (Perez Martin *et al.*, 1994). This review provides a discussion of some selected examples of the influence of DNA conformation on transcription but it is not intended to be an exhaustive summary of occasions where DNA bending plays an important role in transcription initiation. Because a great amount of work by many groups has been devoted to the issues discussed in this review, the references cited can only be illustrative examples among many. I beg indulgence from researchers who feel that important contributions have been neglected.

The role of imposing a precise conformation on the DNA, i.e. constraining its three-dimensional path, can be explained in a straightforward manner for promoters recognized by the σ^{54}-containing holoenzyme $E\sigma^{54}$ but is much less evident for promoters specific for the more abundant forms of RNA polymerase, such as the housekeeping holoenzyme $E\sigma^{70}$. This difference arises from the distinct mechanisms of transcription initiation at the two types of promoters. $E\sigma^{54}$ is blocked at one particular step of transcription initiation and needs an enhancer-like activator protein to get past this 'bottleneck'. The conformation of the DNA determines the local concentration of the DNA-bound enhancer in the vicinity of $E\sigma^{54}$ and thereby profoundly affects the rate of transcription initiation (see below). The molecular mechanisms of activators of $E\sigma^{70}$ promoters are more diverse, and any one of the steps of initiation can be the target by the activator. In addition, these activators very often bind immediately adjacent to the promoter and DNA bending apparently does not function to approach a far-away site on the DNA. Most of this review is devoted to discussing possible functions of the geometry of DNA at such promoters.

Transcription Initiation

The isolated RNA polymerase holoenzyme is capable of carrying out the entire transcription reaction, from initiation to termination. No additional factors are required. Even though the activity of RNA polymerase is modulated by repressors or activators, there is no evidence to suggest that the reaction pathway of transcription initiation is altered by these regulators. Activators only accelerate, and repressors slow down, particular steps of the reaction of transcription initiation. In order to elucidate the molecular mechanism of transcription regulation it is therefore necessary to describe the structural and kinetic foundation of transcription initiation. A recent review summarizes the large number of experiments that have revealed molecular details of this process for the *Escherichia coli* RNA polymerase (Record *et al.*, 1996). For the sake of clarity unnecessary details are omitted in the description that follows.

The Mechanism of Transcription Initiation

Transcription initiation can be divided into three distinct phases based on structural as well as functional features:

(i) A bimolecular reaction where RNA polymerase binds to the double-stranded promoter,

(ii) isomerization of this 'closed complex' to the 'open complex' where about 10 bases of single-stranded DNA form the so-called transcription bubble, and

(iii) formation of the first phosphodiester bonds and escape of RNA polymerase from the promoter, freeing the promoter for the binding of another holoenzyme (Figure 1).

Additional reaction intermediates have been observed experimentally by adjusting solution conditions (salt concentrations, temperature, etc.) (Suh *et al.*, 1993; Record *et al.*, 1996), but for simplicity we consider only one further such subdivision (see below). The three-dimensional path of the DNA plays a role at each step and I will discuss physical models that could explain the experimental data. These structural and functional underpinnings are presented in the following paragraphs and concrete examples are cited in the second part of this review.

Binding The only step of transcription initiation that depends on the concentration of RNA polymerase is the initial binding to the promoter to form the closed complex (RP_C). Binding is fast, facilitated by sliding of RNA polymerase along the DNA (von Hippel and Berg, 1989), and RP_C is therefore usually in rapid equilibrium with free RNA polymerase. The binding constants measured *in vitro* are often comprised between $10^6\,M^{-1}$ and $10^8\,M^{-1}$ but 'exceptional promoters' extend this range by one order of magnitude in either direction. An *Escherichia coli* cell contains around 2000 molecules of RNA polymerase (Ishihama, 1993) per genome equivalent. Most of these molecules are engaged in transcription or bound non-specifically to DNA and less than 1% is thought to be free in solution (McClure, 1985). The concentration of free holoenzyme within the cell is therefore likely to be on the order of 10 nM to 50 nM. This concentration has never been measured directly, but indirect data are consistent with this estimation. In particular, recent experimental results (Dove *et al.*, 1997) prove that transcription can be activated by recruitment of RNA polymerase to the promoter, implying a 'low' intracellular concentration of free RNA polymerase. Even though the precise concentration of RNA polymerase may depend on the physiological state of the cell, variations are probably small because RNA polymerase concentration is buffered by autoregulation (Dykxhoorn *et al.*, 1996) and non-specific binding to DNA.

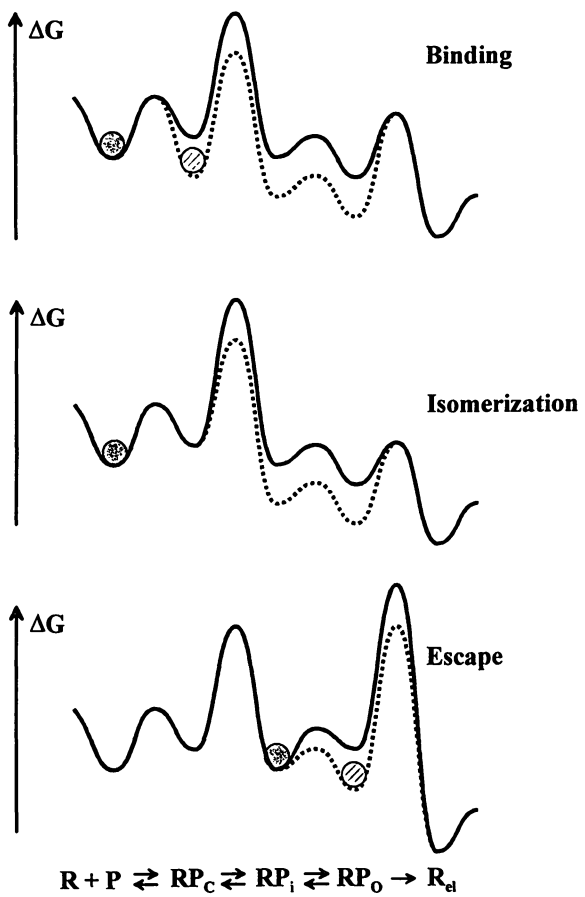


Fig. 1 Free Energy Diagrams of Transcription Initiation.
The free energy of reaction intermediates of transcription initiation (shown at the bottom) are drawn for promoters limited at different steps of initiation. The rate-limiting step is indicated to the right of each diagram. The solid line represents the non-activated promoter, the stippled line the situation in the presence of a transcriptional activator. The most populated complex during steady state transcription is indicated by the circle. The top diagram corresponds to the situation at the *lac* promoter, where CRP favors initial binding; the middle diagram is typical for a *gal*-type promoter, where CRP accelerates isomerization; and the bottom diagram represents the situation at the *malT* promoter, where CRP stabilizes open forms of the promoter and thus accelerates promoter escape.

Most repressors impede the first step of transcription initiation by competing with RNA polymerase for overlapping binding sites on the DNA (Schlax *et al.*, 1995). This type of control is purely thermodynamic. On the other hand, any interactions that increase the affinity of RNA polymerase for a weak promoter will lead to an activation of transcription by increasing the saturation of the promoter with RNA polymerase. For example, direct protein-protein contacts with an activator or binding within a DNA-loop stabilize RNA polymerase on the promoter (Figure 2).

The closed complex is characterized structurally by a DNAseI footprint covering about 50 bp between positions −5 and −55 (Schickor *et al.*, 1990; Kovacic, 1987; Cowing *et al.*, 1989). The DNA remains double-stranded and RNA polymerase covers one face of the helix. The extent of the protected region exceeds the length of the putative DNA-binding channel on RNA polymerase (Darst *et al.*, 1989). The DNA may therefore already be partially wrapped around RNA polymerase (Zinkel and Crothers, 1991; Perez Martin *et al.*, 1994). Indeed, curved DNA upstream of the promoter favors RNA polymerase binding (Lavigne *et al.*, 1992a; Perez Martin and Espinosa, 1994a). The additional upstream contacts are formed by the α-subunits of RNA polymerase (Ebright and Busby, 1995). Since the α-subunits preferentially bind to AT-rich sequences, the so-called UP-element (Rao *et al.*, 1994) and since bent DNA is formed by tracts of As, it is not clear whether improved binding to bent sequences is due to stronger interactions with the α-subunits or to a more favorable DNA conformation or to both.

Isomerization After the initial binding to the promoter RNA polymerase isomerizes to form an open complex

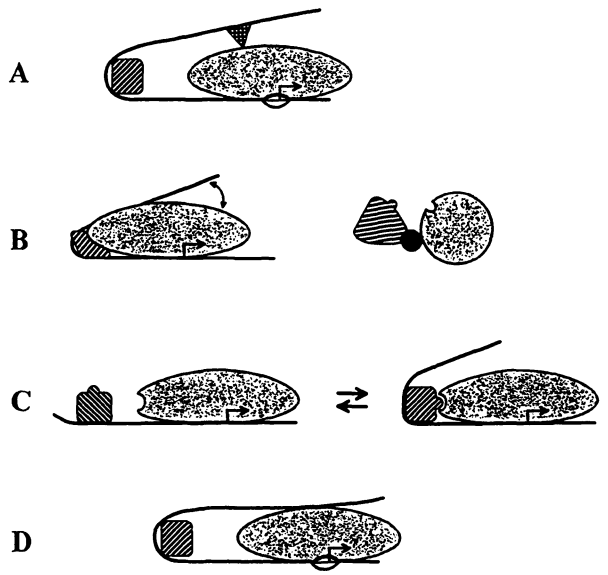


Fig. 2 Structural Models for Different Roles of DNA-Bending in Transcription Activation.
RNA polymerase is symbolized by the gray oval, the DNA is represented by the black line, the transcription start site is indicated by a broken arrow, and the activator is shown as a striped rectangle. (A) The activator bends DNA and increases the local concentration of an enhancer binding protein (triangle) in the vicinity of RNA polymerase. (B) The activator binds close to the promoter but is out of phase with respect to RNA polymerase. The right part of the diagram is rotated by 90° and shows the view down the DNA axis (black circle). In order to make a favorable contact with RNA polymerase the DNA has to be twisted. (C) DNA-bending serves to juxtapose interacting regions on the activator and RNA polymerase. (D). Transcriptional activation is due to contacts between the far-upstream DNA and the back of RNA polymerase, leading to opening of the transcription bubble.

(RP_O) passing through at least two additional reaction intermediates (Craig *et al.*, 1995). We only discuss the first intermediate and refer to it as RP_i (Buc and McClure, 1985). An alternative nomenclature calls it RP_{C2} (Record *et al.*, 1996) to emphasize the fact that the DNA remains closed, as judged by cytosine reactivity with DMS (Duval-Valentin and Ehrlich, 1987), but that the conformation of the complex differs markedly from the initial closed complex. In relaxed DNA the two recognition hexamers are 'over-rotated' by 68° (deHaseth and Helmann, 1995) and simultaneous binding of RNA polymerase to both regions may imply the generation of torsional strain. One attractive model therefore proposes that the torsional strain within the linker between the −10 and −35 hexamers provides part of the energy needed for subsequent DNA-opening. Even though the model is very reasonable and intellectually appealing, experimental data that define the structure of the postulated kinetic intermediate are scarce and circumstantial. The intermediate complex is almost exclusively defined by two properties: the absence of an open region and its rapid equilibration with RP_O.

Irrespective of its precise structure, the formation of this stressed intermediate, RP_i, is the rate-limiting step at many promoters (Record *et al.*, 1996) and thus constitutes a preferred target for activators. Binding measurements on topoisomers of minicircles are consistent with the untwisting model (see above); the data show that the rate-limiting step in the formation of RNA polymerase-promoter complexes has topologically unwound the DNA by about one third of a helical turn (Su and McClure, 1994). Transcriptional activators that accelerate the formation of RP_i may thus do so by assisting the untwisting of DNA. The transition state could be stabilized by direct protein-protein contacts or by the formation of a DNA-loop, provided that these interactions support the untwisting of the spacer DNA (Figure 2). The DNAseI footprint of RP_i covers bases between +20 and −55 (Schickor *et al.*, 1990), a region longer than the largest linear dimension of RNA polymerase (Darst *et al.*, 1989). The DNA therefore has to be wrapped around RNA polymerase as had been proposed earlier (Buc *et al.*, 1987). Atomic force microscopy (Rees *et al.*, 1993) confirms this assertion directly. The intermediate complex isomerizes to the open complex where the DNA strands are separated between bases −9 and +2 (deHaseth and Helmann, 1995). DNA-melting nucleates within the −10 hexamer and proceeds towards the start site of transcription (Chen and Helmann, 1997). The transition from RP_i to RP_O is usually rapid (Buc and McClure, 1985) but the equilibrium may be shifted by transcriptional activators (Figure 1). Stabilization of RP_O could increase transcriptional initiation in cases where the subsequent step, promoter escape, is rate-limiting (see below). The structural characteristics that distinguish RP_O from the preceding intermediates are: the existence of a melted region of DNA and more extensive wrapping of DNA around RNA polymerase (Record *et al.*, 1996). Any influences, such as activator proteins, that favor DNA-bending or DNA melting could thus potentially activate transcription at this step.

Escape The ultimate 'goal' of RNA polymerase is not to tightly and efficiently bind the promoter, but, on the contrary, to leave the promoter as quickly as possible. This dichotomy between strong binding and facile release imposes important constraints on the design of a promoter and the mechanism of activators (see below). Eukaryotic RNA polymerases have circumvented this problem by incorporating irreversible steps of energy expenditure into the pathway of transcription activation. For example, the C-terminal domain of RNA polymerase is phosphorylated before or during promoter escape.

Promoter escape is a complex process. Starting with RP_O, short transcripts (< 12 nt) are produced in the presence of NTPs. Most of these transcripts are released in a reaction termed abortive initiation. After transcript release, RNA polymerase reverts to the open complex conformation. During abortive initiation the RNA polymerase remains bound to the promoter (Straney and Crothers, 1987; Krummel and Chamberlin, 1989; Carpousis and Gralla, 1980). However, it can not be ruled out that a possible translocation of RNA polymerase during abortive cycling has passed undetected by currently used methods,

as was the case for analogous studies of transcription elongation (Komissarova and Kashlev, 1997). Once RNA polymerase has managed to synthesize a sufficiently long transcript (8 to 12 nts) it escapes from the promoter, releasing the sigma subunit, and forms the very stable elongation complex. Elongation is very processive and ceases only when RNA polymerase encounters a transcriptional terminator (von Hippel and Yager, 1992; von Hippel *et al.*, 1996).

The major structural change of the complex during promoter escape is an enlargement of the transcription bubble from ca. 10 nt in RP_O to the full 18 nt of an elongation complex. (Yager and von Hippel, 1991). During promoter escape, the early transcribed region has to be melted, and contacts with the promoter or activator proteins have to be relinquished. It is therefore not surprising that the sequence of the early transcribed region modulates the rate of promoter escape (Kammerer *et al.*, 1986), and that a polymerase that is very tightly anchored to the promoter traverses this last step of initiation only very slowly (Ellinger *et al.*, 1994). Since additional base pairs have to be melted it is expected that activators that assist DNA melting will accelerate promoter escape.

Bending of DNA

DNA bending participates in different ways at all steps of transcription initiation. This feature can be exploited by activators that influence these different steps. Indeed, most activators distort DNA upon binding (Perez Martin *et al.*, 1994). In general, DNA-bending serves to place a particular segment of DNA, and/or proteins bound to this DNA, into a precise spatial position with respect to RNA polymerase. The resulting optimal juxtaposition of an activating region with its target on RNA polymerase leads to a stronger interaction because of the increased local concentration. Such concentration effects can attain considerable magnitudes: the introduction of a 120° bend in the center of a 150 bp fragment of DNA increases 500-fold the local concentration of one end in the vicinity of the other (Rippe *et al.*, 1995). This example demonstrates the importance of DNA conformation in the assembly of a nucleoprotein complex, which will or will not form depending on the path of the DNA-scaffold. DNA-bending not only increases the local concentration of, e.g. an upstream-bound activator, it also constrains the rotation of the activator. The reduction of rotational entropy can further increase the strength as well as the specificity of the interaction with RNA polymerase. Modulating the conformation of the DNA thus permits to dictate which interactions will occur, and, as a consequence, whether transcription will be activated or not.

Energetics of Transcription Initiation

Free energy diagrams (as the ones shown in Figure 1) are a convenient way to visualize the kinetics of transcription initiation. The stability (in terms of free energy) of a state is proportional to the depth of the valleys. The rate of transition (reaction rate) from one state to an adjacent state decreases exponentially with the height of the activation barrier separating them (Fersht, 1977). The cyclic AMP receptor protein, CRP, is a useful example of a transcriptional activator because it bends DNA and acts on all steps of transcription initiation depending on the promoter context. The diagram of figure 1, top represents approximately the situation at the well-characterized *lac* promoter (Busby and Buc, 1987). The largest free energy barrier, corresponding to the slowest step of the reaction is the isomerization to RP_i. The activator, CRP, improves binding of RNA polymerase. The initial situation is almost identical at the *galP1* (Figure 1 middle) or *malT* (Figure 1 bottom) promoters, but now CRP accelerates isomerization or promoter escape respectively (Lavigne *et al.*, 1992; Menendez *et al.*, 1987).

Steady State Rate of Transcription Several barriers have to be surmounted during initiation. The flow through such a linear reaction scheme is essentially determined by the highest activation barrier (the rate-limiting step) and the concentration of substrates before that barrier. At both the *lac* and *galP1* promoters isomerization towards the open promoter complex is rate-limiting (Figure 1). In one case the promoter is activated by stabilizing the substrate of isomerization (RP_C) in the other case the rate-limiting step itself is accelerated. Even when a particular step is accelerated, the overall reaction rate may remain unchanged. Imagine that an activator accelerates promoter escape at the *lac* or *galP1* promoters. Because this step is not rate-limiting the activator apparently has no effect. If the same activator acted at the *malT* promoter (limited at escape, Figure 1 bottom) transcription activation would readily be observed. Conversely, adverse effects of an activator may be hidden if they slow down a previously fast reaction step. Promoter escape is fast at the *galP1* and *lac* promoters. The activator makes direct contacts with RNA polymerase (Niu *et al.*, 1996; Zhou *et al.*, 1994; Busby and Ebright, 1994), contacts that have to be broken during promoter escape. It appears thus plausible that the corresponding energy barrier becomes higher (as indicated in Figure 1). Slowing escape from these promoters does not, however, modify the frequency of initiation because the rate-limiting step remains isomerization. In summary, the combination of all steps of the initiation reaction determines the frequency of transcription initiation. One step is often rate-limiting, but the rate-determining step may change in the presence of an activator.

Representative Experimental Systems

$E\sigma^{54}$-Activated Promoters

A special class of bacterial promoters are activated by enhancer-like proteins bound at distant locations on the DNA (Gralla, 1996; Gralla and Collado-Vides, 1996). All of these promoters depend on alternative forms of RNA polymerase, containing either the σ^{54} subunit (Gralla and Collado-

Vides, 1996) or proteins from the bacteriophage T4 (Brody *et al.*, 1995). $E\sigma^{54}$ forms a stable closed complex but is incapable of isomerizing to the open complex. In terms of the free energy diagrams this corresponds to one large energy barrier at isomerization, and negligible barriers at all other steps of initiation. The activator protein dramatically accelerates isomerization, much like CRP at the *galP1* promoter (Figure 1 middle). The similarity with the *galP1* promoter is only a formal one: activators of σ^{54}-dependent promoters utilize ATP hydrolysis to open the DNA, and the factor of acceleration of the isomerization step is much greater than for activators of $E\sigma^{70}$ promoters. Because there is only one rate-determining step the formal description of such systems is rather simple: no transcription occurs in the absence of activator and transcription proceeds at a maximal rate, intrinsic to the promoter, when the activator binds its target on RNA polymerase, the σ^{54} subunit. The activity of the promoter is thus directly proportional to the binding saturation of $E\sigma^{54}$ with activator. The promoter is therefore regulated by adjusting the local concentration of enhancer near the promoter. The enhancer, bound upstream, approaches RNA polymerase by looping of DNA (Figure 2A), and an activator can favor such loop formation.

The *glnHp2* and *glnAp2* promoters require intrinsic or induced bending of the DNA for activity (Brahms *et al.*, 1995; Carmona and Magasanik, 1996). In the natural situation a binding site for the integration host factor (IHF), a histone-like DNA-bending protein (Rice *et al.*, 1996), is located between the promoter and the enhancer binding sites (Claverie-Martin and Magasanik, 1991). The induced DNA-bend brings the enhancer into the vicinity of RNA polymerase. Activation persists when the enhancer binding sites are moved further upstream as predicted by the looping model. In this case, however, IHF inhibits transcription (Claverie-Martin and Magasanik, 1991) because the DNA bend now presents the wrong region of DNA, devoid of enhancer binding sites, to RNA polymerase. This example illustrates another important feature of transcriptional regulation by rigidifying the DNA scaffold: this mode of regulation provides a mechanism for preventing unwanted 'cross-regulation' by heterologous activators. For example, IHF prevents promiscuous activation of the σ^{54}-dependent Pu promoter by restricting access to the natural activator, XylR (Perez Martin and De Lorenzo, 1995). In summary, at $E\sigma^{54}$ promoters proteins that alter the conformation of DNA act mainly by modifying the local concentration of the enhancer binding protein in the vicinity of RNA polymerase. These proteins take advantage of the fact that DNA conformation is a very sensitive regulator of this local concentration.

Activation by Direct Protein-Protein Interactions

Contrary to the $E\sigma^{54}$ promoters any of the steps of transcriptional initiation can be the target of activation at $E\sigma^{70}$ promoters: binding is often not saturating given the intracellular concentration of RNA polymerase, isomerization is rather slow, and in certain cases promoter escape can become rate-limiting.

Activators that Overlap the Promoter The binding sites of activators of $E\sigma^{70}$ promoters are found immediately upstream of, or overlapping with, the core promoter (Gralla and Collado-Vides, 1996). Even though several systems have been well characterized I will use CRP to illustrate the principle. At so called class II promoters the CRP binding site is centered around −40. The −35 hexamer is absent and it is thought that contacts between CRP and RNA polymerase compensate for the missing contacts between RNA polymerase and the promoter. The mechanism of activation is known in great detail (for a recent review see Busby and Ebright, 1997). CRP directly interacts via two different regions with at least two subunits of RNA polymerase. The contact between activating region 2 (AR2) of CRP and the amino-terminal domain of the α-subunit of RNA polymerase specifically increases the rate of isomerization to the open complex (Niu *et al.*, 1996). As discussed above, the activation energy for isomerization may in part be provided by the distortion of the spacer DNA. CRP could stabilize this torsion. A possible molecular mechanism is sketched in Figure 2B. RNA polymerase, bound to the −10 hexamer, and CRP bound to its site centered at −41.5 are slightly out of phase. In order for AR2 to reach its target on the α-subunit, CRP has to rotate clockwise (looking in the direction of transcription). This motion leads to an untwisting of the spacer DNA, thus stabilizing the transition state towards the open complex. This model would naturally explain how CRP can exert a 'long range effect' in the sense that the AR2-α interaction takes place at a long distance from DNA or the active site of RNA polymerase (Busby and Ebright, 1997). This model predicts that CRP should no longer activate transcription when its binding site is rotated towards the RNA polymerase. Indeed, deletion of one bp (leading to a 34° rotation) between the −10 region and the CRP binding site at the *galP1* promoter, abolishes activation and even results in a slight inhibition by CRP (Lavigne *et al.*, 1992b).

DNA-bending is important in this model only to ensure a precise juxtaposition of the interaction regions on CRP and α (Figure 2C). A second role for DNA-bending in this situation is suggested from experiments where activation of *galP1* by CRP has been measured using fragments of DNA that are truncated at different positions upstream (Lavigne *et al.*, 1992a). CRP can accelerate isomerization only on a DNA fragment carrying at least 130 bp upstream of the transcription start site. A possible explanation is sketched in Figure 2B. Interactions between the far-upstream DNA and the 'back' of RNA polymerase could stabilize the untwisting of the spacer DNA in the same way as the direct protein-protein contacts. Since wrapping of DNA around RNA polymerase is more extensive in the open complex than in the closed complex, the same kind of interactions (contacts between the far-upstream DNA and the back of RNA polymerase) could favor initial binding (K_B) as well as isomerization (k_2).

DNA Bending Re-Positions a Proximal Activator The two activators MalT and AraC bind to sites roughly equivalent in location to the CRP sites of class II promoters, i.e. overlapping the −35 region of a classical promoter (Schleif, 1996). Their mechanism of action is based on direct protein-protein contacts with RNA polymerase, similar to the ones described in the previous paragraph for CRP. In both cases however, higher order nucleoprotein structures are formed that dictate the local binding geometry near the promoter. In the case of AraC a repression loop involving an upstream operator site is formed in the uninduced state. This repression loop prevents the correct positioning of AraC on the promoter proximal site and thus impedes transcription activation (Schleif, 1996). CRP binding to a site within the loop bends the DNA in the opposite direction and thus breaks the repression loop (Lobell and Schleif, 1991). A similar mechanism operates at the *malKp* promoter. Several binding sites for the transcriptional activator MalT and three CRP binding sites are spread over a 271 bp region located between two divergently transcribed promoters. The MalT proteins bound to their sites form a higher order nucleoprotein structure, probabley via protein-protein interactions. This structure is organized by strategically placed DNA bends introduced by bound CRP. Experiments that replace CRP by another DNA-bending protein demonstrate that CRP plays a purely structural role (Richet and Sogaard Andersen, 1994). The establishment of the defined nucleoprotein complex leads to the displacement of the promoter proximally bound MalT proteins to alternative binding sites that are correctly phased for an optimal interaction with RNA polymerase (Richet *et al.*, 1991). The DNA conformation plays the important role, in both cases, of organizing a higher order structure. The formation of the structure is possible because of the increased local concentration of the constituents as a result of appropriate deformations of the DNA.

Indirect Transcriptional Activation

DNA Wrapping around RNA Polymerase It is more difficult to imagine direct protein-protein contacts between the activator and RNA polymerase when the activator binding site is moved further upstream. The target of a number of activators is the C-terminal domain (CTD) of the α-subunit of RNA polymerase (Ebright and Busby, 1995). This domain is attached to the rest of the protein via a flexible linker of a least 13 amino-acids and could thus follow the activator to a certain extent (Blatter *et al.*, 1994; Jeon *et al.*, 1997). Experiments using variants of the CRP-activated *malT* promoter show that transcription activation is hardly diminished when the CRP-binding site is moved from −60 to −80 (Dethiollaz *et al.*, 1996). Footprinting experiments using these same promoters show directly that at least one of the α-subunits remains bound to the DNA immediately upstream of the promoter (Eichenberger *et al.*, 1996). DNA upstream of the CRP binding site is necessary for activation (Dethiollaz *et al.*, 1996). We have therefore proposed that additional contacts between this far-upstream DNA and the back of RNA polymerase (Figure 2D) are responsible for activation. UV-laser footprinting signals in the far-upstream DNA confirm interactions in this region (Eichenberger *et al.*, 1996). Such an extended nucleoprotein structure is not unique to the *malT* promoter, but had also been observed at the *lac* promoter (Buckle *et al.*, 1992).

Kinetic UV-laser footprinting (Eichenberger, unpublished) demonstrates that these far-upstream contacts are formed at a late stage of transcription, after formation of RP_O. *MalT* is an example of a promoter limited at escape (Menendez *et al.*, 1987). We therefore propose that these far-upstream contacts accelerate promoter escape. Indeed, CRP stabilizes open forms of the promoter complex, including the open complex, but also the abortively initiating complexes (Eichenberger, unpublished), as shown in the free energy diagram of Figure 1C. The steady state rate of transcription is increased by providing more substrate (RP_O) for the rate-limiting step (escape) without directly affecting promoter escape. Similar interactions are present at the *lac* promoter (see above) but do not function to activate transcription because a non rate-limiting step is activated (see Figure 1A). Indeed, CRP is dispensable after the formation of the open complex at *lac* (Tagami and Aiba, 1995).

Other examples of indirect activation via bending of DNA include promoters activated by IHF (Goosen and van de Putte, 1995). One crucial hallmark that distinguishes this indirect mode of activation from direct protein-protein contacts is the orientation-dependence of activation: the activating region of only one subunit of a dimeric activator touches RNA polymerase (Zhou *et al.*, 1994b). Inversion of the subunits abolishes activation. IHF is a heterodimer with only 25% identity between the subunits, making it unlikely that an activation patch present on one subunit is conserved on the other subunit. However, the bacteriophage Mu early promoter is activated by IHF in both orientations (van Ulsen, 1997). Variants of the *malT* promoter are also activated by IHF in a distance-independent (Dethiollaz *et al.*, 1996) and orientation-independent manner (Engelhorn, unpublished). The simplest explanation of these data invokes indirect activation via back side contacts as shown in Figure 2D or by Goosen and van de Putte (1995).

Structural Transmission The information to accelerate transcription initiation could also be transmitted to RNA polymerase via a conformational change of the DNA linking the activator and the promoter. Evidence for such a mode of activation has been obtained at the *lac* promoter (Ryu *et al.*, 1994). Introducing a short stretch of single-stranded DNA between the CRP-binding site and the promoter abolishes activation. Because the single-stranded region increases the conformational entropy of the system it is difficult to distinguish between an incorrectly positioned activating region on CRP and structural transmission via the DNA from these experiments. Direct demonstration of structural transmission of a DNA deformation was obtained at the *ilvGMEDA* promoter (Parekh and Hatfield, 1996). Binding of IHF to a site upstream causes limited

base-pair opening in the –10 region of the promoter *in vitro* on supercoiled templates. It had been shown previously that the placement of a CRP-induced DNA bend fixes the orientation of a closed circular DNA molecule (Lavigne *et al.*, 1994). The positioning of the –10 region at the inside or the outside of a superhelical structure may affect its propensity to base-pair opening. Distortion of the DNA at a site far upstream may thus be transmitted to the promoter via the intervening DNA. DNA-bending in such a model serves to provide the initial distortion of the DNA.

Conclusions

Transcription initiation is a multi-step process involving extensive transconformation reactions of the DNA. Transcriptional regulators have evolved to take advantage of the particular DNA geometry at each step of the reaction by favoring or impeding their formation. DNA-bending serves to correctly align interacting groups in space, adjust the local concentration of an activator to the appropriate level, or favor more extensive interactions of promoter DNA with RNA polymerase by assisting the wrapping of DNA around RNA polymerase. Distortions of the DNA can be transmitted over a large distance by making use of the dynamics of topologically closed domains of DNA. The examples cited were intended to illustrate the importance of the conformation of DNA for transcription initiation and the great regulatory potential offered by 'simple' manipulations of DNA conformation. Each individual step of initiation is sensitive to DNA conformation and constitutes a target for transcriptional activators that manipulate the conformation of DNA.

Acknowledgement

This work was partly supported by grant 31-42501.94 from the Swiss National Fund.

References

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., and Ebright, R.H. (1994). Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell *78*, 889–896.

Brahms, G., Brahms, S., and Magasanik, B. (1995). A sequence-induced superhelical DNA segment serves as transcriptional enhancer. J. Mol. Biol. *246*, 35–42.

Brody, E.N., Kassavetis, G.A., Ouhammouch, M., Sanders, G.M., Tinker, R.L., and Geiduschek, E.P. (1995). Old phage, new insights: two recently recognized mechanisms of transcriptional regulation in bacteriophage T4 development. FEMS Microbiol. Lett. *128*, 1–8.

Buc, H., and McClure, W.R. (1985). Kinetics of open complex formation between *Escherichia coli* RNA polymerase and *lacUV5* promoter. Evidence for a sequential mechanism involving three steps. Biochemistry *24*, 2712–2723.

Buc, H., Amouyal, M., Buckle, M., Herbert, M., Kolb, A., Kotlarz, D., Menendez, M., Rimsky, S., Spassky, A., and Yeramian, E. (1987). Activation of transcription by the cyclic AMP receptor protein. In: RNA-polymerase and the Regulation of Transcription. W.S. Reznikoff, R.R. Burgess, J.E. Dahlberg, C.A. Gross, M.T. Record and M.P. Wickens, eds. (Amsterdam: Elsevier), pp. 115–125.

Buckle, M., Buc, H., and Travers, A.A. (1992). DNA deformation in nucleoprotein complexes between RNA polymerase, cAMP receptor protein and the lac UV5 promoter probed by singlet oxygen. EMBO J. *11*, 2619–2625.

Busby, S., and Buc, H. (1987). Positive regulation of gene expression by cyclic AMP and its receptor protein in *Escherichia coli*. Microbiol. Sci. *4*, 371–375.

Busby, S., and Ebright, R.H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell *79*, 743–746.

Busby, S., and Ebright, R.H. (1997). Transcription activation at class II CAP-dependent promoters. Mol. Microbiol. *23*, 853–859.

Carmona, M., and Magasanik, B. (1996). Activation of transcription at sigma 54-dependent promoters on linear templates requires intrinsic or induced bending of the DNA. J. Mol. Biol. *261*, 348–356.

Carpousis, A.J., and Gralla, J.D. (1980). Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation *in vitro* at the *lacUV5* promoter. Biochemistry *19*, 3245–3253.

Chen, Y., and Helmann, J.D. (1997). DNA-melting at the Bacillus subtilis flagellin promoter nucleates near –10 and expands unidirectionally. J. Mol. Biol. *267*, 47–59.

Claverie-Martin, F., and Magasanik, B. (1991). Role of integration host factor in the regulation of glnHp2 promoter of *Escherichia coli*. Proc. Natl. Acad. Sci. USA *88*, 1631–1635.

Cowing, D.W., Mecsas, J., Record, M.T., Jr., and Gross, C.A. (1989). Intermediates in the formation of the open complex by RNA polymerase holoenzyme containing the sigma factor s^{32}. J. Mol. Biol. *210*, 521–530.

Craig, M.L., Suh, W.C., and Record, M.T., Jr. (1995). HO. and DNase I probing of E sigma 70 RNA polymerase-lambda PR promoter open complexes: Mg^{2+} binding and its structural consequences at the transcription start site. Biochemistry *34*, 15624–15632.

Darst, S.A., Kubalek, E.W., and Kornberg, R.D. (1989). The three-dimensional structure of Escherichia coli RNA polymerase holoenzyme determined by electron crystallography. Nature *340*, 730–732.

deHaseth, P.L., and Helmann, J.D. (1995). Open complex formation by Escherichia coli RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol. Microbiol. *16*, 817–824.

Dethiollaz, S., Eichenberger, P., and Geiselmann, J. (1996). Influence of DNA geometry on transcriptional activation in Escherichia coli. EMBO J. *15*, 5449–5458.

Dove, S.L., Joung, J.K., and Hochschild, A. (1997). Activation of prokaryotic transcription through arbitrary protein-protein contacts. Nature *386*, 627–630.

Duval-Valentin, G., and Ehrlich, R. (1987). Dynamic and structural characterization of multiple steps during complex formation between *E. coli* RNA polymerase and *tetR* promoter from pSC101. Nucleic Acids Res. *15*, 575–594.

Dykxhoorn, D.M., St. Pierre, R., and Linn, T. (1996). Synthesis of the beta and beta' subunits of *Escherichia coli* RNA polymerase is autogenously regulated *in vivo* by both transcriptional and translational mechanisms. Mol. Microbiol. *19*, 483–493.

Ebright, R.H., and Busby, S. (1995). *Escherichia coli* RNA polymerase alpha subunit: structure and function. Curr. Op. Genet. Dev. *5*, 197–200.

Eichenberger, P., Déthiollaz, S., Fujita, N., Ishihama, A., and Geiselmann, J. (1996). Influence of the Location of the cAMP Receptor Protein Binding Site on the geometry of a Transcriptional Activation Complex in *Escherichia coli*. Biochemistry *35*, 15302–15312.

Ellinger, T., Behnke, D., Bujard, H., and Gralla, J.D. (1994). Stalling of *Escherichia coli* RNA polymerase in the +6 to +12 region *in vivo* is associated with tight binding to consensus promoter elements. J. Mol. Biol. *239*, 455–465.

Fersht, A. (1977). Enzyme Structure and Mechanism (San Francisco: W.H. Freeman and Company).

Goosen, N., and van de Putte, P. (1995). The regulation of transcription initiation by integration host factor. Mol. Microbiol. *16*, 1–7.

Gralla, J.D. (1996). Activation and repression of *E. coli* promoters. Curr. Op. Genet. Dev. *6*, 526–530.

Gralla, J.D., and Collado-Vides, J. (1996). Organization and function of transcription regulatory elements. In *Escherichia coli* and *Salmonella*. F. Neidhardt, ed. (Washington: American Society of Microbiology), pp. 1232–1246.

Ishihama, A. (1993). Protein-protein communication within the transcription apparatus. J. Bacteriol. *175*, 2483–2489.

Jeon, Y.H., Yamazaki, T., Otomo, T., Ishihama, A., and Kyogoku, Y. (1997). Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. J. Mol. Biol. *267*, 953–962.

Kammerer, W., Deuschle, U., Gentz, R., and Bujard, H. (1986). Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. EMBO J. *5*, 2995–3000.

Komissarova, N., and Kashlev, M. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. Proc. Natl. Acad. Sci. USA *94*, 1755–1760.

Kovacic, R.T. (1987). The 0°C closed complex between *Escherichia coli* RNA polymerase and two promoters, T7-A3 and lacUV5. J. Biol. Chem. *262*, 13654–13661.

Krummel, B., and Chamberlin, M.J. (1989). RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in early ternary complexes. Biochemistry *28*, 7829–7842.

Lavigne, M., Herbert, M., Kolb, A., and Buc, H. (1992a). Upstream curved sequences influence the initiation of transcription at the Escherichia coli galactose operon. J. Mol. Biol. *224*, 293–306.

Lavigne, M., Kolb, A., and Buc, H. (1992b). Transcription activation by cAMP receptor protein (CRP) at the Escherichia coli gal P1 promoter. Crucial role for the spacing between the CRP binding site and the –10 region. Biochemistry *31*, 9647–9656.

Lavigne, M., Kolb, A., Yeramian, E., and Buc, H. (1994). CRP fixes the rotational orientation of covalently closed DNA molecules. EMBO J. *13*, 4983–4990.

Lobell, R.B., and Schleif, R. (1991). AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein. J. Mol. Biol. *218*, 45–54.

McClure, W.R. (1985). Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. *54*, 171–204.

Menendez, M., Kolb, A., and Buc, H. (1987). A new target for CRP action at the *malT* promoter. EMBO J. *6*, 4227–4234.

Niu, W., Kim, Y., Tau, G., Heyduk, T., and Ebright, R.H. (1996). Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. Cell *87*, 1123–1134.

Parekh, B.S., and Hatfield, G.W. (1996). Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model. Proc. Natl. Acad. Sci. USA *93*, 1173–1177.

Perez Martin, J., and De Lorenzo, V. (1995). Integration host factor suppresses promiscuous activation of the s^{54}-dependent promoter *Pu* of *Pseudomonas putida*. Proc. Natl. Acad. Sci. USA *92*, 7277–7281.

Perez Martin, J., and Espinosa, M. (1994). Correlation between DNA bending and transcriptional activation at a plasmid promoter. J. Mol. Biol. *241*, 7–17.

Perez Martin, J., Rojo, F., and De Lorenzo, V. (1994). Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol. Rev. *58*, 268–290.

Rao, L., Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., Record, M.T., Jr., and Gourse, R.L. (1994). Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. J. Mol. Biol. *235*, 1421-1435.

Record, M.T., Jr., Reznikoff, W.S., Craig, M.L., McQuade, K.L., and Schlax, P.J. (1996). *Escherichia coli* RNA polymerase (Es^{70}), promoters, and the kinetics of the steps of transcription initiation. In *Escherichia coli* and *Salmonella*. F. Neidhardt, ed. (Washington: ASM Press), pp. 792–820.

Rees, W.A., Keller, R.W., Vesenka, J.P., Yang, G., and Bustamante, C. (1993). Evidence of DNA bending in transcription complexes images by scanning force microscopy. Science *260*, 1646–1649.

Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell *87*, 1295–1306.

Richet, E., and Sogaard Andersen, L. (1994). CRP induces the repositioning of MalT at the Escherichia coli malKp promoter primarily through DNA bending. EMBO J. *13*, 4558–4567.

Richet, E., Vidal Ingigliardi, D., and Raibaud, O. (1991). A new mechanism for coactivation of transcription initiation: repositioning of an activator triggered by the binding of a second activator. Cell *66*, 1185–1195.

Rippe, K., von Hippel, P.H., and Langowski, J. (1995). Action at a distance: DNA looping and initiation of transcription. Trends. Biochem. Sci. *20*, 500–506.

Ryu, S., Garges, S., and Adhya, S. (1994). An arcane role of DNA in transcription activation Proc. Natl. Acad. Sci. USA *91*, 8582–8586. [published erratum appears in Proc Natl Acad Sci USA 1994 Nov 8;91(23): 11281].

Schickor, P., Metzger, W., Werel, W., Lederer, H., and Heumann, H. (1990). Topography of intermediates in transcription initiation of *E. coli*. EMBO J. *9*, 2215–2220.

Schlax, P.J., Capp, M.W., and Record, M.T., Jr. (1995). Inhibition of transcription initiation by lac repressor. J. Mol. Biol. *245*, 331–350.

Schleif, R. (1996). Two positively regulated systems, *ara* and *mal*. In *Escherichia coli* and *Salmonella*. F. Neidhardt, ed. (Washington: ASM Press), pp. 1300–1309.

Straney, D.C., and Crothers, D.M. (1987). A stressed intermediate in the formation of stably initiated RNA chains at the *Escherichia coli lacUV5* promoter. J. Mol. Biol. *193*, 267–278.

Su, T.T., and McClure, W.R. (1994). Selective binding of Escherichia coli RNA polymerase to topoisomers of minicircles carrying the TAC16 and TAC17 promoters. J. Biol. Chem. *269*, 13511–13521.

Suh, W.C., Ross, W., and Record, M.T., Jr. (1993). Two open complexes and a requirement for Mg2+ to open the lambda PR transcription start site. Science *259*, 358–361.

Tagami, H., and Aiba, H. (1995). Role of CRP in transcription activation at Escherichia coli lac promoter: CRP is dispensable after the formation of open complex. Nucleic Acids Res. *23*, 599–605.

van Ulsen, P. (1997). Regulation of transcription initiation from the early promoter of bacteriophage Mu. 125–132. (Abstract)

von Hippel, P.H., and Berg, O.G. (1989). Facilitated target location in biological systems. J. Biol. Chem. *264*, 675–678.

von Hippel, P.H., and Yager, T.D. (1992). The elongation-termination decision in transcription. Science *255*, 809–812.

von Hippel, P.H., Rees, W.A., Rippe, K., and Wilson, K.S. (1996). Specificity mechanisms in the control of transcription. Biophys. Chem. *59*, 231–246.

Yager, T.D., and von Hippel, P.H. (1991). A thermodynamic analysis of RNA transcript elongation and termination in Escherichia coli. Biochemistry *30*, 1097–1118.

Zhou, Y., Merkel, T.J., and Ebright, R.H. (1994a). Characterization of the activating region of Escherichia coli catabolite gene activator protein (CAP). II. Role at Class I and class II CAP-dependent promoters. J. Mol. Biol. *243*, 603–610.

Zhou, Y., Pendergrast, P.S., Bell, A., Williams, R., Busby, S., and Ebright, R.H. (1994b). The functional subunit of a dimeric transcription activator protein depends on promoter architecture. EMBO J. *13*, 4549–4557.

Zinkel, S.S., and Crothers, D.M. (1991). Catabolite activator protein-induced DNA bending in transcription initiation. J. Mol. Biol. *219*, 201–215.

Biol. Chem., Vol. 378, pp. 609 – 616, July 1997

Review

Oxygen(es) and the Hypoxia-Inducible Factor-1

Roland H. Wenger and Max Gassmann

Institute of Physiology, University of Zürich-Irchel, Winterthurerstr. 190, CH-8057 Zürich, Switzerland

The hypoxia-inducible factor-1 (HIF-1) is a basic-helix-loop-helix (bHLH) heterodimeric transcription factor activated by reductions in oxygen concentration (hypoxia). Activated HIF-1 upregulates expression of genes involved in the adaptation of higher organisms to hypoxic conditions, caused by e.g. high altitude, anemia, wound healing or during development. Examples of these oxy-genes include erythropoietin, a hormone regulating erythropoiesis and hence the oxygen transport capacity, and vascular endothelial growth factor, a potent inducer of angiogenesis leading to increased blood capillary density. The HIF-1 heterodimer is composed of a HIF-1α and an ARNT subunit, both belonging to the explosively growing PAS subfamily of bHLH transcription factors. Closely related, but differentially expressed, factors have recently been cloned, at least one of which can also be activated by hypoxia. In this review, we present a survey of the bHLH-PAS family as well as of the genes regulated by HIF-1, and we summarize our current knowledge on the oxygen-dependent activation of this fascinating transcription factor.

Key words: p300/CBP / Dioxin receptor / Erythropoietin / Glucose metabolism / Transcription factor / Vascular endothelial growth factor.

Introduction

Almost two years ago, the cloning of a novel transcription factor that is activated by reductions in oxygen (O_2) partial pressure (hypoxia) attracted not only the attention of physiologists but also of scientists working in such diverse fields as dioxin toxicology and basic molecular biology. As happens often in biological research, the hunt for answers to completely different questions culminated in the identification of one and the same protein. Physiologists have been (and still are) investigating the molecular mechanisms by which reduced oxygenation influences the expression of specific genes such as erythropoietin (Epo), the main regulator of erythropoiesis. Although far from completely understanding how cells sense changes in O_2 concentration and transduce this signal to the nucleus, a hypoxia-inducible factor (HIF-1) was identified that transactivates Epo gene expression following binding to a hypoxia response element located in the 3′ flanking region of the Epo gene. Semenza and co-workers succeeded in the molecular identification of this factor and demonstrated that it consists of two subunits termed HIF-1α and HIF-1β (Wang *et al.*, 1995a). Whereas HIF-1α represents a newly detected protein, it came quite as a surprise that HIF-1β is identical to the heterodimerization partner of the dioxin receptor/aryl hydrocarbon receptor (AhR), called AhR nuclear translocator (ARNT). Subsequently, the cloning of a HIF-1α cDNA was also reported as a result of the search for new heterodimerization partners of ARNT (Li *et al.*, 1996), as well as by virtue of binding the transcriptional co-activator p300/CBP (Arany *et al.*, 1996).

Hypoxia-Inducible Epo Expression and the Discovery of HIF-1

Oxygen serves as the terminal electron acceptor in mitochondria and hence is essential for the life of almost all higher organisms. One of the most dramatic and best-studied mechanisms by which vertebrates adapt to hypoxia is the increased rate of erythrocyte production, leading to an elevation of the O_2 transport capacity. This response is mediated by the glycoprotein Epo (reviewed by Jelkmann, 1992). Upon hypoxic exposure, the kidney, and to a lesser extent also the liver, increase Epo synthesis at the level of both transcription and mRNA stability. Once released into the circulation, Epo is transported to the bone marrow where it binds to the Epo receptor present on erythroid progenitor cells. A signalling cascade is then initiated that prevents apoptosis of these erythroid precursor cells allowing them to differentiate to mature erythrocytes.

Analyses of both, human hepatoma cell lines capable of expressing Epo in an O_2-dependent manner, and of transgenic mouse models, allowed the identification of a hypoxia-response element (HRE) located downstream of the Epo gene, which is crucial for hypoxic induction of Epo transcription. DNA binding studies revealed that several factors interact with this HRE, one of which (HIF-1) bound solely after hypoxic exposure (reviewed by Bunn and Poyton, 1996; Guillemin and Krasnov, 1997). Subsequently, Ratcliffe and co-workers (Maxwell *et al.*, 1993), as well as Semenza and co-workers (Wang and Semenza, 1993a), found this O_2-dependent DNA binding and transactivation activity also in various non-Epo producing cell lines, including insect cells (Nagao *et al.*, 1996), suggesting that HIF-1 is part of a widespread O_2-sensing and signal transducing mechanism. Biochemical studies revealed that HIF-1 is a heterodimer composed of an α subunit (120 kDa) and a β subunit (91 to 94 kDa), both directly

contacting DNA in the major groove (Wang and Semenza, 1993b). Protein purification and microsequencing allowed the identification of HIF-1α as a novel protein and HIF-1β as ARNT (Wang *et al.*, 1995a).

In contrast to other transcription factors, including NF-κB, AP-1 and the tumor suppressor p53, all of them being induced by either (near-) anoxic stress conditions (where the O_2 concentration is below approximately 0.02% v/v) or even reoxygenation-dependent processes, HIF-1 is operating at physiologically relevant O_2 concentrations. A recent report demonstrated that HIF-1α (and to a lesser extent also ARNT) protein levels and HIF-1 DNA binding activity increased exponentially upon exposure to decreased O_2 concentrations (Jiang *et al.*, 1996a). Half maximal induction was found between 1.5% and 2% O_2, and maximal response was at 0.5% O_2. Below 0.5% O_2, HIF-1 protein levels decreased again. For comparison, the corresponding normoxic O_2 concentration in kidney and liver tissues is approximately 3 to 5%. Thus, HIF-1 is the first ubiquitously expressed transcription factor known to be activated within a physiological range of O_2 concentrations.

HIF-1α Is a bHLH-PAS Family Member

Both HIF-1α and ARNT belong to a subfamily of the basic-helix-loop-helix (bHLH) class of transcription factors (Figure 1) harboring a region of homology termed PAS (**P**er, **A**RNT, **S**im). The PAS domain was first found in the *Drosophila* transcription factors Per and Sim, and contains direct internal repeats (A and B) of approx. 50 to 60 amino acids (reviewed by Schmidt and Bradfield, 1996). Per (period) forms a heterodimer with Tim (timeless) which is involved in control of the circadian rhythm of *Drosophila*. Sim (single-minded) is a regulator of neuronal development, and a mammalian Sim homolog was identified as a candidate gene responsible for Down's syndrome (Chen *et al.*, 1995; Dahmane *et al.*, 1995; Fan *et al.*, 1996). Interestingly, two recently detected mouse Sim proteins (mSim1 and mSim2) can form heterodimers with ARNT and act as transcriptional repressors (Ema *et al.*, 1996; Probst *et al.*, 1997). Two additional *Drosophila* bHLH-PAS genes have recently been cloned: Trh (trachealess) and Sima (similar). Trh is required for embryonic tube formation (Isaac and Andrew, 1996; Wilk *et al.*, 1996). Sima is most similar to HIF-1α and might therefore represent the *Drosophila* HIF-1α homolog (Nambu *et al.*, 1996).

Mammalian ARNT is a nuclear protein known to heterodimerize with AhR following activation by xenobiotic ligands such as dioxin and other aryl hydrocarbons. This complex binds to the xenobiotic response element that controls expression of genes involved in xenobiotic metabolism including cytochrome P-450IA1 and glutathione *S*-transferase Ya. In addition, a homolog of ARNT (ARNT2), whose expression is restricted to brain and kidney, has also been reported (Hirose *et al.*, 1996). Very recently, three novel bHLH-PAS factors, termed endothelial

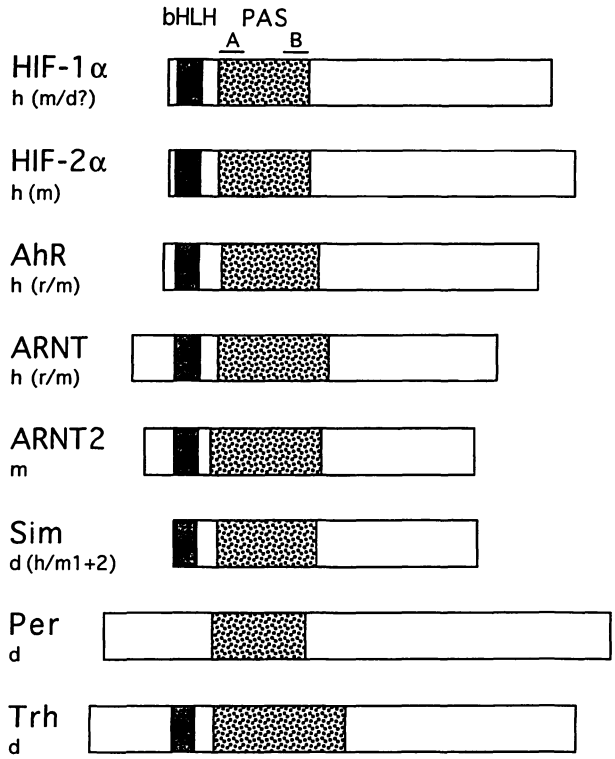


Fig. 1 Schematic Representation of Some bHLH-PAS Family Members.
The bHLH and PAS domains are indicated by filled and hatched boxes, respectively. Whereas the PAS domain depicted here comprises only the A and B direct repeat, the PAS homology domain extends somewhat beyond this region. The letters on the left indicate the species from which the respective genes were cloned, and letters in parentheses indicate other cloned homologs. HIF-2α is also called EPAS1, HLF or HRF. The species are: h, human; r, rat; m, mouse; d, *Drosophila*.

PAS domain protein 1 (EPAS1, Tian *et al.*, 1997), HIF-1α-*like* factor (HLF, Ema *et al.*, 1997), and HIF-related factor (HRF, Flamme *et al.*, 1997) with marked sequence similarity to HIF-1α have been published. Comparison of the amino acid sequences suggested that EPAS1, HLF and HRF are identical (R.H.W. and M.G., unpublished observation). In contrast to the ubiquitous expression of HIF-1α, EPAS1/HLF/HRF mRNA expression is restricted to endothelial cells during embryogenesis and has predominantely been observed in highly vascularized tissues in adults and in postnatal alveolar epithelial cells. Since EPAS1/HLF/HRF

(i) is very similar to HIF-1α,
(ii) forms functional heterodimers with ARNT,
(iii) is activated by hypoxia, and
(iv) transactivates genes via the same DNA recognition sites as HIF-1,

we propose that this novel bHLH-PAS family member should be termed 'HIF-2α' and the corresponding HIF-2α-ARNT heterodimeric complex 'HIF-2'. The endothelial cell-specific receptor tyrosine kinase Tie-2 (Tian *et al.*, 1997) as well as the vascular endothelial growth factor (VEGF, Ema *et al.*, 1997) were identified as targets of HIF-2

function. To date, besides HIF-1α, HIF-2α is the only bHLH-PAS family member known to be induced by hypoxia.

In summary, the bHLH-PAS family is growing rapidly and additional members are continuously found in the EST treasury (Hogenesch *et al.*, 1997; Zhou *et al.*, 1997). Functional characterization of the corresponding gene products certainly will provide many interesting insights into the complex network of heterodimer formation between AhR/HIF-1α/HIF-2α/Sim-type and ARNT/ARNT2-type of bHLH-PAS proteins, and will identify the target genes regulated by these factors.

Functional Domains of HIF-1

Much of our current knowledge about the functional aspects of bHLH-PAS proteins is derived from investigations on the AhR/ARNT complex. Both proteins contain N-terminal basic and HLH domains responsible for DNA binding and dimerization, respectively, whereas the transactivation domains reside in a glutamine-rich region near the C-terminus. The role of the PAS domain is less clear, but it is involved in heterodimerization, influences DNA binding and transactivation, and bears the ligand as well as the heat shock protein 90 (hsp90) binding sites of the AhR (Schmidt and Bradfield, 1996). *In vitro* translation of deletion mutants coupled with co-immunoprecipitation and DNA binding experiments revealed that the presence of the HLH and PAS(A) domains in HIF-1α as well as in ARNT is indispensable for HIF-1 heterodimer formation. On the other hand, the basic domain and a region including the PAS(B) domain are required for DNA binding (Gradin *et al.*, 1996; Jiang *et al.*, 1996b). These findings were confirmed by reconstitution experiments in an ARNT-deficient mouse hepatoma cell line (Hepa1C4), demonstrating that the bHLH-PAS domain but not the C-terminal part of ARNT is necessary for hypoxic induction of transactivation (Li *et al.*, 1996; Wood *et al.*, 1996). Thus, although ARNT was shown to be principally able to homodimerize when expressed at high concentrations and displays a certain transactivation capability, its primary function might be to serve as a heterodimerization partner for other bHLH-PAS members, including AhR, HIF-1α, HIF-2α and Sim.

Co-expression of fusion proteins between HIF-1α and a heterologous DNA binding domain together with a reporter gene construct in Hepa1C4 cells, revealed that the transactivation function of HIF-1α can also be induced by hypoxia in the absence of ARNT, underlining the predominant role of HIF-1α compared to ARNT in hypoxic gene regulation (Li *et al.*, 1996). A more detailed dissection of the C-terminal region of HIF-1α revealed that deletion of the last 13 amino acids resulted in a reduced hypoxic inducibility, and that the last 83 amino acids were sufficient to convey hypoxia-inducible transactivation to a heterologous reporter gene system (Li *et al.*, 1996; Jiang *et al.*, 1996b).

Oxy-genes Regulated by HIF-1

We recently showed that transcripts from the genes encoding HIF-1α and ARNT are expressed ubiquitously in human and mouse tissues (Gradin *et al.*, 1996; Wenger *et al.*, 1996). This is in agreement with the widespread hypoxia-inducible HIF-1 DNA binding and transactivation activity observed in virtually all permanent cell lines investigated so far (Maxwell *et al.*, 1993; Wang and Semenza, 1993a). Figure 2 lists the genes that have been identified to date as targets of HIF-1 function. These include Epo, VEGF, several glycolytic enzymes, glucose-transporter 1, inducible nitric oxide synthase, a VL30 retrotransposon, heme oxygenase 1 and transferrin. By aligning the known HIF-1 binding sites (HBSs) described from these genes, a tentative HIF-1 consensus DNA recognition site was obtained with CGTG as the conserved core motif, usually preceeded by an adenosine residue and an underrepresentation of adenosine residues in the 3′ flanking part (Figure 2). However, determination of the significance and probability of a given position flanking the inner core motif CGTG awaits a more complete list of functional HIF-1 DNA binding sites.

The ubiquitous expression of functional HIF-1, as well as the growing number of genes that are regulated by this transcription factor, strongly suggest that HIF-1 plays a key role in O_2-dependent gene expression. Further evidence for this notion arises from experiments with wild-type Hepa1 cells and its ARNT-deficient subline Hepa1C4. This subline was originally selected for deficiency in the induction of benzo[a]pyrene-metabolizing enzyme activity. As expected, hypoxic exposure of these cells no longer resulted in induced HIF-1 DNA binding activity nor did it induce expression from reporter gene constructs containing HREs. Furthermore, hypoxic upregulation of most endogenous genes was abrogated in the mutant Hepa1C4 cells, but this function was rescued by forced expression of exogenous ARNT (Forsythe *et al.*, 1996; Gradin *et al.*, 1996; Li *et al.*, 1996; Salceda *et al.*, 1996; Wood *et al.*, 1996; Gassmann *et al.*, 1997). Interestingly, in some reports (Gradin *et al.*, 1996; Wood *et al.*, 1996; Gassmann *et al.*, 1997) VEGF and Glut-1 mRNAs were found to be still induced in hypoxic Hepa1C4 cells, albeit to a reduced level compared to the parental Hepa1 cells. The reason for this phenomenon is currently unknown, but one might speculate that more than one O_2-dependent signalling pathway exists, one of which being ARNT (HIF-1)-independent. It is a striking coincidence that for both VEGF and Glut-1, post-transcriptional mRNA stabilization has been reported to contribute to the increased steady-state mRNA levels under hypoxia (Ikeda *et al.*, 1995; Stein *et al.*, 1995), a process that involves the von Hippel-Lindau tumor suppressor gene VHL (Iliopoulos *et al.*, 1996; Levy *et al.*, 1996).

Hypoxia-inducible gene	Species	Location	HIF-1 binding site	Function	Reference
erythropoietin	human	3'FS +3065/+3082	gccc TACGTGCTGT ctca	+	Wang *et al.*, 1993b
erythropoietin	mouse	3'FS +359/+376	gccc TACGTGCTGC ctcg	+	Firth *et al.*, 1994
erythropoietin	human	5'FS -163/-180	tgca GACGTGCGTG tggg	+	Semenza *et al.*, 1994
phosphofructokinase L	mouse	IVS1 +346/+361	ggcg TACGTGCTGC ag	+	Semenza *et al.*, 1994
aldolase A	human	5'FS -184/-201	ctcg GACGTGACTC ggac	+	Semenza *et al.*, 1996
aldolase A	human	IVS4 +124/+141	tctt CACGTGCGGG gacc	-	Semenza *et al.*, 1994
phosphoglycerate kinase 1	human	5'FS -173/-190	gtga GACGTGCGGC ttcc	+	Semenza *et al.*, 1994
phosphoglycerate kinase 1	human	5'FS -206/-189	gccg GACGTGACaa acgg	N.D.	Semenza *et al.*, 1996
phosphoglycerate kinase 1	mouse	5'FS -290/-307	ttgt CACGTcCTGC acga	+	Firth *et al.*, 1994
phosphoglycerate kinase 1	human	5'UT +30/+13	tgcc GACGTGCGCT ccgg	N.D.	Semenza *et al.*, 1994
enolase 1	human	5'FS -590/-607	gggc CACGTGCGCC gcct	-	Semenza *et al.*, 1994
enolase 1	human	5'FS -413/-396	gccg GACGTGGGGC ccca	-	Semenza *et al.*, 1996
enolase 1	human	5'FS -386/-369	tgag TGCGTGCGGG actc	+	Semenza *et al.*, 1996
enolase 1	human	5'FS -368/-351	ggag TACGTGACGG agcc	+	Semenza *et al.*, 1996
lactate dehydrogenase A	mouse	5'FS -89/-72	ccta CACGTGGGTT cccg	+	Firth *et al.*, 1995
lactate dehydrogenase A	mouse	5'FS -62/-79	agcg GACGTGCGGG aacc	+	Firth *et al.*, 1995
glucose transporter 1	mouse	5'FS +273/+290	caca GGCGTGCCGT ctga	+	Ebert *et al.*, 1995
vascular endothelial growth factor	human	5'FS -978/-961	tgca TACGTGGGCT ccaa	+	Liu *et al.*, 1995
vascular endothelial growth factor	rat	5'FS +61/+78	tgca TACGTGGGCT tcca	+	Levy *et al.*, 1995
inducible nitric oxide synthase	human	5'FS -229/-212	tgac TACGTGCTGC ctag	+	Melillo *et al.*, 1995
retrotransposon VL30	rat	5'FS -101/-113	agag TACGTGCTa	+	Estes *et al.*, 1995
heme oxygenase 1	mouse	5'FS +322/+339	agcg GACGTGCTGG cgtg	+	Lee *et al.*, 1997
heme oxygenase 1	mouse	5'FS +350/+333	agag GACGTGCCaC gcca	+	Lee *et al.*, 1997
transferrin	human	5'FS -201/-184	gaaa TACGTGCGCT ttgt	+	Rolfs *et al.*, 1997
transferrin	human	5'FS -185/-168	tgtg TACGTGCaGG aaag	+	Rolfs *et al.*, 1997

Tentative HIF-1 consensus DNA recognition site:

$$\begin{array}{ccccccccc} T_{11} & A_{23} & \mathbf{CGTG} & C_{18} & G_{12} & G_{14} & C_9 \\ G_{10} & G_2 & & G_4 & T_7 & T_3 & T_7 \\ C_4 & & & A_3 & C_5 & C_5 & G_7 \end{array}$$

Fig. 2 Compilation of HIF-1 DNA Binding Sites.
Nucleotides that match the tentative HIF-1 consensus DNA binding sequence are in capitals. The core motif is in bold face. Note that the mouse phosphoglycerate kinase 1 HIF-1 site corresponds to the complementary human site shown above it; mismatches to the tentative consensus site were found in both cases. Cases where binding of HIF-1 to the corresponding sites leads to functional upregulation of a reporter gene are marked by a (+) sign. HIF-1 binding sites that are, despite proven HIF-1 binding, transactivation-inactive are marked by a (–) sign. FS, flanking sequence; IVS, intervening sequence; UT, untranslated; N.D., not determined.

Complex Protein Interactions at the HIF-1 DNA Binding Site

Despite its critical importance in hypoxic gene regulation, it should be noted that a single HBS within a HRE is apparently not sufficient to confer hypoxic induction of gene transcription. Additional factors binding to a HRE seem to be necessary for optimal hypoxic induction (Ebert *et al.*, 1995; Firth *et al.*, 1995; Forsythe *et al.*, 1996, Semenza *et al.*, 1996; and references therein). As in the case of the phosphoglycerate kinase 1, enolase 1, lactate dehydrogenase A, heme oxygenase 1 and transferrin genes, two functional HBSs in close vicinity may also constitute a HRE (Fig. 2). This hypothesis is further supported by the finding that a HBS in isolation did not confer hypoxic induction of reporter gene expression, whereas three concatemerized HBS copies were hypoxia-responsive (Firth *et al.*, 1994; Kvietikova *et al.*, 1995; Gradin *et al.*, 1996; Gassmann *et al.*, 1997). Furthermore, there is evidence that ATF and CREB family members are capable of constitutively binding the HBS (Kvietikova *et al.*, 1995).

Interestingly, in an approach to clone gene products interacting with the transcriptional co-activator p300/CBP, HIF-1α was identified as one of the binding partners of p300/CBP (Arany *et al.*, 1996). Inhibition of p300/CBP by the viral oncoprotein E1A blocked hypoxic gene activation, whereas ectopic expression of p300 enhanced hypoxia-induced transcription, demonstrating the critical importance of p300/CBP in HIF-1 function (Arany *et al.*, 1996). p300/CBP display histone acetyltransferase activities and are thus probably involved in chromatin rearrangement (Bannister *et al.*, 1996; Ogryzko *et al.*, 1996), which might be crucial for the accessibility of HIF-1 to its DNA target site.

How Is HIF-1 Regulated?

So far, very little is known about the mechanisms of O_2 sensing and subsequent conditional regulation of HIF-1. Initially, it was reported that HIF-1α, and to a lesser extent also ARNT, are hypoxically up-regulated at both the level

of mRNA and protein concentrations (Wang *et al.*, 1995a). However, the strong hypoxic increase in HIF-1α and ARNT mRNA could not be observed in other laboratories, despite a concomitant increase in HIF-1α and ARNT protein levels and DNA binding activity (Gradin *et al.*, 1996; Huang *et al.*, 1996; Salceda *et al.*, 1996; Wood *et al.*, 1996; Wenger *et al.*, 1996; 1997a). Therefore, even if variations in HIF-1α and ARNT mRNA levels may be found in some experimental settings, they do not account for the dramatic hypoxic increase in protein levels. How, then, is HIF-1 activated? The answer to this question is not only central to the understanding of HIF-1 function itself, but will also provide new insights into the enigma of the O_2-dependent signal transducing mechanisms, since HIF-1 activation may be the ultimate target of this process. Possible mechanisms include:

(i) translational upregulation and protein stability,
(ii) phosphorylation,
(iii) reduction-oxydation (redox) processes; and
(iv) ligand binding or intracellular localization.

Evidence for all of these putative mechanisms has been provided and more than one of them might be involved in hypoxic HIF-1 activation. In the following section we will briefly discuss our current knowledge about HIF-1 activation. We will concentrate mainly on HIF-1α, since this subunit is apparently unique to the O_2-dependent signal transduction pathway. In contrast, ARNT serves also as a heterodimerization partner for various other transcription factors, and hence is involved in other signal transduction pathways including xenobiotic metabolism. In addition, hypoxic ARNT activation is usually much less pronounced than regulation of HIF-1α (Wang *et al.*, 1995a; Huang *et al.*, 1996; Jiang *et al.*, 1996a). It cannot be ruled out, however, that hypoxic modulation of ARNT activity might also influence HIF-1 regulation.

(i) Intracellular protein levels might be regulated by translational control or by altering protein stability. It is well established that pre-treatment with the protein synthesis inhibitor cycloheximide blocks hypoxic induction of Epo (and other O_2-regulated) mRNA expression as well as HIF-1 activity (Bunn and Poyton, 1996). A plausible explanation for this observation points towards translational upregulation of HIF-1α protein. Upon reoxygenation, HIF-1α protein levels are rapidly downregulated indicating a very short half life (around 15 minutes) of HIF-1α under normoxic conditions (Wang *et al.*, 1995a). As reported recently by Bunn and colleagues, the stability of HIF-1α is markedly increased following hypoxic exposure (Huang *et al.*, 1996). Moreover, this group found that following anaerobic administration of cycloheximide to hypoxically growing cells, HIF-1α protein level was not affected, suggesting that an increase in protein stability rather than *de novo* translation might be the primary cause for hypoxic induction. However, the possibility that translational is also involved in this process cannot be excluded. In keeping with this, we recently showed that mouse HIF-1α is expressed from two alternative promoters resulting in two mRNA isoforms with different 5′ untranslated regions and ATG translation initiation codons. These 5′ untranslated regions vary markedly in their G+C contents which might influence the translation efficiency of each mRNA isoform (Wenger *et al.*, 1997b).

(ii) HIF-1 DNA binding activity depends on its phosphorylation status: addition of a protein kinase inhibitor (2-aminopurine) blocked both hypoxic HIF-1 and Epo mRNA induction, and treatment of nuclear extracts with phosphatase *in vitro* eliminated HIF-1 DNA binding (Wang *et al.*, 1995b). Moreover, addition of a phosphatase inhibitor (NaF) abolished hypoxic upregulation of HIF-1 activity, implying that phosphorylation plays an important role at multiple steps of the hypoxia signal transduction pathway. The c-Src tyrosine kinase had been suggested to be involved in hypoxic VEGF induction (Mukhopadhyay *et al.*, 1995). However, more detailed studies on HIF-1 regulation involving c-Src-deficient cells could not confirm a role for c-Src in the mechanism of transcriptional activation (Gleadle and Ratcliffe, 1997). Phosphorylation (and other post-translational protein modifications) could also account for the marked difference between the calculated molecular mass predicted from the cDNA (93 and 85/87 kDa for HIF-1α and ARNT, respectively) and the apparent molecular mass observed in polyacrylamide gels (120 and 91/93/94 kDa for HIF-1α and ARNT, respectively) (Wang and Semenza, 1993b; Wang *et al.*, 1995a).

(iii) Another important modification affecting HIF-1 activity involves the alteration of the cellular redox state. This point is of particular interest since superoxide (O_2^-) production is dependent on the O_2 concentration and since the superoxide dismutase conversion product hydrogen peroxide (H_2O_2) could serve as a second messenger. Therefore, redox processes initiated by a postulated heme O_2 sensor represent a candidate system for the O_2 sensing and signalling mechanism (Bunn and Poyton, 1996). Indeed, exogenous H_2O_2 added to the culture medium prior to hypoxia inhibited hypoxic induction of Epo expression (Fandrey *et al.*, 1994), and also blocked hypoxic HIF-1 activation, probably by preventing HIF-1α protein accumulation (Wang *et al.*, 1995c; Huang *et al.*, 1996). Interestingly, administration of H_2O_2 during hypoxia did not affect HIF-1 activity, once more suggesting an increase in protein stability (Huang *et al.*, 1996). However, in addition to the oxidizing agent H_2O_2, the reducing agent DTT also inhibited hypoxic HIF-1 activation, and the antioxidant ascorbate failed to induce HIF-1 (Wang *et al.*, 1995c), challenging the hypothesis that H_2O_2 is the second messenger of hypoxia signal transduction. *In vitro* treatment of purified HIF-1 and nuclear extracts with oxidizing (H_2O_2 or diamide) and alkylating (N-ethylmaleimide) agents completely abolished HIF-1 DNA binding, indicating that a functionally important cysteine sulfhydryl group is sensitive to oxidation. The presence of a reducing agent (DTT or β-mercaptoethanol) preserved HIF-1 function in these experiments (Wang *et al.*, 1995c; Huang *et al.*, 1996). Moreover, purified oxidized thioredoxin in the presence of DTT enhanced HIF-1 DNA binding and overexpression of exogenous thioredoxin potentiated hypoxic

HIF-1 transactivation function of a co-expressed reporter gene construct (Huang *et al.*, 1996). Thus, as for phosphorylation, these results suggest that redox processes are involved at multiple steps of O_2 signalling. However, the finding that forced expression of exogenous HIF-1α and ARNT resulted in increased transactivation of reporter gene expression already under normoxic conditions (Forsythe *et al.*, 1996; Gradin *et al.*, 1996; Li *et al.*, 1996) provides some evidence that hypoxia-dependent post-translational protein modifications are not the sole cause for HIF-1 activation.

(iv) Several transcription factors are known to be retained in an inactive form in the cytoplasm and, upon activation, are translocated to the nucleus. In the case of the AhR for example, the unliganded form resides in the cytoplasm as a complex with two molecules of hsp90. Following binding of xenobiotic ligands such as dioxins, hsp90 is released and the liganded AhR forms a heterodimeric complex with ARNT that activates genes involved in xenobiotic metabolism (Schmidt and Bradfield, 1996). It is tempting to speculate that a similar mechanism is also involved in hypoxic HIF-1 activation. In this model, following hypoxic stimulation, a so far undefined ligand would bind to HIF-1α, that in turn translocates to the nucleus and heterodimerizes with ARNT to form HIF-1. In support of this hypothesis, *in vitro* experiments showed that the affinity of hsp90 for HIF-1α binding seems to be even higher than the affinity for AhR binding (Gradin *et al.*, 1996). However, HIF-1α protein could only be detected in nuclear extracts of hypoxically treated cells but not in the cytoplasmic fraction (Wang *et al.*, 1995a), suggesting that, following synthesis, HIF-1α is either rapidly degraded in normoxia or stabilized and immediately translocated to the nucleus in hypoxia.

Perspectives

Clearly, the elucidation of the mechanisms leading to hypoxic activation of HIF-1 is the most demanding topic for the near future. Some of the possibilities raised above are rather speculative and further functional dissection of HIF-1α, HIF-2α and ARNT will be necessary to delineate a more precise picture. Moreover, the molecules transmitting the hypoxic signal to HIF-1, once identified, should provide useful tools to tackle the O_2 sensor itself by following the signalling pathway in the upward direction. On a more general level, HIF-1α, HIF-2α (EPAS1/HLF/HRF), ARNT and ARNT2 transgenic and 'knock-out' mouse models will provide invaluable information on the physiological role of these factors not only in the adaptation to hypoxia but also during development as well as in cardiac and neoplastic diseases. Indeed, in a very recent report, Maltepe *et al.* (1997) have shown that ARNT null mutant mouse embryos die at midgestation due to defective blood vessel formation, a phenotype reminiscent of that observed in VEGF +/– and tissue factor –/– mouse embryos. The finding that HIF-2α mRNA is expressed in vascular endothelial cells at midgestation, whereas HIF-1α mRNA is almost undetectable (Ema *et al.*, 1997), suggests that the lethal phenotype is probably due to the lack of HIF-2 (rather than HIF-1) complex formation. Even if a definitive answer awaits the targeted disruption of the HIF-1α and HIF-2α genes, the results on the ARNT –/– phenotype already imply that a hypoxia-inducible complex (HIF-1 and/or HIF-2) most probably plays an indispensable role during embryonic development.

Note Added in Proof

While this manuscript was processed for publication, Pugh *et al.* (1997) found that HIF-1α amino acids 549–582 and 775–826 could independently convey hypoxia-inducible responses. Of note, the 549–582 domain was also associated with a marked reduction of expressed protein levels in normoxic cells, which was relieved by exposure to hypoxia. Moreover, based on *in vitro* data, Kallio *et al.* (1997) presented a new model of HIF-1 regulation in which HIF-1α acquires a new conformational state upon dimerization with ARNT, rendering HIF-1α more resistant to proteolytic attack.

Acknowledgements

We thank H. F. Bunn, G. L. Semenza, P. J. Ratcliffe and L. Poellinger for providing preprints, and P. J. Ratcliffe, L. Poellinger, A. Görlach, I. Forster and J. Silke for critically reading the manuscript. Our work was supported by the Swiss National Science Foundation Grant 31-47111.96. R. H. W. is a recipient of the "Sondermassnahmen des Bundes zur Förderung des akademischen Nachwuchses".

References

Arany, Z., Huang, L.E., Eckner, R., Bhattacharya, S., Jiang, C., Goldberg, M.A., Bunn, H.F., and Livingston, D.M. (1996). An essential role for p300/CBP in the cellular response to hypoxia. Proc. Natl. Acad. Sci. USA *93*, 12969–12973.

Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. Nature *384*, 641–643.

Bunn, H.F., and Poyton, R.O. (1996). Oxygen sensing and molecular adaptation to hypoxia. Physiol. Rev. *76*, 839–885.

Chen, H., Chrast, R., Rossier, C., Gos, A., Antonarakis, S.E., Kudoh, J., Yamaki, A., Shindoh, N., Maeda, H., Minoshima, S., and Shimizu, N. (1995). Single-minded and Down syndrome? Nature Genet. *10*, 9–10.

Dahmane, N., Charron, G., Lopes, C., Yaspo, M.-L., Maunoury, C., Decorte, L., Sinet, P.-M., Bloch, B., and Delabar, J.-M. (1995). Down syndrome-critical region contains a gene homologous to Drosophila sim expressed during rat and human central nervous system development. Proc. Natl. Acad. Sci. USA *92*, 9191–9195.

Ebert, B.L., Firth, J.D., and Ratcliffe, P.J. (1995). Hypoxia and mitochondrial inhibitors regulate expression of glucose transporter-1 via distinct cis-acting sequences. J. Biol. Chem. *270*, 29083–29089.

Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H., Kikuchi, Y., and Fujii-Kuriyama, Y. (1996). Two new members of the murine sim gene family are transcriptional repressor and show different expression patterns during mouse embryogenesis. Mol. Cell. Biol. *16*, 5865–5875.

Ema, M., Taya, S., Yokotani, N., Sogawa, K., Matsuda, Y., and Fujii-Kuriyama, Y. (1997). A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1α regulates the VEGF expression and is potentially involved in lung and vascular development. Proc. Natl. Acad. Sci. USA *94*, 4273–4278.

Estes, S.D., Stoler, D.L., and Anderson, G.R. (1995). Anoxic induction of a sarcoma virus-related VL30 retrotransposon is mediated by a cis-acting element which binds hypoxia-inducible factor 1 and an anoxia-inducible factor. J. Virol. *69*, 6335–6341.

Fan, C.-M., Kuwana, E., Bulfone, A., Fletcher, C.F., Copeland, N.G., Jenkins, N.A., Crews, S., Martinez, S., Puelles, L., Rubenstein, J.L.R., and Tessier-Lavigne, M. (1996). Expression patterns ot two murine homologs of Drosophila single-minded suggest possible roles in embryonic patterning and in the pathogenesis of Down syndrome. Mol. Cell. Neurosci. *7*, 1–16.

Fandrey, J., Frede, S., and Jelkmann, W. (1994). Role of hydrogen peroxide in hypoxia-induced erythropoietin production. Biochem. J. *303*, 507–510.

Firth, J.D., Ebert, B.L., Pugh, C.W., and Ratcliffe, P.J. (1994). Oxygen-regulated control elements in the phosphoglycerate kinase 1 and lactate dehydrogenase A genes: similarities with the erythropoietin 3′ enhancer. Proc. Natl. Acad. Sci. USA *91*, 6496–6500.

Firth, J.D., Ebert, B.L., and Ratcliffe, P.J. (1995). Hypoxic regulation of lactate dehydrogenase A. Interaction between hypoxia-inducible factor 1 and cAMP response elements. J. Biol. Chem. *270*, 21021–21027.

Flamme, I., Fröhlich, T., von Reutern, M., Kappel, A., Damert, A., and Risau, W. (1997). HRF, a putative basic helix-loop-helix-PAS domain transcription factor is closely related to hypoxia-inducible factor-1α and developmentally expressed in blood vessels. Mech. Dev. *63*, 51–60.

Forsythe, J.A., Jiang, B.-H., Iyer, N.V., Agani, F., Leung, S.W., Koos, R.D., and Semenza, G.L. (1996). Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. Mol. Cell. Biol. *16*, 4604–4613.

Gassmann, M., Kvietikova, I., Rolfs, A., and Wenger R.H. (1997). Oxygen- and dioxin-regulated gene expression in mouse hepatoma cells. Kidney Int. *51*, 567–574.

Gleadle, J.M., and Ratcliffe, P.J. (1997). Induction of hypoxia-inducible factor-1, erythropoietin, vascular endothelial growth factor, and glucose transporter-1 by hypoxia: evidence against a regulatory role for Src kinase. Blood *89*, 503–509.

Gradin, K., McGuire, J., Wenger, R.H., Kvietikova, I., Whitelaw, M.L., Toftgard, R., Tora, L., Gassmann, M., and Poellinger, L. (1996). Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor. Mol. Cell. Biol. *16*, 5221–5231.

Guillemin, K., and Krasnow, M. A. (1997). The hypoxic response: huffing and HIFing. Cell *89*, 9–12.

Hirose, K., Morita, M., Ema, M., Mimura, J., Hamada, H., Fujii, H., Saijo, Y., Gotoh, O., Sogawa, K., and Fujii-Kuriyama, Y. (1996). cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt). Mol. Cell. Biol. *16*, 1706–1713.

Hogenesch, J. B., Chan, W. K., Jackiw, V. H., Brown, R. C., Gu, Y.-Z., Pray-Grant, M., Perdew, G. H., and Bradfield, C. A. (1997). Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway. J. Biol. Chem. *272*, 8581–8593.

Huang, L. E., Arany, Z., Livingston, D. M., and Bunn, H. F. (1996). Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its α subunit. J. Biol. Chem. *271*, 32253–32259.

Ikeda, E., Achen, M.G., Breier, G., and Risau, W. (1995). Hypoxia-induced transcriptional activation and increased mRNA stability of vascular endothelial growth factor in C6 glioma cells. J. Biol. Chem. *270*, 19761–19766.

Iliopoulos, O., Levy, A.P., Jiang, C., Kaelin Jr., W.G., and Goldberg, M.A. (1996). Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. Proc. Natl. Acad. Sci. USA *93*, 10595–10599.

Isaac, D.D., and Andrew, D.J. (1996). Tubulogenesis in Drosophila: a requirement for the trachealess gene product. Genes Dev. *10*, 103–117.

Jelkmann, W. (1992). Erythropoietin: structure, control of production, and function. Physiol. Rev. *72*, 449–489.

Jiang, B.-H., Semenza, G.L., Bauer, C., and Marti, H.H. (1996a). Hypoxia-inducible factor 1 levels vary exponentially over a physiologically relevant range of O_2 tension. Am. J. Physiol. *271*, C1172–C1180.

Jiang, B.-H., Rue, E., Wang, G.L., Roe, R., and Semenza, G.L. (1996b). Dimerization, DNA binding and transactivating properties of hypoxia-inducible factor 1. J. Biol. Chem. *271*, 17771–17778.

Kallio, P.J., Pongratz, I., Gradin, K., McGuire, J., and Poellinger, L. (1997). Activation of hypoxia-inducible factor 1α: posttranscriptional regulation and conformational change by recruitment of the Arnt transcription factor. Proc. Natl. Acad. Sci. USA *94*, 5667–5672.

Kvietikova, I., Wenger, R.H., Marti, H.H., and Gassmann, M. (1995). The transcription factors ATF-1 and CREB-1 bind constitutively to the hypoxia-inducible factor-1 (HIF-1) DNA recognition site. Nucl. Acids Res. *23*, 4542–4550.

Lee, P. J., Jiang, B.-H., Chin, B. Y., Iyer, N. V., Alam, J., Semenza, G. L., and Choi, A. M. K. (1997). Hypoxia-inducible factor-1 mediates transcriptional activation of the heme oxygenase-1 gene in response to hypoxia. J. Biol. Chem. *272*, 5275–5381.

Levy, A.P., Levy, N.S., Wegner, S., and Goldberg, M.A. (1995). Transcriptional regulation of the rat vascular endothelial growth factor gene by hypoxia. J. Biol. Chem. *270*, 13333–13340.

Levy, A.P., Levy, N.S., and Goldberg, M.A. (1996). Hypoxia-inducible protein binding to vascular endothelial growth factor mRNA and its modulation by the von Hippel-Lindau protein. J. Biol. Chem. *271*, 25492–25497.

Li, H., Ko, H. P., and Whitlock Jr., J. P. (1996). Induction of phosphoglycerate kinase 1 gene expression by hypoxia. Roles of ARNT and HIF1α. J. Biol. Chem. *271*, 21262–21267.

Liu, Y., Cox, S.R., Morita, T., and Kourembanas, S. (1995). Hypoxia regulates vascular endothelial growth factor gene expression in endothelial cells. Circ. Res. *77*, 638–643.

Maltepe, E., Schmidt, J. V., Baunoch, D., Bradfield, C. A., and Simon, M. C. (1997). Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT. Nature *386*, 403–406.

Maxwell, P.H., Pugh, C.W., and Ratcliffe, P.J. (1993). Inducible operation of the erythropoietin 3′ enhancer in multiple cell lines: Evidence for a widespread oxygen-sensing mechanism. Proc. Natl. Acad. Sci. USA *90*, 2423–2427.

Melillo, G., Musso, T., Sica, A., Taylor, L.S., Cox, G.W., and Varesio, L. (1995). A hypoxia-responsive element mediates a

novel pathway of activation of the inducible nitric oxide synthase promoter. J. Exp. Med. *182*, 1683–1693.

Mukhopadhyay, B., Tsiokas, L., Zhou, X.-M., Foster, D., Brugge, J.S., and Sukhatme, V.P. (1995). Hypoxic induction of human vascular endothelial growth factor expression through c-Src activation. Nature *375*, 577–580.

Nagao, M., Ebert, B.L., Ratcliffe, P.J., and Pugh, C.W. (1996). Drosophila melanogaster SL2 cells contain a hypoxically inducible DNA binding complex which recognises mammalian HIF-1 binding sites. FEBS Letters *387*, 161–166.

Nambu, J.R., Chen, W., Hu, S., and Crews, S.T. (1996). The Drosophila melanogaster similar bHLH-PAS gene encodes a protein related to human hypoxia-inducible factor 1 alpha and Drosophila single-minded. Gene *172*, 249–254.

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953–959.

Probst, M.R., Fan, C.-M., Tessier-Lavigne, M., and Hankinson, O. (1997). Two murine homologs of the Drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. J. Biol. Chem. *272*, 4451–4457.

Pugh, C. W., O'Rourke, J. F., Nagao, M., Gleadle, J. M., and Ratcliffe, P. J. (1997). Activation of hypoxia-inducible factor-1; definition of regulatory domains within the α subunit. J. Biol. Chem. *272*, in press.

Rolfs, A., Kvietikova, I., Gassmann, M., and Wenger, R.H. (1997). Oxygen-regulated transferrin expression is mediated by hypoxia-inducible factor-1. J. Biol. Chem., in press.

Salceda, S., Beck, I., and Caro, J. (1996). Absolute requirement of aryl hydrocarbon receptor nuclear translocator protein for gene activation by hypoxia. Arch. Biochem. Biophys. *334*, 389–394.

Schmidt, J.V., and Bradfield, C.A. (1996). Ah receptor signalling pathways. Annu. Rev. Cell Dev. Biol. *12*, 55–89.

Semenza, G.L., Roth, P.H., Fang, H.-M., and Wang, G.L. (1994). Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. J. Biol. Chem. *269*, 23757–23763.

Semenza, G.L., Jiang, B.-H., Leung, S.W., Passantino R., Concorde, J.-P., Maire, P., and Giallongo, A. (1996). Hypoxia response elements in the aldolase A, enolase 1, and lactate dehydrogenase A gene promoters contain essential binding sites for hypoxia-inducible factor 1. J. Biol. Chem. *271*, 32529–32537.

Stein, I., Neeman, M., Shweiki, D, Itin, A., and Keshet, E. (1995). Stabilization of vascular endothelial growth factor mRNA by hypoxia and hypoglycemia and coregulation with other ischemia-induced genes. Mol. Cell. Biol. *15*, 5363–5368.

Tian, H., McKnight, S.L., and Russel, D.W. (1997). Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. Genes Dev. *11*, 72–82.

Wang, G.L., and Semenza, G.L. (1993a). General involvement of hypoxia-inducible factor 1 in transcriptional response to hypoxia. Proc. Natl. Acad. Sci. USA *90*, 4304–4308.

Wang, G.L., and Semenza, G.L. (1993b). Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia. J. Biol. Chem. *268*, 21513–21518.

Wang, G.L., Jiang, B.-H., Rue, E.A., and Semenza, G.L. (1995a). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O_2 tension. Proc. Natl. Acad. Sci. USA *92*, 5510–5514.

Wang, G.L., Jiang, B.-H., and Semenza, G.L. (1995b). Effect of protein kinase and phosphatase inhibitors on expression of hypoxia-inducible factor 1. Biochem. Biophys. Res. Commun. *216*, 669–675.

Wang, G.L., Jiang, B.-H., and Semenza, G.L. (1995c). Effect of altered redox states on expression and DNA-binding activity of hypoxia-inducible factor 1. Biochem. Biophys. Res. Commun. *212*, 550–556.

Wenger, R.H., Rolfs, A., Marti, H.H., Guénet, J.-L., and Gassmann, M. (1996). Nucleotide sequence, chromosomal assignment and mRNA expression of mouse hypoxia-inducible factor 1α. Biochem. Biophys. Res. Commun. *223*, 54–59.

Wenger, R.H., Kvietikova, I., Rolfs, A., Gassmann, M., and Marti, H.H. (1997a). Hypoxia-inducible factor-1α is regulated at the post-mRNA level. Kidney Int. *51*, 560–563.

Wenger, R.H., Rolfs, A., Kvietikova, I., Spielmann, P., Zimmermann, D.R., and Gassmann, M. (1997b). The mouse gene for hypoxia-inducible factor-1α. Genomic organization, expression, and characterization of an alternative first exon and 5′ flanking sequence. Eur. J. Biochem. *246*, 155–165.

Wilk, R., Weizman, I., and Shilo, B.-Z. (1996). Trachealess encodes a bHLH-PAS protein that is an inducer of tracheal cell fates in Drosophila. Genes Dev. *10*, 93–102.

Wood, S.M., Gleadle, J.M., Pugh, C.W., Hankinson, O., and Ratcliffe, P.J. (1996). The role of the aryl hydrocarbon receptor nuclear translocator (ARNT) in hypoxic induction of gene expression. J. Biol. Chem. *271*, 15117–15123.

Zhou, Y.-D., Barnard, M., Tian, H., Li, X., Ring, H. Z., Francke, U., Shelton, J., Richardson, J., Russel, D. W., and McKnight, S. L. (1997). Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. Proc. Natl. Acad. Sci. USA *94*, 713–718.

Biol. Chem., Vol. 378, pp. 617 – 634, July 1997

Review

2-Oxo Acid Dehydrogenase Multienzyme Complexes. The Central Role of the Lipoyl Domain

Axel Berg and Aart de Kok*

Department of Biochemistry, Agricultural University, NL-6703 HA Wageningen, The Netherlands

* Corresponding author

2-Oxo acid dehydrogenase complexes are composed of multiple copies of at least three different enzymes, 2-oxo acid dehydrogenase, dihydrolipoyl acyltransferase and dihydrolipoamide dehydrogenase. The acyltransferase component harbours all properties required for multienzyme catalysis: it forms a large multimeric core, it contains binding sites for the peripheral components, the acyltransferase active site and mobile substrate carrying lipoyl domains that couple the active sites. In the past years these complexes have disclosed many of their secrets, providing currently a wealth of information on macromolecular structure, assembly and symmetry, active-site coupling, conformational mobility, substrate specificity and metabolic regulation. In this review we will discuss developments concerning the structural and mechanistic features of the 2-oxo acid dehydrogenase complexes, with special emphasis on the structure and role of the lipoyl domains in the complex.

Key words: Acyltransferase / Assembly / Catalysis / Lipoyl domain / Mobility.

Introduction

The 2-oxo acid dehydrogenase complexes are commonly regarded as classic examples of multienzyme complexes. Ever since the pioneering work of the group of Lester J. Reed (Reed, 1974), these complexes have expansively disclosed many of their secrets. In particular, by effectively combining protein biochemistry with modern biophysical and genetic techniques, which are undeniable indispensable in contemporary biochemistry, the knowledge of structural and mechanistic properties of these complexes of very large size (0.7– ~14 MDa) has expanded largely during the last decade (Perham, 1991; Guest *et al.*, 1989; Mattevi *et al.*, 1992a; Patel *et al.*, 1996).

Multienzyme complexes are defined as noncovalent aggregates of enzymes that catalyse two or more consecutive steps in a metabolic sequence (Reed, 1974). The purpose of multienzyme complexes appear to be manifold. Reaction intermediates can be forced to complete the intended reaction sequence instead of escaping to conversion by enzymes that compete for the same reaction intermediate (Hammes, 1981). This substrate channelling is particular efficient if intermediates are covalently bound to the complex, like in the cases of the 2-oxo acid dehydrogenase complexes and the fatty acid synthases. Besides, by sequestering reactive intermediates, their conversion by undesired chemical reactions is also prevented (Perham, 1975). Another advantage of multienzyme complexes is that the catalytic activity can be enhanced because the local substrate concentrations are increased significantly (Reed, 1974). Since the enzyme activities in the complex are coupled, also better and more efficient regulation of the overall reaction is possible.

The large attention for the family of 2-oxo acid dehydrogenase multienzyme complexes originates certainly, at least in part, from the key positions they occupy in energy metabolism. The pyruvate dehydrogenase complex (PDHC) catalyses the irreversible oxidative decarboxylation of pyruvate to acetyl-CoA, linking the glycolysis with the tricarboxylic acid cycle. The 2-oxoglutarate dehydrogenase complex (OGDHC) converts 2-oxoglutarate into succinyl-CoA as part of the tricarboxylic acid cycle itself, and the branched-chain 2-oxo acid dehydrogenase complex (BCDHC) catalyses an irreversible step in the catabolism of the branched-chain amino acids by converting the 2-oxo acids derived from valine, leucine and isoleucine.

In this review we will discuss mainly developments concerning the structural and mechanistic features of the 2-oxo acid dehydrogenase complexes, with special emphasis on the structure and role of the lipoyl domains in the complex. Developments on genetic defects and regulation of the eukaryotic complexes have been reviewed elsewhere (Yeaman, 1989; Patel and Roche, 1990; Chuang *et al.*, 1991; Patel *et al.*, 1992; Behal *et al.*, 1993; Patel and Harris, 1995).

Structure and Mechanism

2-Oxo acid dehydrogenase complexes are composed of multiple copies of at least three different enzymes, which are the major catalytic components: a substrate specific 2-oxo acid dehydrogenase (E1), a dihydrolipoyl acyltransferase (E2), and a common lipoamide dehydrogenase (E3). The complex-specific components are commonly abbreviated according to the origin of their complex,

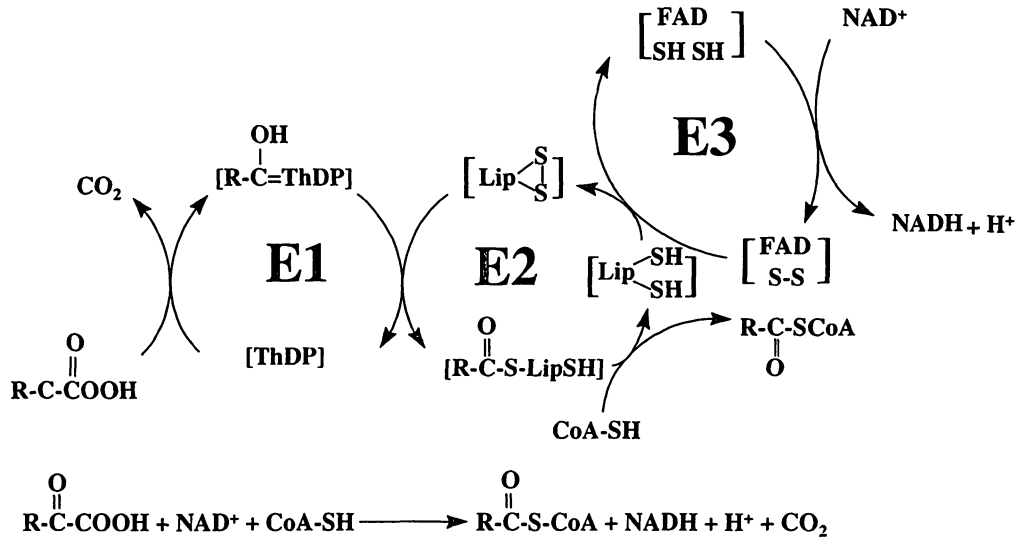


Fig. 1 Reaction Scheme for the Oxidative Decarboxylation of 2-Oxo Acids by the 2-Oxo Acid Dehydrogenase Complexes (Reed, 1974). R = CH_3 for the pyruvate dehydrogenase complex (PDHC), R = CH_2CH_2COOH for the 2-oxoglutarate dehydrogenase complex (OGDHC), and R = $CH(CH_3)_2$, $CH_2CH(CH_3)_2$ or $CH(C_2H_5)(CH_3)$ for the branched-chain 2-oxo acid dehydrogenase complex (BCDHC). ThDP, thiamin diphosphate; Lip, lipoic acid.

where E1p, E1o and E1b, and E2p, E2o and E2b, indicate the E1 and E2 components of PDHC, OGDHC, and BCDHC, respectively. The three components catalyse the set of sequential reactions as shown in Figure 1, involving three prosthetic groups [thiamin diphosphate (ThDP), lipoic acid and flavin adenine dinucleotide (FAD)], and two cofactors (NAD^+ and CoA). The E1 component catalyses the oxidative decarboxylation of the 2-oxo acid and the subsequent reductive acylation of lipoic acid. The lipoic acid is covalently bound in an amide linkage to the N^ϵ group of a specific lysine residue of the E2 component (Nawa *et al.*, 1960), forming the so-called lipoyl group. The acyl group is then transferred to CoA, catalysed by the E2 component. Finally, the reduced lipoyl group is reoxidised by the E3 component with the concomitant reduction of NAD^+.

In addition to the three major catalytic components, mammalian and yeast PDHCs also contain a component called protein X (De Marcucci and Lindsay, 1985), which is involved in the binding of the E3 component to E2 (Neagle and Lindsay, 1991; Lawson *et al.*, 1991a). Furthermore, eukaryotic PDHC and BCDHC contain an E1-specific kinase and phosphatase, which are involved in regulation of the complex activity by a phosphorylation/dephosphorylation mechanism (Linn *et al.*, 1969; Patel and Roche, 1990).

The structural core of all 2-oxo acid dehydrogenase complexes is formed by an aggregate of the E2 component. The E2 components of PDHC from Gram-negative bacteria, and the OGDHC and BCDHC from all sources, except BCDHC from *Bacilli* (Perham and Packman, 1989), are assemblies of 24 identical subunits arranged with octahedral symmetry (Perham, 1991). The PDHC of mammals, yeast and Gram-positive bacteria form a core of 60 E2 subunits with icosahedral symmetry. Multiple copies of the peripheral components E1 and E3 bind tightly but non-covalently to the E2 core. This results in multienzyme complexes of enormous size (~ 5–14 MDa), which can easily be observed as large particles in electron micrographs (Oliver and Reed, 1982). A well-known exception is the PDHC core from *Azotobacter vinelandii*, which dissociates into functional trimers upon binding of the peripheral components to the E2 component (Bosma *et al.*, 1984).

The E1 component (2-oxo acid dehydrogenase) catalyses the rate-limiting step in the overall complex reaction, the reductive acylation of lipoyl groups (Cate *et al.*, 1980). It is the least-characterised enzymatic component of the complex, and no structural information of any E1 component at atomic resolution is yet available. The E1 component exists in two forms, dependent on the type of complex and its symmetry (Perham, 1991). A homodimeric form (α_2) is found in PDHC and OGDHC with octahedral symmetry, and a heterodimeric ($\alpha_2\beta_2$) form is found in all BCDHCs, and in PDHCs with icosahedral cores. Despite a remarkable absence of sequence similarity among the different E1 subunits [e.g. E1p and E1o from *Escherichia coli* (Darlison *et al.*, 1984)], a common structural motif for a ThDP-binding site was proposed in all sequences of E1 components and other ThDP-dependent enzymes (Hawkins *et al.*, 1989). From the recent determination of the three-dimensional structures of a number of different ThDP-dependent enzymes, i.e. transketolase (Lindqvist *et al.*, 1992), pyruvate oxidase (Muller and Schulz, 1993) and pyruvate decarboxylase (Dyda *et al.*, 1993), this motif was shown to be involved in binding the metal ion and the diphosphate group (Lindqvist and Schneider, 1993; Muller *et al.*, 1993). The lack of structural information of E1 components seems partly due to the limited availability of a stable form of the enzyme. There are only a small number of reports on the recombinant expression of functional

heterodimeric E1 (Wynn *et al.*, 1992; Lessard and Perham, 1994; Hester *et al.*, 1995), and recently on the expression of a functional homodimeric E1 (Berg *et al.*, 1996a). However, the work mentioned above seems promising enough to expect new structural information shortly.

In marked contrast to the E1 component, many structural and mechanistic details for the E3 component (lipoamide dehydrogenase) are available. The E3 component is usually common to all 2-oxo acid dehydrogenase complexes from the same source. Exceptions occur e.g. in *Pseudomonas putida*, where three different E3 enzymes have been demonstrated (Palmer *et al.*, 1991), and in *Enterococcus faecalis* where two *lpd* genes have been found, one related to the PDHC (Allen and Perham, 1991) and one of unknown function but probably related to BCDHC (Claiborne *et al.*, 1994). On the other hand, it has been shown that in pea leaf mitochondria the PDHC and the glycine decarboxylase complex share the same lipoamide dehydrogenase (Bourguignon *et al.*, 1996). Lipoamide dehydrogenase belongs to the family of flavin-dependent disulphide oxidoreductases and is a homodimeric enzyme. Four three-dimensional structures of E3 components have been solved by means of X-ray crystallography (Mattevi *et al.*, 1991, 1992b, 1993a; Mande *et al.*, 1996). The enzyme catalyses the regeneration (oxidation) of the dihydrolipoyl group of the E2 component, using a ping-pong mechanism. In the first step the electrons are transferred from the reduced lipoyl group to a reactive disulphide group of the enzyme. In the second step the electrons are transferred via the FAD group to the final electron acceptor NAD^+. Further details regarding structural and mechanistic properties of this enzyme are beyond the scope of this review and can be obtained elsewhere (Massey, 1960; Williams, 1992; De Kok and Van Berkel, 1996).

The Acyltransferase Component

The acyltransferase (E2) components of all 2-oxo acid dehydrogenase complexes are multidomain proteins sharing a common but extraordinary design (Figure 2). Three different types of functional and separately folded domains have been disclosed from limited proteolysis studies (Bleile *et al.*, 1979, 1981; Packman *et al.*, 1984, Chuang, 1985; Hanemaaijer *et al.*, 1987; Packman and Perham, 1987), amino acid sequence comparisons (Russell and Guest, 1991) and functional expression of separate domains by means of genetic engineering (Miles and Guest, 1987b; Dardel *et al.*, 1990; Schulze *et al.*, 1991a, 1992; Hipps and Perham, 1992; Berg *et al.*, 1994; Meng and Chuang, 1994).

The E2 polypeptide chain contains at the N-terminus one to three lipoyl domains (~ 80 amino acid residues), each containing a covalently bound lipoyl group, followed by a peripheral subunit-binding domain (~ 35 amino acid residues) involved in the binding of the E3 and/or E1 components to the E2 core. The C-terminal catalytic domain (~ 29 kDa) accommodates the acyltransferase active site and the intersubunit binding sites responsible for the formation of the multimeric (24 or 60 subunit) core of the complex. The domains are linked by long (15 to 40 amino acid residues) flexible linker segments rich in alanine, proline and charged residues.

All OGDHCs and BCDHCs known so far contain a single lipoyl domain per E2 chain. In PDHCs, E2 components with one, two or three lipoyl domains are found (Zhu and Peterkofsky, 1996). The lipoyl domains of the same E2p chain show a very high amino acid sequence identity. It appears that there is no obvious correlation between the number of lipoyl domains per E2 chain and the source or symmetry of the E2 core (Perham, 1991). The structure,

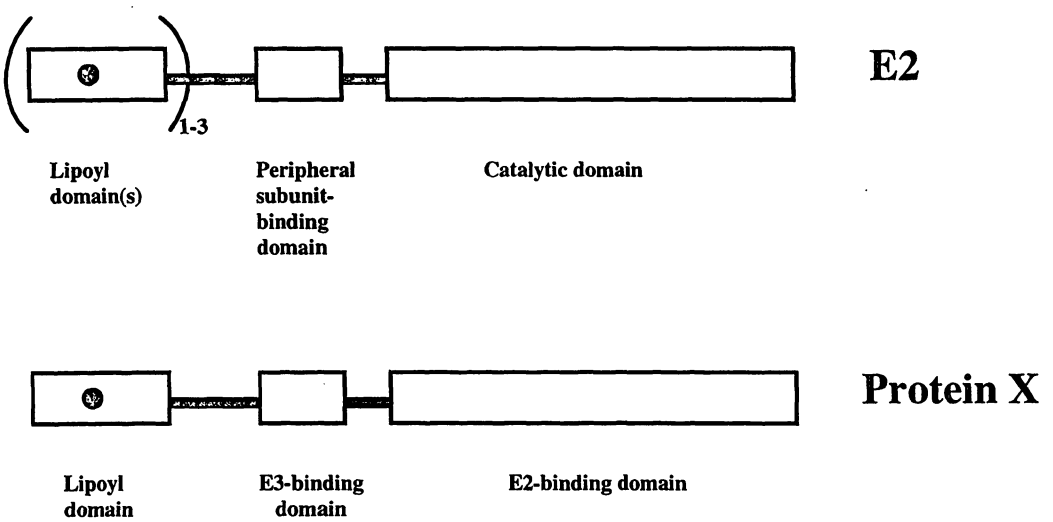


Fig. 2 Schematic Representation of the Structural Domains of the E2 Components and Protein X.
The domains are connected by flexible linker segments (thick line). The approximate position of the lipoylation site in the lipoyl domains is indicated by a dot. PDHCs contain one, two or three lipoyl domains depending on the source, OGDHCs and BCDHCs contain only one lipoyl domain per E2 chain.

function, expression and lipoylation of lipoyl domains are discussed below.

The isolated peripheral subunit-binding domain (~ 35 amino acid residues) is one of the smallest proteins having a stable globular fold without the help of disulphide bridges or prosthetic groups (Brocklehurst *et al.*, 1994). The solution structures of the chemically synthesised binding domains of *E. coli* E2o (Robien *et al.*, 1992) and *B. stearothermophilus* E2p (Kalia *et al.*, 1993), and the crystal structure of the latter binding domain in complex with the E3 component (Mande *et al.*, 1996), have been solved. The global fold of the domain comprises two almost parallel α-helices, separated by a short loop, a short helix and a long, more disordered loop (Figure 3). The binding domain of *A. vinelandii, E. coli* and *B. stearothermophilus* E2p seems to be involved in binding of both the E3 and E1 components (Hanemaaijer *et al.*, 1987; Packman *et al.*, 1988; Schulze *et al.*, 1992), as is the case for all BCDHCs (Wynn *et al.*, 1992). In other complexes with octahedral cores, like OGDHC from *E. coli* (Packman and Perham, 1986), the binding domain is responsible for binding only the E3 component. In eukaryotic PDHCs, the binding domain is involved in binding of the E1 component (Rahmatullah *et al.*, 1989a; Lawson *et al.*, 1991b), whereas the E3 component is bound to the complex by protein X (Rahmatullah *et al.*, 1989a; Neagle and Lindsay, 1991; Lawson *et al.*, 1991a).

The structural core of all 2-oxo acid dehydrogenase complexes is formed by aggregation of the C-terminal catalytic domain of the E2 polypeptide chain. From early electron microscopy studies (Reed, 1974) and crystallographic symmetries observed in X-ray diffraction analysis (DeRosier *et al.*, 1971; Fuller *et al.*, 1979), a cubic or dodecahedral core formed by trimeric E2 units was already proposed (Reed and Hackert, 1990). Although for some time tetrameric E2 building blocks were suggested for the *A. vinelandii* E2p (Bosma *et al.*, 1984; Hanemaaijer *et al.*, 1989), the determination of the crystal structure of the catalytic domain of *A. vinelandii* E2p (Mattevi *et al.*, 1992c, 1993b) clearly showed a hollow 24-meric truncated cube with an edge of 125 Å, with trimers at its vertices (Figure 4). Based on the sequence similarity among E2p amino acid sequences, the same trimeric building blocks are also assumed for complexes with 60-meric dodecahedral cores (Mattevi *et al.*, 1992a, c). This assumption is confirmed by the recently determined crystal structure of the dodecahedral cores from *B. stearothermophilus* and *E. faecalis* (Izard *et al.*, 1997).

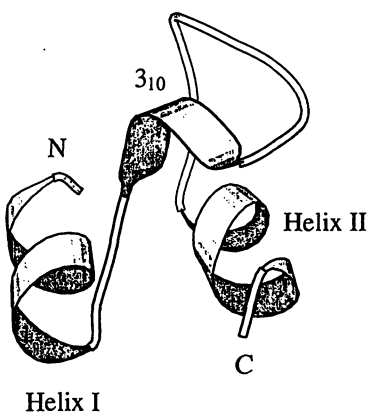


Fig. 3 Schematic Drawing (Kraulis, 1991) of the Peripheral Subunit-Binding Domain of *B. stearothermophilus* E2p (Kalia *et al.*, 1993), Indicating Its Overall Fold and the Elements of Secondary Structure.

The determination of the crystal structure of the catalytic domain revealed a high structural similarity between the trimers and chloramphenicol acetyltransferase (CAT), as predicted earlier on the basis of sequence homology between CAT and *E. coli* E2p (Guest, 1987). The active site is located at each interface of two E2 subunits in a trimer, forming a channel where lipoamide enters from the outside and CoA arrives from the inside of the cube (Figure 4A). Guest (1987) also suggested for the E2p acetyltransferase reaction a similar reaction mechanism to CAT, with a histidine residue acting as a general base and a serine which stabilises the tetrahedral intermediate. Since then, a number of site-directed mutagenesis experiments on different E2 components of various sources have confirmed and refined the proposed reaction mechanism (Griffin and Chuang, 1990; Russell and Guest, 1990; Russell *et al.*, 1992; Meng and Chuang, 1994; Hendle *et al.*, 1995), with the single exception of yeast E2p, where substitution of the proposed active-site histidine by alanine or asparagine did not have a significant effect on the activity (Niu *et al.*, 1990). The involvement of the histidine and serine in the reaction mechanism has been confirmed by the crystal structures of binary and ternary complexes of the *A. vinelandii* catalytic domain with the substrates lipoamide and CoA (Mattevi *et al.*, 1993c) and by structures in which the histidine was replaced by cysteine and the serine by alanine (Hendle *et al.*, 1995).

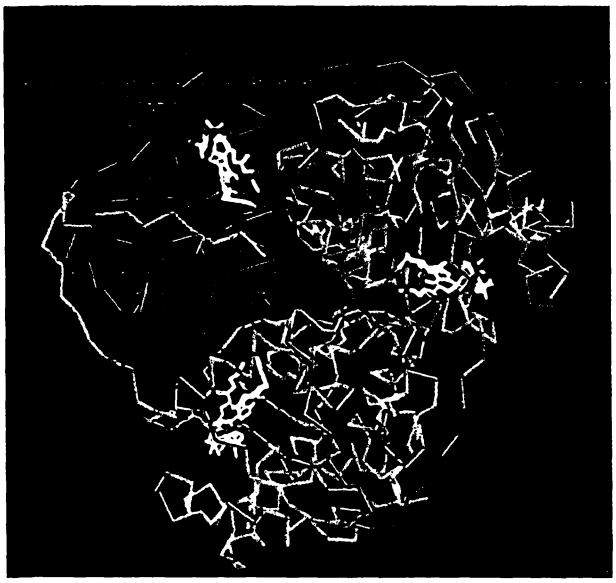
The separate domains in the acyltransferase chain are connected to each other by flexible polypeptide segments of unusual composition with a large majority of alanine, proline, and usually charged amino acid residues. Their structure, mobility and role in active-site coupling will be discussed in the paragraph 'active-site coupling and role of linkers'.

Related Complexes

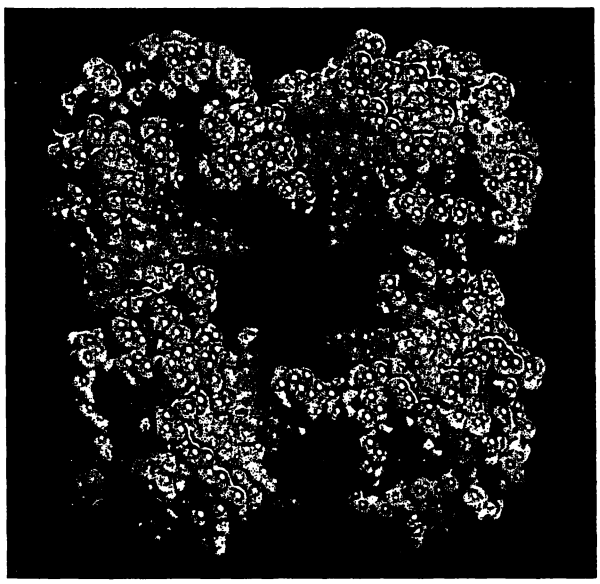
In this intervening paragraph, a brief description of several multienzyme complexes and multicomponent enzymes that show some structural or functional relationship to the 2-oxo acid dehydrogenase complexes is given, aimed at a better understanding of several comparisons that are made with these complexes elsewhere in this review. The glycine decarboxylase complex, the acetoin dehydrogenase enzyme system, the acetyl-coenzyme A carboxylase and the fatty acid synthase are considered.

The glycine decarboxylase complex, also known as the glycine cleavage system, catalyses the reversible oxidative decarboxylation and deamination of glycine yielding

A



B



C

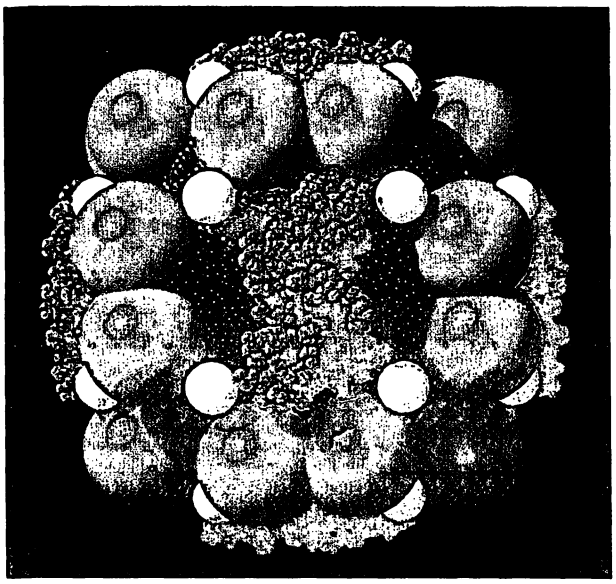


Fig. 4 The Pyruvate Dehydrogenase Complex from *Azotobacter vinelandii*.
(A) backbone trace of the catalytic trimer with bound Coenzyme A (yellow). (B) space filling model (only Cα atoms) of the cubic catalytic core domain consisting of eight trimers, as viewed along the fourfold axis (Mattevi *et al.*, 1992c), (C) space filling model of the octahedral pyruvate dehydrogenase complex as viewed along the fourfold axis. In red and green space filling models (only Cα atoms) of *A. vinelandii* E2p core and E3 are shown, respectively. The structure of E1 is approximated by a sphere (blue) with a radius of 29 Å, while each lipoyl domain is indicated by a sphere (yellow) with a radius of 17 Å. For simplicity the other two lipoyl domains of each E2p chain have been omitted.

carbon dioxide, ammonia, NADH, and 5,10-methylene-5,6,7,8-tetrahydropteroyl-glutamic acid ($CH_2H_4PteGlu_n$) (Douce *et al.*, 1994). The latter component is recycled by serine hydroxymethyltransferase, which is linked to the complex. The glycine decarboxylase complex consists of four protein components, named P-, H-, T- and L-protein. The P-protein, a homodimer containing pyridoxal phosphate, catalyses the decarboxylation of glycine and transfer of the remaining methylamine to the lipoyl group of the H-protein. The T-protein catalyses the transfer of the methylene carbon from the H-protein to $H_4PteGlu_n$ forming $CH_2H_4PteGlu_n$ with the release of ammonia. Finally, the reduced lipoyl group of the H-protein is reoxidised by the L-protein, a lipoamide dehydrogenase. The H-protein couples the activities of the multienzyme complex enzymes with its covalently bound lipoyl group, analogous to the lipoyl domains of 2-oxo acid dehydrogenase complexes. The structure of the H-protein has shown to be similar to the lipoyl domain structures (see below).

The acetoin dehydrogenase enzyme system catalyses the cleavage of acetoin to acetyl-CoA and acetaldehyde, comparable to the oxidative decarboxylation of 2-oxo acids by the 2-oxo acid dehydrogenase complexes (Opperman *et al.*, 1989). This multienzyme complex also comprises multiple copies of three enzymes, a ThDP-dependent acetoin dehydrogenase (E1), a dihydrolipoamide acetyltransferase (E2), and lipoamide dehydrogenase (E3). The nucleotide sequence-derived amino acid sequences of the acetoin dehydrogenase complex components of the Gram-negative bacteria *Pelobacter carbinolicus* (Opperman and Steinbüchel, 1994) and *Alcaligenes euthrophus* (Pfiefert *et al.*, 1991), and the Gram-positive bacterium

Clostridium magnum (Krüger *et al.*, 1994) show a high similarity with each other and with the components of the 2-oxo acid dehydrogenase complexes. The E1 components of acetoin dehydrogenase enzyme systems are heterodimers, the E2 components have one or two lipoyl domains, and the E3 component was found to have an N-terminal lipoyl domain in *C. magnum*. No structural information on any of these components is yet available.

Acetyl-CoA carboxylase catalyses the first committed step in long-chain fatty acid biosynthesis, converting acetyl-CoA to malonyl-CoA. It belongs to the family of biotin-dependent carboxylases, which include also propionyl-CoA carboxylase, oxaloacetate decarboxylase, pyruvate decarboxylase and transcarboxylase (Toh *et al.*, 1993). They are composed of three subunits: a biotin carboxylase, a biotin carboxyl carrier protein, and a carboxyltransferase. The biotin carboxylase catalyses the ATP-dependent carboxylation of biotin using bicarbonate. Biotin is linked to the N^{ϵ} group of a specific lysine residue of the biotin carboxyl carrier protein. It acts as a swinging arm to transfer the carboxyl group to the carboxyltransferase, much like the lipoyl domains of the 2-oxo acid dehydrogenase complexes. In acetyl-CoA carboxylase, the carboxyl group is finally transferred to acetyl-CoA to form malonyl-CoA, catalysed by the carboxyltransferase subunit. The three-dimensional structures of the *E. coli* biotin carboxylase and the biotin carboxyl carrier protein of acetyl-CoA carboxylase have been determined by X-ray crystallography (Waldrop *et al.*, 1994; Athappilly and Hendrickson, 1995). It shows that the structures of the biotin carboxyl carrier protein and lipoyl domains are strikingly similar (see below).

The fatty acid synthase system is considered here, because a swinging arm is involved in the catalytic mechanism (Hammes, 1981). The enzymes of fatty acid synthesis constitute a multienzyme complex, in which all reaction intermediates are bound to an acyl carrier protein. The acyl carrier protein contains a phosphopantetheine moiety covalently linked to the hydroxyl group of a serine residue. During catalysis, the acyl groups are bound via a thioester linkage to the -SH group of this prosthetic group, which serves as a swinging arm to deliver the acyl group to the different active sites. With the presence of a swinging arm the similarity with 2-oxo acid dehydrogenase complexes ends. There is no structural or sequential homology between the acyl carrier protein and lipoyl domains.

Subunit Assembly and Quaternary Structure

Despite the fact that tertiary structures of individual components and domains (Mattevi *et al.*, 1992a) have provided a wealth of new information and insights in many aspects of 2-oxo acid dehydrogenase complexes, a quaternary structure of the complex is of course not obtained by simply adding up these tertiary structures. But the combination of these data with other biochemical data and electron microscopy studies could lead to an increased insight into subunit assembly of these complexes and have led to models such as pictured in Figure 4C.

As discussed above, the quaternary structure of the E2 component is based on an assembly of 24 or 60 subunits in octahedral and icosahedral complexes, respectively. Multiple copies of the E1 and E3 components bind to the E2 cores, but not in stoichiometric amounts. In *E. coli* PDHC, 12 E1p dimers bind to the edges of the octahedral core, while 6 E3 dimers bind onto the 6 faces of the cube (Koike *et al.*, 1963). The optimal E1o:E2o:E3 chain-stoichiometry for *E. coli* and *A. vinelandii* OGDHC have been estimated at 12:24:12 (Pettit *et al.*, 1973; Bosma, 1984). An unusual exception to the organisation of PDHCs from Gram-negative bacteria is the PDHC from *A. vinelandii*. As isolated, it consists of a trimeric E2 core to which two E1 dimers and a single E3 dimer are bound (Schulze *et al.*, 1992). With that it is the smallest 2-oxo acid dehydrogenase complex known ($M_r \sim 700000$). Upon removal of the peripheral components the E2 component aggregates to the common 24-meric cubic core with 432 symmetry. However, in the presence of CoA or acetyl-CoA, the cubic core does not dissociate upon addition of the peripheral components. Therefore we cannot exclude the possibility that *in vivo* this PDHC can exist as a 24-meric E2p core (Schulze *et al.*, 1993).

The icosahedral cores of mammalian, Gram-positive bacterial and yeast PDHC bind about 30 E1 tetramers ($\alpha_2\beta_2$) at the edges and 6 E3 dimers at the faces of the pentagonal dodecahedral E2 core (Henderson *et al.*, 1979; Wu and Reed, 1984). The E1p components are bound to the peripheral subunit-binding domain of the E2p component via their β-subunits (Wynn *et al.*, 1992; Lessard and Perham, 1995). In addition, mammalian and yeast PDHC contain 6 or 12 protein X subunits (Jilka *et al.*, 1986; Maeng *et al.*, 1994; Sanderson *et al.*, 1996), and one to three copies of a specific kinase and phosphatase (Reed, 1974; Reed and Hackert, 1990). Protein X has a similar domain structure as the E2 components (Figure 2), with one N-terminal lipoyl domain, a small E3-binding domain, and a C-terminal domain that binds protein X to the E2 core (Behal *et al.*, 1989; Rahmatullah *et al.*, 1989b; Lawson *et al.*, 1991a). The main function of protein X seems to be the binding of the E3 dimers to the complex (Neagle and Lindsay, 1991; Lawson *et al.*, 1991a; Maeng *et al.*, 1994), although it has been suggested that the lipoyl domain of protein X can also play a role in the catalytic mechanism (Rahmatullah *et al.*, 1990; Lawson *et al.*, 1991b; Sanderson *et al.*, 1996). The E1 kinases were shown to associate to the mammalian PDHC via binding to, preferentially, the inner lipoyl domain of the E2 component (Radke *et al.*, 1993; Liu *et al.*, 1995a), which also appears to play a role in kinase stimulation (Ono *et al.*, 1993; Ravindran *et al.*, 1996). Finally, mammalian OGDHC lacks protein X, but the E2o component also lacks a sequence motif of the putative peripheral-subunit binding domain (Nakano *et al.*, 1991). However, in the N-terminal region of the E1o component a sequence similar to protein X is found to be involved in E3 binding (Rice *et al.*, 1992).

Many 2-oxo acid dehydrogenase complexes have shown to be self-assembling (Koike *et al.*, 1963; Reed *et al.*, 1975; Bates *et al.*, 1977; Bosma *et al.*, 1984), at least *in vitro*, which means that they can be functionally reconstituted from their individual components. Interestingly, reconstitution experiments have revealed that in cases where the binding domain is involved in binding of both peripheral components, there seems to be competition for binding sites during assembly, possibly caused by steric hindrance (Reed *et al.*, 1975). For example, the prokaryotic E2p components from *A. vinelandii* and *E. coli* can bind one peripheral component in a chain ratio of 1:2 for E2p:E1 or E2p:E3 in the absence of the other peripheral component (Reed *et al.*, 1975; Bosma *et al.*, 1984). Addition of the other component causes displacement of the bound component. For the *B. stearothermophilus* E2p it has been shown that the E1p and E3 components cannot bind simultaneously to the same isolated binding domain (Lessard and Perham, 1995). This suggests that one peripheral subunit-binding domain is involved in the binding of a dimer of E1 (β subunit) or E3. Other studies on the binding of E3 to an isolated di-domain (lipoyl domain plus binding domain, Hipps *et al.*, 1994) or the intact E2p or E2o (Westphal *et al.*, 1995) confirmed this suggestion. Furthermore, it has been shown that a dimeric E3 is essential for binding to the E2 component (Schulze *et al.*, 1991b).

An exciting observation by Westphal and co-workers was that binding to the *A. vinelandii* E2p stabilised the E3 by tightening the intersubunit interaction, making this component less sensitive to over-reduction (Westphal *et al.*, 1995). From this study it was also concluded that the E2p or E2o component must bind to the E3 component at its subunit interface near the dyad axis, thereby preventing sterically the binding of a second E2. This mode of interaction between E3 and the binding domain has recently been justified by the elucidation of the X-ray crystal structure of the *B. stearothermophilus* E3 component with the binding domain of E2p (Mande *et al.*, 1996). The E3 component was shown to bind mainly to the N-terminal part of the binding domain, via predominantly electrostatic interactions with both E3 subunits.

The interaction between the E1 and the E2 component has not been studied in such large detail. Reconstitution experiments of PDHCs based on chimeric E2p components from *A. vinelandii* and *E. coli* have shown that E1p interacts with both the binding domain and the catalytic domain of E2p (Schulze *et al.*, 1992). Site-directed mutagenesis experiments of *A. vinelandii* E2p already had indicated that the binding sites for E1p are located on the C-terminal part of the binding domain and the N-terminal part of the catalytic domain (Schulze *et al.*, 1991c).

The spatial distribution of the complex components has been probed mainly by different electron microscopy techniques. From electron microscopy studies on negatively stained *E. coli* complexes (Oliver and Reed, 1982), and confirmed by cryoelectron microscopy of the frozen-hydrated complexes (Wagenknecht *et al.*, 1990), it has been shown that the E1 and E3 subunits are separated from the E2 core by a gap of 3–5 nm. These results have been interpreted by a flexible mode of attachment of E1 and E3 to the E2 core, conferred by the linker sequence connecting the peripheral subunit-binding domain with the core-forming catalytic domain. Similar results have been obtained for the mammalian PDHC, where the E1p and E3 components also appear to not be bound directly to the E2p-X core (Wagenknecht *et al.*, 1991). Whether the suggested mobility of the peripheral components plays a role in active-site coupling remains to be determined. It could well be that a function as a spacer rather than being flexible is more important for this linker segment. It should be noted that scanning transmission electron microscopy (STEM) studies of cross-linked *E. coli* PDHC suggest more distinct E1 and E3 binding sites on the E2 core (CaJacob *et al.*, 1985a; Yang *et al.*, 1986). Finally, various electron microscopy studies of *E. coli* PDHC, e.g. of negatively stained E2p cores (Bleile *et al.*, 1979), using cryoelectron microscopy (Wagenknecht *et al.*, 1990, 1992), or using STEM of gold cluster labelled lipoyl groups (Yang *et al.*, 1994), all indicate that the lipoyl domains extend from the surface of the E2p core. This has been confirmed by cross-linking studies of *E. coli* PDHC with avidin, which tightly binds lipoyl groups (Hale *et al.*, 1992).

Structure and Role of Lipoyl Domains

The lipoyl groups, which are covalently attached to lipoyl domains, are essential for the coupling of the activities of the separate multienzyme components, by acting as reaction intermediate carriers. As such, they are substrates for the three different active sites in the multienzyme complex, and are indispensable for the efficient functioning of the complex. Lipoyl domains are independently folded and functioning protein units, as has been shown by reductive acylation of lipoyl domains obtained by limited proteolysis (Bleile *et al.*, 1981; Packman *et al.*, 1984) or by expression of sub-genes encoding them (Ali and Guest, 1990; Dardel *et al.*, 1990; Quinn *et al.*, 1993; Berg *et al.*, 1994; 1995; Liu *et al.*, 1995b). In combination with amino acid sequence comparisons of many acyltransferases (Russell and Guest, 1991; Matuda *et al.*, 1992; Dardel *et al.*, 1993; Zhu and Peterkofsky, 1996), it was concluded that lipoyl domains comprise approximately 80 residues, each containing one fully conserved lysine residue as potential lipoylation site.

A very intriguing question regarding lipoyl domains is why a number of PDHCs have more than one (two or three) lipoyl domain per E2 chain? It has been shown for the *E. coli* PDHC, containing three lipoyl domains per E2p chain, that nearly half of the lipoyl domains can be removed by limited proteolysis without significant loss in overall complex activity (Berman *et al.*, 1981; Stepp *et al.*, 1981). Likewise, the rate of chemical modification of enzymatic excision of lipoyl groups was shown to be faster than the rate at which complex activity decreased (Ambrose-Griffin *et al.*, 1980; Berman *et al.*, 1981; Danson *et al.*, 1981;

Stepp *et al.*, 1981). Furthermore, by genetic engineering two of the three *E. coli* PDHC lipoyl domains can be removed with no apparent effect on overall complex activity or active-site coupling (Guest *et al.*, 1985; Graham *et al.*, 1986). This is explained by an active-site coupling mechanism in which the rate-limiting E1 component can serve many lipoyl domains, and in which acyl groups can rapidly transfer among different lipoyl domains (Bates *et al.*, 1977; Collins and Reed, 1977; Danson *et al.*, 1978a, b). In this way the function of the removed or inactivated lipoyl domains can be taken over by the remainders. However, this mechanism does still not explain the apparent excess of lipoyl domains in a number of PDHCs. Only recently, an alternative approach of comparing isogenic strains of *E. coli* containing PDHCs with one, two or three lipoyl domains per E2p chain, showed that the maximum growth rates of these strains in minimal medium are directly correlated with the number of lipoyl domains (Dave *et al.*, 1995). These results show at least the advantage for *E. coli* having PDHC with three lipoyl domains per E2p chain for efficient balanced growth on carbon sources that need this complex in their metabolic route.

Lipoyl domains are not exclusively found at the N-terminal part of the E2 chains of 2-oxo acid dehydrogenase complexes and acetoin dehydrogenase enzyme systems. As mentioned earlier, the protein X component of eukaryotic PDHCs also contains an N-terminal lipoyl domain, and it has been shown that this lipoyl domain is able to function in the overall complex reaction (Rahmatullah *et al.*, 1990; Lawson *et al.*, 1991b). The lipoyl domain of protein X of *S. cerevisiae* shows about 50% amino acid sequence identity to the lipoyl domains of its E2p (Behal *et al.*, 1989). Only recently, the PDHCs from *Alcaligenes eutrophus* (Hein and Steinbüchel, 1994) and *Neisseria meningitidis* (De la Sierra *et al.*, 1994, Ala'Aldeen *et al.*, 1996), and the acetoin dehydrogenase enzyme system from *Clostridium magnum* (Krüger *et al.*, 1994), were found to have an E3 component containing a lipoyl domain connected to its N-terminus by a linker segment. Although the role of these lipoyl domains in the multienzyme complex has not yet been established, their high amino acid sequence identity with the lipoyl domains of the E2 component suggests that they could take part in the overall reaction.

An additional role for the lipoyl domains of mammalian PDHCs has found to be involvement in binding of the pyruvate dehydrogenase kinase. It was shown that the kinase selectively binds to the inner lipoyl domain of the two lipoyl domains of the mammalian E2 component (Liu *et al.*, 1995a), and that this association involves the hydrophobic inner portion of the lipoyl group (Radke *et al.*, 1993). The kinase activity is regulated through the redox state of the inner lipoyl domain, showing an increased activity upon reduction or acetylation of the lipoyl group (Ravindran *et al.*, 1996). To explain the rapid phosphorylation of many E1 components by a limited number of kinase molecules, a mechanism has been proposed in which the bound kinase directly moves between the different inner lipoyl domains without dissociating from the complex (Ono *et al.*, 1993; Liu *et al.*, 1995a).

Recently, several three-dimensional structures of lipoyl domains have been solved, all by means of NMR spectroscopy. These are the single lipoyl domain of *B. stearothermophilus* PDHC (Dardel *et al.*, 1993), a non-native hybrid lipoyl domain of *E. coli* PDHC (Green *et al.*, 1995a) and the N-terminal lipoyl domains of *A. vinelandii* OGDHC and PDHC (Berg *et al.*, 1996b, 1997). All lipoyl domain structures show a very similar overall fold, which is now considered a new class of all-β folds called β-barrel-sandwich hybrids (Chothia and Murzin, 1993), or flattened β-barrels (Green *et al.*, 1995a). The structure of the lipoyl domain is formed by two very similar four-stranded antiparallel β-sheets, which are packed around a core of hydrophobic residues in a sandwich-like manner (Figure 5). The lipoylation site is exposed in a β-turn at the far end of one of the sheets, while the N-terminus and C-terminus meet at the opposite side of the domain, in two adjacent β-strands in the other β-sheet. The lipoyl domain displays a remarkable internal symmetry, relating the two halves of the molecule by a two-fold rotational axis. Regarding the lipoyl domain structures determined so far, in combination with the conservation of key-residues in all lipoyl domain amino acid sequences, it is concluded that all lipoyl domains will have highly similar folds (Dardel *et al.*, 1993; Green *et al.*, 1995a; Berg *et al.*, 1996b, 1997).

Another noticeable feature of lipoyl domains is that their structure is not altered by lipoylation. In the NMR spectra

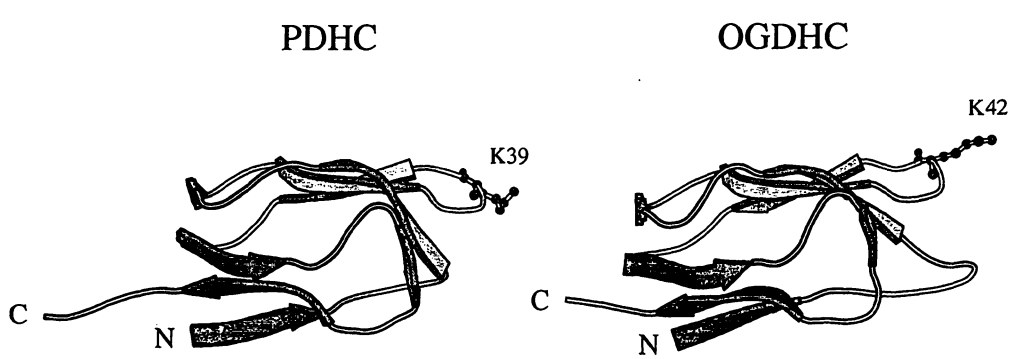


Fig. 5. Schematic Drawing (Kraulis, 1991) of the Lipoyl Domain of *A. vinelandii* PDHC (Berg *et al.*, 1997) and OGDHC (Berg *et al.*, 1996b). Lys39 and Lys42 are the lipoylation sites.

only very small differences in chemical shifts of residues close to the lipoylation site are observed between the unlipoylated and lipoylated forms of the lipoyl domain (Dardel *et al.*, 1991; Berg *et al.*, 1994). These chemical shift differences are considered too small to suggest binding of the hydrophobic lipoyl group back to the protein surface of the lipoyl domain, and indicate that the lipoyl groups of the 2-oxo acid dehydrogenase complexes are likely to swing freely.

The structure of the lipoyl domain is not only required for the specific attachment of the lipoic acid prosthetic group (Wallis and Perham, 1994), but also increases the efficiency of reductive acylation of its lipoyl group dramatically. While free lipoamide or lipoic acid are good substrates for the E2 and E3 components, lipoamide is an extremely poor substrate for the E1 component (Reed *et al.*, 1958a). Likewise, reductive acetylation by E1p of a lipoylated decapeptide, with an amino acid sequence matching that surrounding the lipoylation site of *E. coli* E2p, is barely detectable (Graham *et al.*, 1989). However, lipoyl groups, when bound to the lipoyl domains, are readily and efficiently reductively acylated by their appropriate E1 components (Bleile *et al.*, 1981; Packman *et al.*, 1984, Berg *et al.*, 1994, 1995). It has been suggested, on the basis of the large difference between the K_m (~ 33 μM) and the K_s (> 0.3 mM), that the enlarged efficiency of reductive acylation of the lipoyl group by a folded lipoyl domain is not directly a matter of enhanced binding to the E1 component (Graham and Perham, 1990).

The lipoyl domain is also responsible, at least in part, for the specificity of the reductive acylation reaction. Lipoyl domains are only efficiently reductively acylated by the E1 component of their parent complex, as has been shown for the *E. coli* (Graham *et al.*, 1989) complexes, and subsequently for the *A. vinelandii* complexes (Berg *et al.*, 1996a). Reduced overall activity of reconstituted *E. coli* PDHC containing *A. vinelandii* E1p (De Kok and Westphal, 1985), and of *E. coli* and *A. vinelandii* PDHCs containing each others engineered lipoyl domains (Schulze *et al.*, 1992), is also ascribed to reduced efficiency of reductive acylation. Together this indicates that molecular recognition occurs between lipoyl domains and E1 components. The conclusion that all lipoyl domains, and in particular the *A. vinelandii* PDHC and OGDHC lipoyl domains for which their specificity in the reductive acylation reactions has been demonstrated, have very similar structures, indicates that this molecular recognition is a result of delicate differences among these domains.

On the basis of a careful comparison of lipoyl domain structures and sequences, potential residues that could be important for molecular recognition have been proposed (Dardel *et al.*, 1993; Green *et al.*, 1995a; Berg *et al.*, 1994, 1996b, 1997). These include residues of an exposed loop connecting the first two β-strands in the sequence (Figure 5), and which lie close in space to the lipoylation site. Site-directed mutagenesis experiments of the exposed loop of the *A. vinelandii* OGDHC lipoyl domain combined with cross acylation experiments indicate that this loop is very likely involved in the interaction with the E1 component. Recently a similar conclusion was drawn for the *B. stearothermophilus* PDHC lipoyl domain by a combination of site-directed mutagenesis and NMR experiments (Wallis *et al.*, 1996). It is obvious, however, that a complete picture of the specific molecular interactions involved in recognition of lipoyl domains is impaired by the lack of structural information of the E1 component at atomic resolution.

Lipoyl domains derived from 2-oxo acid dehydrogenase complexes show high structural homology to the H-protein of the glycine decarboxylase system and the biotinyl domain of acetyl-CoA carboxylase, as had been predicted on the basis of (low) sequence similarity to lipoyl domains (Brocklehurst and Perham, 1993; Toh *et al.*, 1993). The structure of the H-protein has been solved by X-ray crystallography (Pares *et al.*, 1994; Cohen-Addad *et al.*, 1995), and consists of a β-barrel-sandwich structure similar to the lipoyl domain (Figure 6). The lipoyl-lysine is analogously presented in a β-hairpin turn and is rather flexible, as concluded from the relatively high B-factors. The N-terminal exposed loop of the lipoyl domain is replaced by a helix in the H-protein, which is also in proximity of the lipoylation site. The H-protein (~ 130 amino acid residues) is larger than a lipoyl domain (~ 80 amino acid residues) and contains two additional β-strands at the N-terminal end, and a short and a long C-terminal helix. Very interestingly, the X-ray crystal structure of the methylamine loaded form of the H-protein shows that the lipoyl-methylamine group interacts with several specific conserved residues, located in a cleft formed by the β-sandwich and the N-terminal helix (Cohen-Addad *et al.*, 1995). The strong interactions between the protein and the methylamine group explain why the methylamine-loaded form of the H-protein is stable (Neuberger *et al.*, 1991), and show that in this form the lipoyl group is not free to rotate.

The X-ray crystal structure of the biotinyl domain of acetyl-CoA carboxylase from *E. coli* is very similar to the structure of lipoyl domains (Athappilly and Hendrickson,

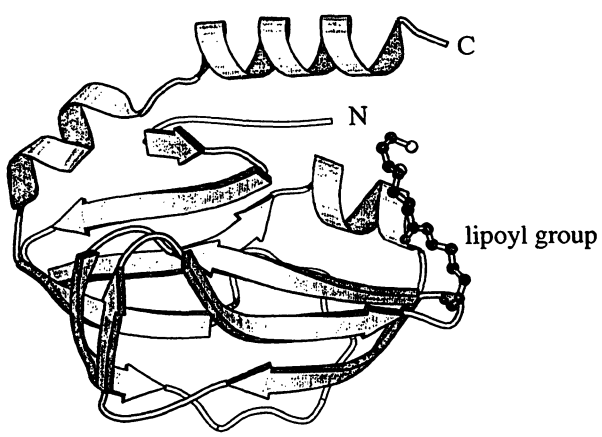


Fig. 6 Schematic Drawing (Kraulis, 1991) of the H-Protein of the Glycine Decarboxylase Complex from Pea Leaves (Pares *et al.*, 1994).
The lipoyl group is represented as a ball-and-stick model.

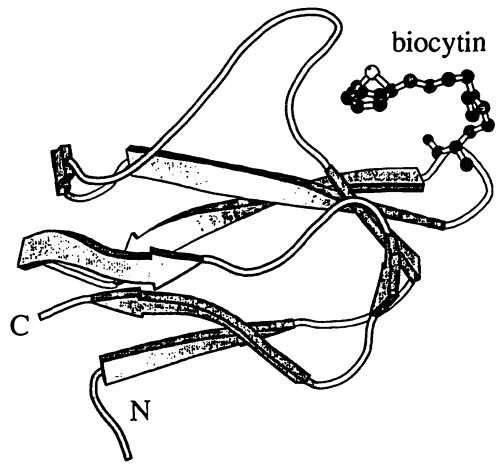


Fig. 7 Schematic Drawing (Kraulis, 1991) of the Biotinyl Domain of Acetyl-CoA Carboxylase from *E. coli* (Athappilly and Hendrickson, 1995).
The biocytin is shown in ball-and-stick representation.

1995). The only difference in overall fold is the absence of the exposed N-terminal loop of the lipoyl domain in the biotinyl domain, which in turn possesses a large loop between the second and third β-strand (Figure 7). The biotinylated lysine residue resides in a β-hairpin turn, a structural feature which seems conserved in all proteins containing lipoic acid or biotin. The biotinyl group is well defined in the electron-density map, and interacts with residues of the large loop that is absent in lipoyl domains. This indicates that the biotinyl group is not completely free to swing, at least in its noncarboxylated form, but is partially buried below the surface of the domain.

Lipoylation

Lipoic acid (6,8-thioctic acid or 1,2-dithiolane-3-pentanoic acid) is the prosthetic group of the lipoyl domains of 2-oxo acid dehydrogenase complexes and acetoin dehydrogenase enzyme systems, and of the H-protein of the glycine decarboxylase complex. The carboxyl group of lipoic acid is bound in an amide linkage to the N^{ϵ} group of a specific lysine residue in a posttranslational modification process called lipoylation. Until recently, very little was known about the enzyme(s) and mechanism(s) involved in lipoylation of apo-lipoyl domains, and about the biosynthesis of lipoic acid, which will not be discussed here. Early work by the group of Lester J. Reed on the lipoylating systems of *E. coli* and *S. faecalis* showed that incorporation of lipoic acid into apo-PDHC required lipoic acid, ATP, inorganic phosphate and a divalent metal ion, and that lipoic acid and ATP could be replaced by lipoyladenylate (Reed *et al.*, 1958a, 1958b). Only after the observation of unlipoylated and mismodified (octanoylated) lipoyl domains after overexpression of their subgenes in *E. coli* under certain conditions, the attention to the lipoylation process seemed renewed.

Unlipoylated and octanoylated lipoyl domains have been observed in a number of cases. Expression of an *E. coli* PDHC lipoyl domain in *E. coli* produced a mixture of the lipoylated and unlipoylated form of the domain (Miles and Guest, 1987b; Ali and Guest, 1990). Likewise, lipoylated and unlipoylated forms were detected when the human PDHC inner lipoyl domain (Quinn *et al.*, 1993), the lipoyl domains of *A. vinelandii* PDHC and OGDHC (Berg *et al.*, 1994, 1995), and the H-protein from pea (Macherel *et al.*, 1996) were overproduced in *E. coli*. Addition of exogenous lipoic acid to the growth medium commonly resulted in increased amounts of the lipoylated form, indicating that the overexpression exceeded the cell's capacity for lipoylation. In addition to the lipoylated and unlipoylated forms, small amounts of an octanoylated form were observed when the lipoyl domains of *B. stearothermophilus* and human PDHC, and the H-protein of bovine glycine decarboxylase complex were expressed in *E. coli* (Dardel *et al.*, 1990; Fujiwara *et al.*, 1992; Hipps and Perham, 1992; Liu *et al.*, 1995b), and when *E. coli* PDHC lipoyl domains were expressed in a lipoate-deficient *E. coli* strain (Ali *et al.*, 1990). The identification of the different forms of lipoyl domains and H-proteins was greatly facilitated by the application of new and advanced mass-spectrometry techniques, e.g. electrospray mass spectrometry.

It is interesting that lipoyl domains and H-proteins from different sources are being lipoylated by the lipoylating system of *E. coli*, with the exception of the lipoyl domain of bovine BCDHC which is not lipoylated (Griffin *et al.*, 1990). The lipoyl domain and its particular lipoyl-lysine residue are selected specifically for lipoylation, and the question arises how this site is recognised by the lipoylating enzyme(s)? This question was addressed by site-directed mutagenesis experiments of residues around the lipoyl-lysine residue of the *B. stearothermophilus* PDHC lipoyl domain (Wallis and Perham, 1994). They showed that the position of the lipoyl-lysine residue in the β-turn where it is found is essential for lipoylation, rather than the residues directly surrounding it.

Lipoate protein ligases, responsible for the lipoylation reaction, have been isolated from *E. coli* (Brookfield *et al.*, 1991; Green *et al.*, 1995b) and from bovine liver mitochondria (Fujiwara *et al.*, 1994). In *E. coli* two distinct genes have been cloned (Morris *et al.*, 1994, 1995), the products of which are involved in lipoylation. The *lplA* gene product is responsible for the incorporation of exogenous lipoic acid, via a mechanism using a lipoyl-AMP intermediate, which is consistent with the early observations by Reed (Reed *et al.*, 1958a, 1958b), and analogous to the biotin protein ligase (Cronan, 1989). The other lipoate protein ligase, the *lipB* gene product, utilises lipoyl groups generated via endogenous lipoic acid biosynthesis. This indicates that two redundant pathways with two different lipoate protein ligases for lipoylation exist in *E. coli*. Two isoforms of lipoyltransferase were purified from bovine liver mitochondria that could use lipoyl-AMP but not lipoic acid plus MgATP for lipoylation (Fujiwara *et al.*, 1994), sug-

gesting that two enzymes are involved in the complete lipoylation reaction.

From the amino acid sequences of E2 components the number of potential lipoylation sites can be determined. The extent of lipoylation of the potential sites has been a subject of controversy for a long time. Various studies using different methods of determination, mainly on the E2p component of *E. coli*, resulted in numbers ranging from 1.7 to 2.0 lipoyl groups per E2p chain (Packman *et al.*, 1984). However, re-assessment of the number of functional lipoyl groups per E2p and E2o chain by a combination of protein-chemical and modern mass-spectrometric techniques clearly showed that all potential lipoylation sites contain functional lipoyl groups (Packman *et al.*, 1991).

Active-Site Coupling and Role of Linkers

The coupling of the activities of the three enzymatic components of the 2-oxo acid dehydrogenase complexes is brought about by their lipoyl groups. A swinging arm mechanism, in which the long and flexible lipoyl groups rotate among the different active sites, had been proposed responsible for active-site coupling (Koike *et al.*, 1963). However, by fluorescence energy transfer measurements it was shown that the distances between the active sites in PDHC were at least 4 nm (Shepherd and Hammes, 1977), a gap that cannot be narrowed by a single rotating lipoyl group (~ 2.8 nm). Thus, more than just a swinging arm is required for active-site coupling, and the participation of two or more lipoyl groups in a catalytic cycle, or additional movement of protein parts, was suggested to be involved in the mechanism.

The mechanism of active-site coupling is more complicated than the simple direct transfer of an intermediate from an active site via the lipoyl group to the subsequent active site. The observation that only a few E1 dimers can reductively acylate many lipoyl groups in *E. coli* PDHC and OGDHC (so-called servicing experiments) (Bates *et al.*, 1977; Collins and Reed, 1977; Danson *et al.*, 1978a), and at a rate comparable to the overall complex activity (Danson *et al.*, 1978b), suggests that acyl-transfer reactions among lipoyl groups need to be accounted for in the active-site coupling mechanism. The involvement of intramolecular transacylation reactions can be used to explain the observations that the removal of lipoyl domains by limited proteolysis (Berman *et al.*, 1981; Stepp *et al.*, 1981), genetic engineering (Guest *et al.*, 1985), or enzymatic release or chemical inactivation of the lipoyl groups (Ambrose-Griffin *et al.*, 1980; Berman *et al.*, 1981; Danson *et al.*, 1981; Stepp *et al.*, 1981) proceed faster than the accompanying loss of overall complex activity. The kinetics of inactivation have been simulated by a computer model, using a multiple random coupling mechanism for active-site coupling (Hackert *et al.*, 1983a, b). This mechanism was supported by site-directed mutagenesis experiments of *E. coli* PDHC, which showed that various permutations of functional and non-functional (by lipoyl-lysine to glutamine mutations) lipoyl domains did not effect active-site coupling (Allen *et al.*, 1989). These elegant experiments strongly suggest that the three lipoyl domains function completely independent and that the reductive acetylation of them is random.

The multiple random coupling mechanism suggests a rapid transfer of acyl groups among lipoyl groups. From the X-ray crystal structures of the E2 and E3 components it is clear that a direct enzyme catalysed transfer of acyl-groups and reducing equivalents between lipoyl domains cannot take place. However, the E2 and E3 components can catalyse exchange reactions. In the absence of CoA and NADH, such as under conditions used in servicing experiments, the only enzymatic component that might participate in an exchange reaction is the E1 component. An exchange reaction between hydroxyethylThDP and acetyllipoate has not been reported. Rather, acetylThDP is formed from acetyllipoate in a slow reaction and subsequently hydrolyzed (CaJacob *et al.*, 1985b). A non-catalyzed chemical reaction seems unlikely as free lipoamide cannot function as an acylgroup acceptor (unpublished observation). Furthermore it is expected that a mixture of 6-S-acyldihydrolipoate and 8-S-acyldihydrolipoate will be formed. Since the 6-S-acyldihydrolipoate is not converted by the E2 component, a rapid decrease of the complex activity during servicing experiments could be expected, but is, however, not observed. Chemical intramolecular isomerisation of the inactive enzyme-bound 6-S-acyldihydrolipoate to 8S-acyldihydrolipoate is too slow during normal turn-over (Yang and Frey, 1986) to compensate for this. Finally, the negative charge on the lipoyl domains may prevent any direct interaction between the domains. In several recent textbooks this feature of random coupling is treated incorrectly (Voet and Voet, 1995, Mathews and van Holde, 1996).

The observation of unexpectedly sharp resonances in the ^{1}H-NMR spectrum of the *E. coli* PDHC (~ 5 MDa) indicated the presence of conformational mobile regions in E2p (Perham *et al.*, 1981). Later on, similar signals have also been detected in other complexes, like *E. coli* OGDHC (Perham and Roberts, 1981), *B. stearothermophilus* PDHC (Duckworth *et al.*, 1982), and *A. vinelandii* PDHC, OGDHC and E2p (Hanemaaijer *et al.*, 1988b). These signals were ascribed to the alanine and proline rich linker segments in the E2 chain, which could be important for facilitating movement of lipoyl domains for active-site coupling. This was supported by the similarity of ^{1}H-NMR spectra of a 32-residue synthetic peptide representing the amino acid sequence of a linker (Radford *et al.*, 1986), and by the reduction of sharp signals in the spectrum of *E. coli* PDHC with two deleted lipoyl domains and linkers (Radford *et al.*, 1987). Direct evidence for mobility of the linker segments came from the observation of sharp ^{1}H-NMR signals assigned to a histidine residue that was introduced in the interdomain segment of a mutant *E. coli* PDHC with only one lipoyl domain (Texter *et al.*, 1988).

A more detailed analysis, by means of NMR spectroscopy and circular dichroism, of several synthetic peptides

with amino acid sequences representing *E. coli* PDHC linkers, showed that these peptides are very flexible in solution (Radford *et al.*, 1989; Green *et al.*, 1992). Moreover, it was shown that their structures were disordered but not random coil, and in particular that all Ala-Pro peptide bonds were in the *trans* conformation. This suggests a certain stiffening of the flexible linkers, which would facilitate protrusion from the core of the complex.

The importance of the flexibility of the linkers for active-site coupling was shown by introducintg deletions in the 32-residue linker segment of an *E. coli* PDHC mutant containing one lipoyl domain (Miles *et al.*, 1988). Reduction of the linker to less than 13 residues caused significant loss in active-site coupling. Studies in which different linkers, varying in length and composition, were engineered in this complex confirmed the importance of the linker size (Turner *et al.*, 1993). They also showed that the amino acid composition of the linker is of less importance, although highly charged linkers were not allowed. Finally, the flexibility and importance of the short linker segment connecting the peripheral subunit-binding domain with the catalytic domain for active-site coupling needs to be established unequivocally (Schulze *et al.*, 1993), and may vary between different 2-oxo acid dehydrogenase complexes (Perham, 1991).

Acknowledgement

This work was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for Scientific Research (NWO).

References

Ala'Aldeen, D.A.A., Westphal, A.H., Weston, V., Atta, M., Baldwin, T.J., Bartley, J., de Kok, A., and Borriello, S.P. (1996). Cloning, sequencing, characterisation and implications on vaccine design of the novel dihydrolipoyl acetyltransferase of *Neisseria meningitidis*. J. Med. Microbiol. *45*, 1–14.

Ali, S.T., and Guest, J.R. (1990). Isolation and characterization of lipoylated and unlipoylated domains of the E2p subunit of the pyruvate dehydrogenase complex of *Escherichia coli*. Biochem. J. *271*, 139–145.

Ali, S.T., Moir, A.J.G., Ashton, P.R., Engel, P.C., and Guest, J.R. (1990). Octanoylation of the lipoyl domains of the pyruvate dehydrogenase complex in a lipoyl-deficient strain of *Escherichia coli*. Mol. Microbiol. *4*, 943–950.

Allen, A.G., Perham, R.N., Allison, N., Miles, J.S., and Guest, J.R. (1989). Reductive acetylation of tandemly repeated lipoyl domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli* is random order. J. Mol. Biol. *208*, 623–633.

Allen, A.G., and Perham, R.N. (1991). Two lipoyl domains in the dihydrolipoamide acetyltransferase chain of the pyruvate dehydrogenase multienzyme complex of *Streptococcus faecalis*. FEBS Lett. *287*, 206–210.

Ambrose-Griffin, M.C., Danson, M.J., Griffin, W.G., Hale, G., and Perham, R.N. (1980). Kinetic analysis of the role of lipoic acid residues in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Biochem. J. *187*, 393–401.

Athappilly, F.K., and Hendrickson, W.A. (1995). Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. Structure *3*, 1407–1419.

Bates, D.L., Danson, M.J., Hale, G., Hooper, E.A., and Perham, R.N. (1977). Self-assembly and catalytic activity of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Nature *268*, 313–316.

Behal, R.H., Browning, K.S., Hall, T.B., and Reed, L.J. (1989). Cloning and nucleotide sequence of the gene for protein X from *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA *86*, 8732–8736.

Behal, R.H., Buxton, D.B., Robertson, J.G., and Olson, M.S. (1993). Regulation of the pyruvate dehydrogenase multienzyme complex. Annu. Rev. Nutr. *13*, 497–520.

Berg, A., de Kok, A., and Vervoort, J. (1994). Sequential ^{1}H and ^{15}N nuclear magnetic resonance assignments and secondary structure of the N-terminal lipoyl domain of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Eur. J. Biochem. *221*, 87–100.

Berg, A., Smits, O., de Kok, A., and Vervoort, J. (1995). Sequential ^{1}H and ^{15}N nuclear magnetic resonance assignments and secondary structure of the lipoyl domain of the 2-oxoglutarate dehydrogenase complex from *Azotobacter vinelandii*. Evidence for high structural similarity with the lipoyl domain of the pyruvate dehydrogenase complex. Eur. J. Biochem. *234*, 148–159.

Berg, A., Bosma, H., Westphal, A.H., Hengeveld, J., Vervoort, J., and de Kok, A. (1996a). The pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Structure and function of the lipoyl domain. In: Biochemistry and Physiology of Thiamin Diphosphate Enzymes, H. Bisswanger and A. Schellenberger, eds. (Prien, Germany: A. u. C. Intemann, Wissenschaflicher Verlag), pp. 278–291.

Berg, A., Vervoort, J., and de Kok, A. (1996b). Solution structure of the lipoyl domain of the 2-oxoglutarate dehydrogenase complex from *Azotobacter vinelandii*. J. Mol. Biol. *261*, 432–442.

Berg, A., Vervoort, J., and de Kok, A. (1997). Three-dimensional structure in solution of the N-terminal lipoyl domain of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Eur. J. Biochem., in press.

Berman, J.N., Chen, G.-X., Hale, G., and Perham, R.N. (1981). Lipoic acid residues in a take-over mechanism for the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Biochem. J. *199*, 513–520.

Bleile, D.M., Munk, P., Oliver, R.M., and Reed, L.J. (1979). Subunit structure of dihydrolipoyl transacetylase component of pyruvate dehydrogenase complex from *Escherichia coli*. Proc. Natl. Acad. Sci. USA *76*, 4385–4389.

Bleile, D.M., Hackert, M.L., Pettit, F.H., and Reed, L.J. (1981). Subunit structure of dihydrolipoyl transacetylase component of pyruvate dehydrogenase complex from bovine heart. J. Biol. Chem. *256*, 514–519.

Bosma, H.J. (1984). Studies on 2-oxoacid dehydrogenase multienzyme complexes of *Azotobacter vinelandii*. (Ph.D. Thesis, Agricultural University Wageningen, The Netherlands).

Bosma, H.J., de Kok, A., Westphal, A.H., and Veeger, C. (1984). The composition of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Does a unifying model exist for the complexes from gram-negative bacteria? Eur. J. Biochem. *142*, 541–549.

Bourguignon, J., Merand, V., Rawsthorne, S., Forest, E., and Douce, R. (1996). Glycine decarboxylase and pyruvate dehydrogenase complexes share the same dihydrolipoamide dehydrogenase in pea leaf mitochondria: evidence from mass spectrometry and primary-structure analysis. Biochem. J. *313*, 229–234.

Brocklehurst, S.M., and Perham, R.N. (1993). Prediction of the three-dimensional structures of the biotinylated domain from yeast pyruvate carboxylase and of the lipoylated H-protein from the pea leaf glycine cleavage system: A new automated method for the prediction of protein tertiary structure. Protein Sci. *2*, 626–639.

Brocklehurst, S.M., Kalia, Y.N., and Perham, R.N. (1994). Protein-protein recognition mediated by a mini-protein domain: possible evolutionary significance. TIBS *19*, 360–361.

Brookfield, D.E., Green, J., Ali, S.T., Machado, R.S., and Guest, J.R. (1991). Evidence for two protein-lipoylation activities in *Escherichia coli*. FEBS Lett. *295*, 13–16.

CaJacob, C.A., Frey, P.A., Hainfeld, J.F., Wall, J.S., and Yang, H. (1985a). *Escherichia coli* pyruvate dehydrogenase complex: particle masses of the complex and component enzymes measured by scanning transmission electron microscopy. Biochemistry *24*, 2425–2431.

CaJacob, C.A., Gavino, G.R., and Frey, P.A. (1985b). Pyruvate dehydrogenase complex of *Escherichia coli*. Thiamin pyrophosphate and NADH-dependent hydrolysis of acetyl-CoA. J. Biol. Chem. *260*, 14610–14615.

Cate, R.L., Roche, T.E., and Davis, L.C. (1980). Rapid intersite transfer of acetyl groups and movement of pyruvate dehydrogenase component in the kidney pyruvate dehydrogenase complex. J. Biol. Chem. *255*, 7556–7562.

Chothia, C., and Murzin, A.G. (1993). New folds for all-β proteins. Structure *1*, 217–222.

Chuang, D.T., Hu, C.C., Ku, L.S., Markovitz, P.J., and Cox, R.P. (1985). Subunit structure of the dihydrolipoyl transacylase component of branched-chain α-keto acid dehydrogenase complex from bovine liver. Characterization of the inner transacylase core. J. Biol. Chem. *260*, 13779–13786.

Chuang, D.T., Fisher, C.W., Lau, K.S., Griffin, T.A., Wynn, R.M., and Cox, R.P. (1991). Maple syrup urine disease: domain structure, mutations and exon skipping in the dihydrolipoyl transacylase (E2) component of the branched-chain α-keto acid dehydrogenase complex. Mol. Biol. Med. *8*, 49–63.

Claiborne, A., Ross, R.P., Ward, D., Parsonage, D., and Crane III, E.J. (1994). Flavoprotein peroxide and disulfide reductases and their roles in Streptococcal oxidative metabolism. In: Flavins and Flavoproteins 1993, K. Yagi, ed. (Berlin: Walter de Gruyter), pp. 587–596.

Cohen-Addad, C., Pares, S., Sieker, L., Neuburger, M., and Douce, R. (1995). The lipoamide arm in the glycine decarboxylase complex is not freely swinging. Nat. Struct. Biol. *2*, 63–68.

Collins, J.H., and Reed, L.J. (1977). Acyl group and electron pair relay system: a network of interacting lipoyl moieties in the pyruvate and α-ketoglutarate dehydrogenase complexes from *Escherichia coli*. Proc. Natl. Acad. Sci. USA *74*, 4223–4227.

Cronan, J.E. (1989). The *E. coli bio* operon: transcriptional repression by an essential protein modification enzyme. Cell *58*, 427–429.

Danson, M.J., Hooper, E.A., and Perham, R.N. (1978a). Intramolecular coupling of active sites in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Biochem. J. *175*, 193–198.

Danson, M.J., Fersht, A.R., and Perham, R.N. (1978b). Rapid intramolecular coupling of active sites in the pyruvate dehydrogenase complex of *Escherichia coli*: Mechanism for rate enhancement in a multimeric structure. Proc. Natl. Acad. Sci. USA *75*, 5386–5390.

Danson, M.J., Hale, G., and Perham, R.N. (1981). The role of lipoic acid residues in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Biochem. J. *199*, 505–511.

Dardel, F., Packman, L.C., and Perham, R.N. (1990). Expression in *Escherichia coli* of a sub-gene encoding the lipoyl domain of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*. FEBS Lett. *264*, 206–210.

Dardel, F., Laue, E.D., and Perham, R.N. (1991). Sequence-specific ^{1}H-NMR assignments and secondary structure of the lipoyl domain of the *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex. Eur. J. Biochem. *201*, 203–209.

Dardel, F., Davis, A.L., Laue, E.D., and Perham, R.N. (1993). Three-dimensional structure of the lipoyl domain from *Bacillus stearothermophilus* pyruvate dehydrogenase multienzy complex. J. Mol. Biol. *229*, 1037–1048.

Darlison, M.G., Spencer, M.E., and Guest, J.R. (1984). Nucleotide sequence of the *sucA* gene encoding 2-oxoglutarate dehydrogenase of *Escherichia coli* K12. Eur. J. Biochem. *141*, 351–359.

Dave, E., Guest, J.R., and Attwood, M.M. (1995). Metabolic engineering in *Escherichia coli*: lowering the lipoyl domain content of the pyruvate dehydrogenase complex adversely affects the growth rate and yield. Microbiology *141*, 1839–1849.

De Kok, A., and Westphal, A.H. (1985). Hybrid pyruvate dehydrogenase complexes reconstituted from components of the complexes from *Escherichia coli* and *Azotobacter vinelandii*. Eur. J. Biochem. *152*, 35–41.

De Kok, A., and van Berkel, W.J.H. (1996). Lipoamide dehydrogenase. In: Alpha-Keto Acid Dehydrogenase Complexes, M.S. Patel, T.E. Roche and R.A. Harris, eds. (Basel: Birkhäuser Verlag), pp. 53–70.

De la Sierra, I.L., Prangé, J.T., Fourme, R., Padrón, G., Fuentes, P., Musacchio, A., and Madrazo, J. (1994). Crystallization and preliminary X-ray investigation of a recombinant outer membrane protein from *Neisseria meningitidis*, J. Mol. Biol. *235*, 1154–1155.

De Marcucci, O., and Lindsay, J.G. (1985). Component X. An immunologically distinct polypeptide associated with mammalian pyruvate dehydrogenase multi-enzyme complex. Eur. J. Biochem. *149*, 641–648.

DeRosier, D.J., Oliver, R.M., and Reed, L.J. (1971). Crystallization and preliminary structural analysis of dihydrolipoyl transsuccinylase, the core of the 2-oxoglutarate dehydrogenase complex. Proc. Natl. Acad. Sci. USA *68*, 1135–1137.

Douce, R., Bourguignon, J., Macherel, D., and Neuburger, M. (1994). The glycine decarboxylase system in higher plant mitochondria: structure, function and biogenesis. Biochem. Soc. Trans. *22*, 184–188.

Duckworth, H.W., Jaenicke, R., Perham, R.N., Wilkie, A.O.M., Finch, J.T., and Roberts, G.C.K. (1982). Limited proteolysis and proton NMR spectroscopy of *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex. Eur. J. Biochem. *124*, 63–69.

Dyda, F., Furey, W., Subramanyam, S., Sax, M., Farrenkopf, B., and Jordan, F. (1993). Catalytic centers in the thiamin diphosphate dependent enzyme pyruvate decarboxylase at 2.4-C resolution. Biochemstry *32*, 6165–6170.

Fujiwara, K., Okamura-Ikeda, K., and Motokawa, Y. (1992). Expression of mature bovine H-protein of the glycine cleavage system in *Escherichia coli* and *in vitro* lipoylation of the apoform. J. Biol. Chem. *267*, 20011–20016.

Fujiwara, K., Okamura-Ikeda, K., and Motokawa, Y. (1994). Purification and characterization of lipoyl-AMP:Nε-lysine lipoyltransferase from bovine liver mitochondria. J. Biol. Chem. *269*, 16605–16609.

Fuller, C.C., Reed, L.J., Oliver, R.M., and Hackert, M.L. (1979). Crystallization of a dihydrolipoyl transacetylase-dihydrolipoyl

dehydrogenase subcomplex and its implications regarding the subunit structure of the pyruvate dehydrogenase complex from *Escherichia coli*. Biochem. Biophys. Res. Commun. *90*, 431–438.

Graham, L.D., Guest, J.R., Lewis, H.M., Miles, J.S., Packman, L.C., Perham, R.N., Radford, F.R.S., and Radford, S.E. (1986). The pyruvate dehydrogenase multi-enzyme complex of *Escherichia coli*: genetic reconstruction and functional analysis of the lipoyl domains. Phil. Trans. R. Soc. Lond. A, *317*, 391–404.

Graham, L.D., Packman, L.C., and Perham, R.N. (1989). Kinetics and specificity of reductive acylation of lipoyl domains from 2-oxo acid dehydrogenase multienzyme complexes. Biochemistry *28*, 1574–1581.

Graham, L.D., and Perham, R.N. (1990). Interactions of lipoyl domains with the E1p subunits of the pyruvate dehydrogenase multienzyme complex from *Escherichia coli*. FEBS Lett. *262*, 241–244.

Green, J.D.F., Perham, R.N., Ullrich, S.J., and Appella, E. (1992). Conformational studies of the interdomain linker peptides in the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Biol. Chem. *267*, 23484–23488.

Green, J.D.F., Laue, E.D., Perham, R.N., Ali, S.T., and Guest, J.R. (1995a). Three-dimensional structure of a lipoyl domain from dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Mol. Biol. *248*, 328–343.

Green, D.E., Morris, T.W., Green, J., Cronan, J.E., and Guest, J.R. (1995b). Purification and properties of the lipoate protein ligase of *Escherichia coli*. Biochem. J. *309*, 853–862.

Griffin, T.A., and Chuang, D.T. (1990). Genetic reconstruction and characterization of the recombinant transacylase (E2b) component of bovine branched-chain α-keto acid dehydrogenase complex. Implication of histidine 391 as an active site residue. J. Biol. Chem. *265*, 13174–13180.

Griffin, T.A., Wynn, R.M., and Chuang, D.T. (1990). Expression and assembly of mature apotransacylase (E2b) of bovine branched-chain α-keto acid dehydrogenase complex in *Escherichia coli*. J. Biol. Chem. *265*, 12104–12110.

Guest, J.R., Lewis, H.M., Graham, L.D., Packman, L.C., and Perham, R.N. (1985). Genetic reconstruction and functional analysis of the repeating lipoyl domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Mol. Biol. *185*, 743–754.

Guest, J.R. (1987). Functional implications of structural homologies between chloramphenicol acetyltransferase and dihydrolipoamide acetyltransferase. FEMS Microbiol. Lett. *44*, 417–422.

Guest, J.R., Angier, S.J., and Russel, G.C. (1989). Structure, expression, and protein engineering of the pyruvate dehydrogenase complex of *Escherichia coli*. In: Alpha-keto Acid Dehydrogenase Complexes. Organization, Regulation, and Biomedical Ramifications, T.E. Roche and M.S. Patel, eds. (New York: The New York Academy of Sciences) Vol. 573, pp. 76–99.

Hackert, M.L., Oliver, R.M., and Reed, L.J. (1983a). Evidence for a multiple random coupling mechanism in the α-ketoglutarate dehydrogenase multienzyme complex of *Escherichia coli*: a computer model analysis. Proc. Natl. Acad. Sci. USA *80*, 2226–2230.

Hackert, M.L., Oliver, R.M., and Reed, L.J. (1983b). A computer model analysis of the active-site coupling mechanism in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Proc. Natl. Acad. Sci. USA *80*, 2907–2911.

Hale, G., Wallis, N.G., and Perham, R.N. (1992). Interaction of avidin with the lipoyl domains in the pyruvate dehydrogenase multienzyme complex: three-dimensional location and similarity to biotinyl domains in carboxylases. Proc. R. Soc. Lond. B *248*, 247–253.

Hammes, G.G. (1981). Processing of intermediates in multienzyme complexes. Biochem. Soc. Symp. *46*, 73–90.

Hanemaaijer, R., de Kok, A., Jollès, J., and Veeger, C. (1987). The domain structure of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Eur. J. Biochem. *169*, 245–252.

Hanemaaijer, R., Vervoort, J., Westphal, A.H., de Kok, A., and Veeger, C. (1988b). Mobile sequences in the pyruvate dehydrogenase complex, the E2 component, the catalytic domain and the 2-oxoglutarate dehydrogenase complex of *Azotobacter vinelandii*, as detected by 600 MHz proton NMR spectroscopy. FEBS Lett. *240*, 205–210.

Hanemaaijer, R., Westphal, A.H., Van der Heiden, T., De Kok, A., and Veeger, C. (1989). The quaternary structure of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. A reconsideration. Eur. J. Biochem. *179*, 287–292.

Hawkins, C.F., Borges, A., and Perham, R.N. (1989). A common structural motif in thiamin pyrophosphate-binding enzymes. FEBS Lett. *255*, 77–82.

Hein, S., and Steinbüchel, A. (1994). Biochemical and molecular characterization of the *Alcaligenes eutrophus* pyruvate dehydrogenase complex and identification of a new type of dihydrolipoamide dehydrogenase. J. Bacteriol. *176*, 4394–4408.

Henderson, C.E., Perham, R.N., and Finch, J.T. (1979). Structure and symmetry of *Bacillus stearothermophilus* pyruvate dehydrogenase multienzyme complex and implications for eukaryote evolution. Cell *17*, 85–93.

Hendle, J., Mattevi, A., Westphal, A.H., Spee, J., de Kok, A., Teplyakov, A., and Hol, W.G.J. (1995). Crystallographic and enzymatic investigations on the role of Ser558, His610, and Asn614 in the catalytic mechanism of *Azotobacter vinelandii* dihydrolipoamide acetyltransferase (E2p). Biochemistry *34*, 4287–4298.

Hester, K., Luo, J., Burns, G., Braswell, E.H., and Sokatch, J.R. (1995). Purification of active E1 α(2)β(2) of *Pseudomonas putida* branched-chain 2-oxo acid dehydrogenase complex. Eur. J. Biochem. *233*, 828–836.

Hipps, D.S., and Perham, R.N. (1992). Expression in *Escherichia coli* of a sub-gene encoding the lipoyl and peripheral subunit-binding domains of the dihydrolipoamide acetyltransferase component of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*. Biochem. J. *283*, 665–671.

Hipps, D.S., Packman, L.C., Allen, M.D., Fuller, C., Sakaguchi, K., Appella, E., and Perham, R.N. (1994). The peripheral subunit-binding domain of the dihydrolipoyl acetyltransferase component of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus*: preparation and characterization of its binding to the dihydrolipoyl dehydrogenase component. Biochem. J. *297*, 137–143.

Izard, T., Sarfaty, S., Westphal, A., de Kok, A., and Hol, W.G.J. (1997). Improvement of diffraction quality upon rehydration of dehydrated icosahedrol *Enterococcus faecalis* pyruvate dehydrogenase core crystals. Protein Science *6*, 913–915.

Jilka, J. M., Rahmatullah, M., Kazemi, M. and Roche, T. E. (1986). Properties of a newly characterised protein of the bovine kidney pyruvate dehydrogenase complex. J. Biol. Chem. *261*, 1858–1867.

Kalia, Y.N., Brocklehurst, S.M., Hipps, D.S., Appella, E., Sakaguchi, K. and Perham, R. N. (1993). The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase

multienzyme complex of *Bacillus stearothermophilus*. J. Mol. Biol. *230*, 323–341.

Koike, M., Reed, L. J., and Carroll, W. R. (1963). α-Keto acid dehydrogenase complexes. IV. Resolution and reconstitution of the *Escherichia coli* pyruvate dehydrogenase complex. J. Biol. Chem. *238*, 30–39.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Krüger, N., Opperman, F.B., Lorenzl, H., and Steinbüchel, A. (1994). Biochemical and molecular characterization of the *Clostridium magnum* acetoin dehydrogenase enzyme system. J. Bacteriol. *176*, 3614–3630.

Lawson, J.E., Behal, R.H., and Reed, L.J. (1991a). Disruption and mutagenesis of the *Saccharomyces cerevisiae* PDX1 gene encoding the protein X component of the pyruvate dehydrogenase complex. Biochemistry *30*, 2834–2839.

Lawson, J.E., Niu, X., and Reed, L.J. (1991b). Functional analysis of the domains of dihydrolipoamide acetyltransferase from *Saccharomyces cerevisiae*. Biochemistry *30*, 11249–11254.

Lessard, I.A.D., and Perham, R.N. (1994). Expression in *Escherichia coli* of genes encoding the E1α and E1β subunits of the pyruvate dehydrogenase complex of *Bacillus stearothermophilus* and assembly of a functional E1 component (α2β2) *in vitro*. J. Biol. Chem. *269*, 10378–10383.

Lessard, I.A.D., and Perham, R.N. (1995). Interaction of component enzymes with the peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*: stoichiometry and specificity in self-assembly. Biochem. J. *306*, 727–733.

Lindqvist, Y., Schneider, G., Ermler, U., and Sundström, M. (1992). Three-dimensional structure of transketolase, a thiamin diphosphate dependent enzyme, at 2.5 Å resolution. EMBO J. *11*, 2373–2379.

Lindqvist, Y., and Schneider, G. (1993). Thiamin diphosphate dependent enzymes: transketolase, pyruvate oxidase and pyruvate decarboxylase. Curr. Opin. Struct. Biol. *3*, 896–901.

Linn, T.C., Pettit, F.H., and Reed, L.J. (1969). α-keto acid dehydrogenase complexes: X. regulation of the activity of pyruvate dehydrogenase complex from beef kidney mitochondria by phosphorylation and dephosphorylation. Proc. Natl. Acad. Sci. USA *62*, 234–241.

Liu, S., Baker, J.C., and Roche, T.E. (1995a). Binding of the pyruvate dehydrogenase kinase to recombinant constructs containing the inner lipoyl domain of the dihydrolipoyl acetyltransferase component. J. Biol. Chem. *270*, 793–800.

Liu, S., Baker, J.C., Andrews, P.C., and Roche, T.E. (1995b). Recombinant expression and evaluation of the lipoyl domains of the dihydrolipoyl acetyltransferase component of the human pyruvate dehydrogenase complex. Arch. Biochem. Biophys. *316*, 926–940.

Macherel, D., Bourguignon, J., Forest, E., Faure, M., Cohen-Addad, C., and Douce, R. (1996). Expression, lipoylation and structure determination of recombinant pea H-protein in *Escherichia coli*. Eur. J. Biochem. *236*, 27–33.

Maeng, C.-Y., Yazdi, M.A., Niu, X.-D., Lee, H.Y., and Reed, L.J. (1994). Expression, purification, and characterization of the dihydrolipoamide dehydrogenase-binding protein of the pyruvate dehydrogenase complex from *Saccharomyces cerevisiae*. Biochemistry *33*, 13801–13807.

Mande, S.S., Sarfaty, S., Allen, M.D., Perham, R.N., and Hol, W.G. J. (1996). Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase. Structure *4*, 277–286.

Massey, V., Gibson, Q.H., and Veeger, C. (1960). Intermediates in the catalytic action of lipoyl dehydrogenase. Biochem. J. *77*, 341–351.

Mathews, C.K., and Van Holde, K.E. (1996). Biochemistry. (Menlo Park, USA: Benjamin/Cummings)

Mattevi, A., Schierbeek, A.J., and Hol, W.G.J. (1991). Refined crystal structure of lipoamide dehydrogenase from *Azotobacter vinelandii* at 2.2 Å resolution. A comparison with the structure of glutathione reductase. J. Mol. Biol. *220*, 975–994.

Mattevi, A., de Kok, A., and Perham, R.N. (1992a). The pyruvate dehydrogenase multienzyme complex. Curr. Opin. Struct. Biol. *2*, 877–887.

Mattevi, A., Obmolova, G., Kalk, K.H., Sokatch, J., Betzel, C.H., and Hol, W.G.J. (1992b). The refined crystal structure of *Pseudomonas putida* lipoamide dehydrogenase complexed with NAD^+ at 2.45 Å resolution. Proteins *13*, 336–351.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K.H., Westphal, A. H., de Kok, A., and Hol, W.G.J. (1992c). Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. Science *255*, 1544–1550.

Mattevi, A., Obmolova, G., Kalk, K.H., van Berkel, W.J.H., and Hol, W.G.J. (1993a). 3-Dimensional structure of lipoamide dehydrogenase from *Pseudomonas fluorescens* at 2.8 Ångstrom resolution – analysis of redox and thermostability properties. J. Mol. Biol. *230*, 1200–1215.

Mattevi, A., Obmolova, G., Kalk, K.H., Westphal, A.H., de Kok, A., and Hol, W.G.J. (1993b). Refined crystal structure of the catalytic domain of dihydrolipoyl transacetylase (E2p) from *Azotobacter vinelandii* at 2.6 Ångstrom resolution. J. Mol. Biol. *230*, 1183–1199.

Mattevi, A., Obmolova, G., Kalk, K.H., Teplyakov, A., and Hol, W. G.J. (1993c). Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p). Biochemistry *32*, 3887–3901.

Matuda, S., Nakano, K., Ohta, S., Shimura, M., Yamanaka, T., Nakagawa, S., Titani, K., and Miyata, T. (1992). Molecular cloning of dihydrolipoamide acetyltransferase of the rat pyruvate dehydrogenase complex: sequence comparison and evolutionary relationship to other dihydrolipoamide acyltransferases. Biochim. Biophys. Acta *1131*, 114–118.

Meng, M., and Chuang, D.T. (1994). Site-directed mutagenesis and functional analysis of the active-site residues of the E2 component of bovine branched-chain α-keto acid dehydrogenase complex. Biochemistry *33*, 12879–12885.

Miles, J.S., and Guest, J.R. (1987b). Subgenes expressing single lipoyl domains of the pyruvate dehydrogenase complex of *Escherichia coli*. Biochem. J. *245*, 869–874.

Miles, J.S., Guest, J.R., Radford, S.E., and Perham, R.N. (1988). Investigation of the mechanism of active site coupling in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli* by protein engineering. J. Mol. Biol. *202*, 97–106.

Morris, T.W., Reed, K.E., and Cronan, J.E. (1994). Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. Molecular cloning and characterization of the *lplA* gene and gene product. J. Biol. Chem. *269*, 16091–16100.

Morris, T.W., Reed, K.E., and Cronan, J.E. (1995). Lipoic acid metabolism in *Escherichia coli*: the *lplA* and *lipB* genes define redundant pathways for ligation of lipoyl groups to apoprotein. J. Bacteriol. *177*, 1–10.

Muller, Y.A., and Schulz, G.E. (1993). Structure of the thiamine- and flavin-dependent enzyme pyruvate oxidase. Science *259*, 965–967.

Muller, Y.A., Lindqvist, Y., Furey, W., Schulz, G.E., Jordan, F., and Schneider, G. (1993). A thiamin diphosphate binding fold re-

vealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase. Structure *1*, 95–103.

Nakano, K., Matuda, S., Yamanaka, T., Tsubouchi, H., Nakagawa, S., Titani, K., Otha, S., and Miyati, T. (1991). Purification and molecular cloning of succinyltransferase of the rat a-ketoglutarate dehydrogenase complex. Absence of a sequence motif of the putative E3 and/or E1 binding site. J. Biol. Chem. *266*, 19013–19017.

Nawa, H., Brady, W.T., Koike, M., and Reed, L.J. (1960). Studies on the nature of protein-bound lipoic acid. J. Am. Chem. Soc. *82*, 896–903.

Neagle, J.C., and Lindsay, J.G. (1991). Selective proteolysis of the protein X subunit of the bovine heart pyruvate dehydrogenase complex. Effects on dihydrolipoamide dehydrogenase (E3) affinity and enzymic properties of the complex. Biochem. J. *278*, 423–427.

Neuberger, M., Jourdain, A., and Douce, R. (1991). Isolation of H-protein loaded with methylamine as a transient species in glycine decarboxylase reactions. Biochem. J. *278*, 765–769.

Niu, X., Stoops, J.K., and Reed, L.J. (1990). Overexpression and mutagenesis of the catalytic domain of dihydrolipoamide acetyltransferase from *Saccharomyces cerevisiae*. Biochemistry *29*, 8614–8619.

Oliver, R.M., and Reed, L.J. (1982). Multienzyme complexes. In: Electron Microscopy of Proteins, J. R. Harris, ed. (London: Academic Press), pp. 1–48.

Ono, K., Radke, G.A., Roche, T.E., and Rahmatullah, M. (1993). Partial activation of the pyruvate dehydrogenase kinase by the lipoyl domain region of E2 and interchange of the kinase between lipoyl domain regions. J. Biol. Chem. *268*, 26135–26143.

Opperman, F.B., Steinbüchel, A., and Schlegel, H.G. (1989). Evidence for oxidative thiolytic cleavage of acetoin in *Pelobacter carbinolicus* analogous to aerobic oxidative decarboxylation of pyruvate. FEMS Microbiol. Lett. *60*, 113–118.

Opperman, F.B., and Steinbüchel, A. (1994). Identification and molecular characterization of the *aco* genes encoding the *Pelobacter carbinolicus* acetoin dehydrogenase enzyme system. J. Bacteriol. *176*, 469–485.

Packman, L.C., Hale, G., and Perham, R.N. (1984). Repeating functional domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. EMBO J. *3*, 1315–1319.

Packman, L.C., and Perham, R.N. (1986). Chain folding in the dihydrolipoyl acetyltransferase components of the 2-oxo-acid dehydrogenase complexes from *Escherichia coli*. Identification of a segment involved in binding the E3 subunit. FEBS Lett. *206*, 193–198.

Packman, L.C., and Perham, R.N. (1987). Limited proteolysis and sequence analysis of the 2-oxo acid dehydrogenase complexes from *Escherichia coli*. Cleavage sites and domains in the dihydrolipoamide acetyltransferase components. Biochem. J. *242*, 531–538.

Packman, L.C., Borges, A., and Perham, R.N. (1988). Amino acid sequence analysis of the lipoyl and peripheral subunit-binding domains in the lipoate acetyltransferase component of the pyruvate dehydrogenase complex from *Bacillus stearothermophilus*. Biochem. J. *252*, 79–86.

Packman, L.C., Green, B., and Perham, R.N. (1991). Lipoylation of the E2 components of the 2-oxo acid dehydrogenase multienzyme complexes of *Escherichia coli*. Biochem. J. *277*, 153–158.

Palmer, J.A., Madhusudhan, K.T., Hatter, K., and Sokatch, J.R. (1991). Cloning, sequence and transcriptional analysis of the structural gene for LPD-3, the third lipoamide dehydrogenase of *Pseudomonas putida*. Eur. J. Biochem. *202*, 231–240.

Pares, S., Cohen-Addad, C., Sieker, L., Neuburger, M., and Douce, R. (1994). X-ray structure determination at 2.6-C resolution of a lipoate-containing protein: The H-protein of the glycine decarboxylase complex from pea leaves. Proc. Natl. Acad. Sci. USA *91*, 4850–4853.

Patel, M.S., and Roche, T.E. (1990). Molecular biology and biochemistry of pyruvate dehydrogenase complexes. FASEB J. *4*, 3224–3233.

Patel, M.S., Kerr, D.S., and Wexler, I.D. (1992). Biochemical and molecular aspects of pyruvate dehydrogenase complex deficiency. Int. Pediatr. *7*, 16–22.

Patel, M.S., and Harris, R.A. (1995). Mammalian α-keto acid dehydrogenase complexes: gene regulation and genetic defects. FASEB J. *9*, 1164–1172.

Patel, M.S., Roche, T.E., and Harris, R.A. (1996). Alpha-Keto Acid Dehydrogenase Complexes. (Basel: Birkhäuser Verlag).

Perham, R.N. (1975). Self-assembly of biological macromolecules. Phil. Trans. R. Soc. Lond., Ser. B *272*, 123–136.

Perham, R.N., and Packman, L.C. (1989). 2-oxo acid dehydrogenase multienzyme complexes: Domains, dynamics and design. Ann. NY Acad. Sci. *573*, 1–20.

Perham, R.N., and Roberts, G.C.K. (1981). Limited proteolysis and proton n.m.r. spectroscopy of the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. Biochem. J. *199*, 733–740.

Perham, R.N., Duckworth, H.W., and Roberts, G.C.K. (1981). Mobility of polypeptide chain in the pyruvate dehydrogenase complex revealed by proton NMR. Nature *292*, 474–477.

Perham, R.N. (1991). Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. Biochemistry *30*, 8501–8512.

Pettit, F.H., Hamilton, L., Munk, P., Namihira, G., Eley, M.H., Willms, C.R., and Reed, L.J. (1973). α-Keto acid dehydrogenase complexes: XIX. subunit structure of the *Escherichia coli* α-ketoglutarate dehydrogenase complex. J. Biol. Chem. *248*, 5282–5290.

Pfiefert, H., Hein, S., Krüger, N., Zeh, K., Schmidt, B., and Steinbüchel, A. (1991). Identification and molecular characterization of the *Alcaligenes eutrophus* H16 *aco* operon genes involved in acetoin metabolism. J. Bacteriol. *173*, 4056–4071.

Quinn, J., Diamond, A. G., Masters, A. K., Brookfield, D. E., Wallis, N.G., and Yeaman, S.J. (1993). Expression and lipoylation in *Escherichia coli* of the inner lipoyl domain of the E2-component of the human pyruvate dehydrogenase complex. Biochem. J. *289*, 81–85.

Radford, S.E., Laue, E.D., and Perham, R.N. (1986). Nuclear-magnetic-resonance spectroscopy studies of the flexible linkages between lipoyl domains in the pyruvate dehydrogenase multienzyme complex from *Escherichia coli*. Biochem. Soc. Trans. *14*, 1231–1232.

Radford, S.E., Laue, E.D., Perham, R.N., Miles, J.S., and Guest, J.R. (1987). Segmental structure and protein domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Genetic reconstruction *in vitro* and ^{1}H-n.m.r. spectroscopy. Biochem. J. *247*, 641–649.

Radford, S.E., Laue, E.D., Perham, R.N., Martin, S.R., and Appella, E. (1989). Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Biol. Chem. *264*, 767–775.

Radke, G.A., Ono, K., Ravindran, S., and Roche, T.E. (1993). Critical role of a lipoyl cofactor of the dihydrolipoyl acetyltransferase in the binding and enhanced function of the pyruvate dehydrogenase kinase. Biochem. Biophys. Res. Commun. *190*, 982–991.

Rahmatullah, M., Gopalakrishnan, S., Andrews, P.C., Chang, C.L., Radke, G.A., and Roche, T.E. (1989a). Subunit associations in the mammalian pyruvate dehydrogenase complex. Structure and role of protein X and the pyruvate dehydrogenase component binding domain of the dihydrolipoyl transacetylase component. J. Biol. Chem. *264*, 2221–2227.

Rahmatullah, M., Golalakrishnan, S., Radke, G.A., and Roche, T.E. (1989b). Domain structure of the dihydrolipoyl transacetylase and the protein X components of mammalian pyruvate dehydrogenase complex. Selective cleavage by protease Arg C. J. Biol. Chem. *264*, 1245–1251.

Rahmatullah, M., Radke, G.A., Andrews, P.C., and Roche, T.E. (1990). Changes in the core of the mammalian pyruvate dehydrogenase complex upon selective removal of the lipoyl domain from the transacetylase component but not from the protein X component. J. Biol. Chem. *265*, 14512–14517.

Ravindran, S., Radke, G.A., Guest, J.R., and Roche, T.E. (1996). Lipoyl domain-based mechanism for the integrated feedback control of the pyruvate dehydrogenase complex by enhancement of pyruvate dehydrogenase kinase activity. J. Biol. Chem. *271*, 653–662.

Reed, L.J., Koike, M., Levitch, M.E., and Leach, F.R. (1958a). Studies on the nature and reactions of protein-bound lipoic acid. J. Biol. Chem. *232*, 143–158.

Reed, L.J., Leach, F.R., and Koike, M. (1958b). Studies on a lipoic acid-activating system. J. Biol. Chem. *232*, 123–142.

Reed, L.J. (1974). Multienzyme complexes. Acc. Chem. Res. *7*, 40–46.

Reed, L.J., Pettit, F.H., Eley, M.H., Hamilton, L., Collins, J.H., and Oliver, R.M. (1975). Reconstitution of the *Escherichia coli* pyruvate dehydrogenase complex. Proc. Natl. Acad. Sci. USA *72*, 3086–3072.

Reed, L.J., and Hackert, M.L. (1990). Structure-function relationships in dihydrolipoamide acyltransferases. J. Biol. Chem. *265*, 8971–8974.

Rice, J.E., Dunbar, B., and Lindsay, J.G. (1992). Sequences directing dihydrolipoamide dehydrogenase (E3) binding are located on the 2-oxoglutarate dehydrogenase (E1) component of the mammalian 2-oxoglutarate dehydrogenase multienzyme complex. EMBO J. *11*, 3229–3235.

Robien, M.A., Clore, G.M., Omichinski, J.G., Perham, R.N., Appella, E., Sakaguchi, K., and Gronenborn, A.M. (1992). Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. Biochemistry *31*, 3463–3471.

Russell, G C., and Guest, J.R. (1990). Overexpression of restructured pyruvate dehydrogenase complexes and site-directed mutagenesis of a potential active-site histidine residue. Biochem. J. *269*, 443–450.

Russell, G.C., and Guest, J.R. (1991). Sequence similarities within the family of dihydrolipoamide acyltransferases and discovery of a previously unidentified fungal enzyme. Biochim. Biophys. Acta *1076*, 225–232.

Russell, G.C., Machado, R.S., and Guest, J.R. (1992). Overproduction of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli* and site-directed substitutions in the E1p and E2p subunits. Biochem. J. *287*, 611–619.

Sanderson, S.J., Miller, C., and Lindsay, J.G. (1996). Stoichiometry, organisation and catalytic function of protein X of the pyruvate dehydrogenase complex from bovine heart. Eur. J. Biochem. *236*, 67–77.

Schulze, E., Westphal, A.H., Obmolova, G., Mattevi, A., Hol, W.G.J., and de Kok, A. (1991a). The catalytic domain of the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii* and *Escherichia coli*. Expression, purification, properties and preliminary X-ray analysis. Eur. J. Biochem. *201*, 561–568.

Schulze, E., Benen, J.A.E., Westphal, A.H., and de Kok, A. (1991b). Interaction of lipoamide dehydrogenase with the dihydrolipoyl transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Eur. J. Biochem. *200*, 29–34.

Schulze, E., Westphal, A.H., Boumans, H., and de Kok, A. (1991c). Site-directed mutagenesis of the dihydrolipoyl transacetylase component (E2p) of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Binding of the peripheral components E1p and E3. Eur. J. Biochem. *202*, 841–848.

Schulze, E., Westphal, A.H., Veeger, C., and de Kok, A. (1992). Reconstitution of pyruvate dehydrogenase multienzyme complexes based on chimeric core structures from *Azotobacter vinelandii* and *Escherichia coli*. Eur. J. Biochem. *206*, 427–435.

Schulze, E., Westphal, A.H., Hanemaaijer, R., and de Kok, A. (1993). Structure/function relationships in the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Role of the linker region between the binding and catalytic domain of the dihydrolipoyl transacetylase component. Eur. J. Biochem. *211*, 591–599.

Schweitzer, E. (1996). Fatty acid synthase – Functional strategies of a multienzyme. Naturwissenschaften *83*, 347–358.

Shepherd, G.B., and Hammes, G.G. (1977). Fluorescence energy transfer measurements in the pyruvate dehydrogenase multienzyme complex from *Escherichia coli* with chemically modified lipoic acid. Biochemistry *16*, 5234–5241.

Spencer, M.E., Darlison, M.G., Stephens, P.E., Duckenfield, I.K., and Guest, J.R. (1984). Nucleotide sequence of the *sucB* gene encoding the dihydrolipoamide succinyltransferase of *Escherichia coli* K12 and homology with the corresponding acetyltransferase. Eur. J. Biochem. *141*, 361–374.

Stepp, L.R., Bleile, D.M., McRorie, D.K., Pettit, F.H., and Reed, L.J. (1981). Use of trypsin and lipoamidase to study the role of lipoic acid moieties in the pyruvate and a-ketoglutarate dehydrogenase complexes of *Escherichia coli*. Biochemistry *20*, 4555–4560.

Texter, F.L., Radford, S.E., Laue, E.D., Perham, R.N., Miles, J.S., and Guest, J.R. (1988). Site-directed mutagenesis and ^{1}H NMR spectroscopy of an interdomain segment in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Biochemistry *27*, 289–296.

Toh, H., Kondo, H., and Tanabe, T. (1993). Molecular evolution of biotin-dependent carboxylases. Eur. J. Biochem. *215*, 687–696.

Turner, S.L., Russell, G.C., Williamson, M.P., and Guest, J.R. (1993). Restructuring an interdomain linker in the dihydrolipoamide acetyltransferase component of the pyruvate dehydrogenase complex of *Escherichia coli*. Protein Eng. *6*, 101–108.

Voet, D., and Voet, J.G. (1995). Biochemistry (New York: Wiley and Sons, Inc).

Wagenknecht, T., Grassucci, R., and Schaak, D. (1990). Cryoelectron microscopy of frozen-hydrated α-ketoacid dehydrogenase complexes from *Escherichia coli*. J. Biol. Chem. *265*, 22402–22408.

Wagenknecht, T., Grassucci, R., Radke, G.A., and Roche, T.E. (1991). Cryoelectron microscopy of mammalian pyruvate dehydrogenase complex. J. Biol. Chem. *266*, 24650–24656.

Wagenknecht, T., Grassucci, R., Berkowitz, J., and Forneris, C. (1992). Configuration of interdomain linkers in pyruvate dehydrogenase complex of *Escherichia coli* as determined by cryoelectron microscopy. J. Struct. Biol. *109*, 70–77.

Waldrop, G.L., Rayment, I., and Holden, H.M. (1994). Three-dimensional structure of the biotin carboxylase subunit of ace-tyl-CoA carboxylase. Biochemistry *33*, 10249–10256.

Wallis, N.G., and Perham, R.N. (1994). Structural dependence of post-translational modification and reductive acetylation of the lipoyl domain of the pyruvate dehydrogenase multienzyme complex. J. Mol. Biol. *236*, 209–216.

Wallis, N.G., Allen, M.D., Broadhurst, R.W., Lessard, I.A.D., and Perham, R.N. (1996). Recognition of a surface loop of the lipoyl domain underlies substrate channelling in the pyruvate dehydrogenase multienzyme complex, J. Mol. Biol. *263*, 463–474.

Westphal, A.H., Fabisz-Kijowska, A., Kester, H., Obels, P.P., and de Kok, A. (1995). The interaction between lipoamide dehydrogenase and the peripheral-component-binding domain from the *Azotobacter vinelandii* pyruvate dehydrogenase complex. Eur. J. Biochem. *234*, 861–870.

Williams, C.H., Jr. (1992). Lipoamide dehydrogenase, glutathione reductase, thioredoxin reductase and mercuric reductase-family of flavoenzyme transhydrogenases. In: Chemistry and Biochemistry of Flavoenzymes, F. Müller, ed. (Boca Raton, USA: CRC Press), pp. 121–211.

Wu, T.L., and Reed, L.J. (1984). Subunit binding in the pyruvate dehydrogenase complex from bovine kidney and heart. Biochemistry *23*, 221–226.

Wynn, R.M., Chuang, J.L., Davie, J.R., Fisher, C.W., Hale, M.A., Cox, R. P., and Chuang, D. T. (1992). Cloning and expression in *Escherichia coli* of mature E1β subunit of bovine mitochondrial branched-chain α-keto acid dehydrogenase complex. Mapping of the E1β-binding region on E2. J. Biol. Chem. *267*, 1881–1887.

Yang, Y.-S., and Frey, P.A. (1986). Dihydrolipoyl transacetylase of *Escherichia coli*. Formation of 8-*S*-acetyldihydrolipoamide. Biochemistry *25*, 8173–8178.

Yang, H., Frey, P.A., Hainfeld, J.F., and Wall, J.S. (1986). Pyruvate dehydrogenase complex of *Escherichia coli*. Radial mass analysis of subcomplexes by scanning transmission electron microscopy. Biophys. J. *49*, 56–58.

Yang, Y.-S., Datta, A., Hainfeld, J.F., Furuya, F.R., Wall, J.S., and Frey, P.A. (1994). Mapping the lipoyl groups of the pyruvate dehydrogenase complex by use of gold cluster labels and scanning transmission electron microscopy. Biochemistry *33*, 9428–9437.

Yeaman, S.J. (1989). The 2-oxo acid dehydrogenase complexes: recent advances. Biochem. J. *257*, 625–632.

Zhu, P.-P., and Peterkofsky, A. (1996). Sequence and organization of genes encoding enzymes involved in pyruvate metabolism in *Mycoplasma capricolum*, Protein Science *5*, 1719–1736.

Biol. Chem., Vol. 378, pp. 635 – 640, July 1997

Review

Aims and Limitations in the Use of Antipeptide Antibodies in Molecular Biology

Alberto Chersi*, Francesca Di Modugno and Laura Rosano'

Istituto Regina Elena for Cancer Research, CRS, Laboratory of Biochemistry, Via delle Messi d'oro 156, I-00158 Rome, Italy

* Corresponding author

Antibodies to peptides obtained by synthesis and, to a much lesser extent, prepared by enzymatic digestion of proteins, have been widely used in the last ten years in a variety of immunochemical and biological investigations. There are however several limitations in the correct utilization of such reagents. In fact, in spite of their 'predetermined specificity', antipeptide antibodies often fail to discriminate related molecules, and their reactivity with native proteins may be scarce or even absent, even if the peptide has been selected from surface regions of the protein. Our critical point of view, concerning two main aspects of antipeptide antibody features, i.e. specificity and reactivity, will be presented here, as confronted with information from the available literature. We have selected a restricted number of references among hundreds of publications dealing with antipeptide antibodies: for sure we neglected outstanding papers on the subject, and we apologize in advance.

Key words: Antibodies / Crossreactivity / Epitopes / Peptides / Protein conformation.

Introduction

In the early 1980's, several investigators from different laboratories around the world reported that synthetic peptides could be successfully employed for the generation of protein-reactive antibodies.

This approach was not completely new, insofar as early as 1963, in the effort to induce protective immunity against infections, Anderer (1963) reported that a natural hexapeptide obtained from the enzymatic digest of TMV, once conjugated to BSA, elicited antibodies with a partial inhibitory effect against the infectivity of the virus. This principle was verified in 1976, when it was described (Langbeheim *et al.*, 1976) that a coliphage could be totally inactivated by antisera raised against a synthetic peptide corresponding to a 20-residue fragment of the viral coat protein. At that time, however, experiments designed to test this approach, and develop the new technique for the production of antipeptide antibodies, were hindered by price and unavailability of synthetic peptides and by the paucity of sequence data on proteins of biological interest. But in the early 80's, the technical improvements in the sequencing of proteins and nucleic acids, and the commercialization of automatic peptide synthesizers and protected amino acids, increased the possibilities for the rapid synthesis of peptides. The interest for antipeptide antibodies was prompted by the finding that they had a remarkable utility in a wide range of immunological as well as biological studies, as the detection of trace amounts of proteins and the study of interactions between macromolecules. Present and future possibilities of this technique were widely examined and anticipated with emphasis in several early reports (Bittle *et al.*, 1982; Green *et al.*, 1982; Jacob *et al.*, 1983).

There is no doubt that, although many laboratories around the world, (including ours), are still using antipeptide antibodies, such reagents, in spite of their 'predetermined specificity', often fail to provide unambiguous results, especially concerning the selective recognition of macromolecules. Even monoclonal antibodies to peptides, which combine the purity of monoclonal reagents with the predetermined target of the antibody, sometimes display scarce specificity and selectivity.

For most molecular biology studies, an ideal antipeptide antibody should display the following two features:

(i) It should exhibit a restricted specificity.

(ii) It should possibly react with native proteins.

Both features appear to be often defective in such reagents. Our critical point of view, concerning specificity and reactivity of antibodies to peptides, will be presented here, as confronted with information from the available literature.

Specificity of Antipeptide Antibodies

Apart from large peptides (30 residues or more) that can be injected uncoupled into animals (Del Giudice *et al.*, 1986; Zamorano *et al.*, 1995), the large majority of peptides must be conjugated to a carrier protein in order to elicit an immune response.

The most common procedure for preparing conjugates of peptides with high-molecular weight carrier proteins is the addition of an aqueous glutaraldehyde solution to a

mixture of the two components, in order to form a bridge between amino groups of the compounds. The technique necessarily introduces a modification into the chemistry of the peptide. In addition, if the peptide contains one or more lysine residues, the ε-amino groups will often participate in the coupling reaction, and the final conjugate is then an undefined chemical entity with severe alterations in the original peptide structure. Almost every chemical modification in the peptide structure results in a change of immunogenicity and thus in a modification of the immune response. This also holds for conjugates prepared by aid of water-soluble carbodiimides, insofar as carboxyl groups of aspartic acid and glutamic acid may participate in the coupling reaction, thus increasing the complexity of the final product. Peptides containing a cysteine residue have the alternative possibility of being coupled through the sulfhydryl group to the amino groups of the carrier by aid of MBS (Liu *et al.*, 1979) or related heterobifunctional reagents (Jones *et al.*, 1989). By this technique, carrier and peptide are coupled in a deliberate manner and in a chemically unambiguous way. Cys-containing peptides, however, are easily subjected to oxidation and dimerization during purification processes, and thus must be handled with care.

On the other side, very long peptides, which can be injected uncoupled, are difficult to isolate and purify, and are often insoluble. A 99% coupling yield during synthesis is compulsory, in order to avoid at the end an intractable mixture of closely related fragments. The complexity and heterogeneity of the immune response closely correlates with the homogeneity of the peptide used as immunogen, as all contaminants may represent additional determinants. Fragments with deletion generate antibodies with new specificites.

Besides rabbits, mice and guinea pigs have been occasionally employed for generating polyclonal antibodies to peptides (Jones *et al.*, 1989; Parry *et al.*, 1988); there are in addition several reports concerning the preparation of mouse monoclonal antipeptide antibodies (Niman *et al.*, 1983; Smith *et al.*, 1993).

When using peptide-carrier conjugates as immunogens, the immune response is also directed against the carrier protein, as well as against several chemical reagents used for the conjugation. Those antibodies, however, can be then easily removed by using suitable immunoadsorbents. An alternative approach for preparing peptide immunogens, known as multiple peptide antigens (MAPs), has been described by Tan and coworkers (Posnett *et al.*, 1988; Tam, 1988): a prototype of a MAP consists of seven lysyl residues as the inner core, and eight copies of the peptide. This polymer is synthesized directly as a single unit by the solid-phase method, and gives rise to an unambiguous, well-characterized conjugate with molecular weight exceeding 10000 daltons. It is claimed that those polymers provide faster and more homogeneous immunological responses, as compared to conventional peptide-KLH conjugates, for the following reasons:

a) They appear to be 'high-density' immunogens, i.e., compounds where the molar ratio peptide/core is very high.
b) Their molecular weight (10000 and above) eliminates the necessity of using a carrier protein, avoiding chemical modifications of the peptide's reactive groups for the preparation of conjugates.
c) The polylysyl core itself functions as the carrier, and does not give rise to the production of anti-carrier antibodies.

In spite of the fact that recent data partially contradict such statements (Briand *et al.*, 1992; Chersi *et al.*, 1995), the method still remains as a valid alternative to conventional immunizations with peptide-KLH complexes. Finally, some investigators have immunized rabbits with peptidyl-polyamide particles, with surprisingly good results (Fisher *et al.*, 1989).

There is no doubt that the immune response against a medium-sized peptide (12 to 20 amino acids) is polyclonal, and the specificity of such a complex mixture of antibodies with different specificites and affinities, each directed to one of the many linear, and sometimes conformational, antigenic determinants of the peptide, may be low. Monoclonal antibodies to peptides offer the advantage that each of them is directed against a single epitope of the fragment. This might be quite advantageous for relatively long peptides (30 residues or more), comprising a high number of sequential antigenic determinants, and different conformations. As an example, in a single fusion experiment, a 36-mer synthetic peptide produced as much as 21 cell lines secreting stable antibodies reactive with the peptide (Niman *et al.*, 1983). It is obvious that for small peptides, where the number of determinants is small, monoclonal antibodies do not offer substantial advantages as compared to rabbit polyclonal antipeptide antibodies, considering their cost and the time necessary for their preparation.

In contrast to antibodies to proteins, antipeptide antibodies appear to be relatively nonspecific and often cross-react with molecules that apparently share little homology with the immunogen: three or four amino acid residues, if placed at key positions, may be sufficient for generating crossreactivities of antipeptide sera with related or even unrelated molecules. This event is not always predictable from the sequence homology between two molecules: as an example, three closely related peptides from the variable region of HLA-DQ3, -DQ1 and -DQ2 membrane glycoproteins:

A) SQKEVLERTRAELDTV
B) SQKEVLEGARASVDRV
C) SQKDILERKRAAVDRV

elicit antibodies with quite different specificites: anti-B heavily crossreacts with peptide C (sequence homology: 68%), but very little with peptide A, with the same sequence homology. Accordingly, anti-A does not react with either B or C, and thus appears to be relatively specific. Those differences in specificity were explained by the find-

ing that the major epitope in all peptides was the carboxy-terminal 7-residue fragment where the homology between B and C was extremely high (Chersi *et al.*, 1986).

A detailed study on the specificity of rabbit antipeptide antibodies has been conducted by Geysen and coworkers. Taking advantage of the multipin synthesis (Geysen *et al.*, 1986), those authors examined a panel of hexa-, hepta- and octapeptides displaying single amino acid substitutions, as compared to the synthetic fragment used for eliciting antipeptide antibodies, and determined position and nature of the residue that could be substituted without affecting the binding of the antibody to the peptide(s). As an example, the antiserum elicited by the hexapeptide G D L G S I crossreacts with several variants of the peptide. Three of the six positions (2, 3, 4) do not practically allow amino acid substitutions (only Asn for Asp, Ile for Leu, and Ala for Gly). The three others, on the contrary, appear to be quite permissive, and allow a wide range of replacements that only marginally affect the ability of the fragment to be recognized by the antibody. This means that this antipeptide antibody might crossreact with any protein fortuitously containing the sequence X-D-L-G-X-X (or its variants X-N-L-G-X-X, X-D-I-G-X-X, X-D-L-A-X-X). In this particular case, with each of the three essential amino acid residues quite common in proteins, the probability of encountering a D-L-G sequence in any related or unrelated protein is relatively high.

In some cases, the crossreactivity can be predicted by analyzing related peptide sequences by aid of a computer program provided with suitable corrections and approximations, as Asn = Asp, Gln = Glu, Ile = Leu, Ala = Gly (conservative substitutions, factor 0.90). For semi-conservative substitutions, the factor is 0.50. In some cases, however, no obvious homology between the immunogen and the crossreactive peptide(s) can be detected (Dibwad *et al.*, 1995), while in others, very closely related peptides generate non-crossreacting antibodies (Nakamura *et al.*, 1995).

All those observations led to the prediction, soon experimentally confirmed, that the majority of rabbit antipeptide antibodies might be unsuitable for recognizing selectively a single component within a complex mixture of closely related proteins. Nevertheless, very good quality, specific antibodies were sometimes obtained, already in the first years of development of this technique. For instance, some investigators (Walker *et al.*, 1985; Choppin *et al.*, 1986), succeeded in producing antibodies against peptides homologous to fragments of HLA-B7 heavy chain, which selectively immunoprecipitated the heavy chain-β2 complex only from cell lines serologically expressing this specificity, notwithstanding the extended sequence homology occurring among Class I histocompatibility antigens. Quite recently, in spite of the scarce (and often unpredictable) specificity displayed by most antipeptide antibodies, several investigators, by using as immunogens Tyr-phosphorylated peptides (Bangalore *et al.*, 1992; Epstein *et al.*, 1992; 1995), have succeeded in producing antipeptide sera that preferentially and quite specifically recognize proteins only when they are tyrosine-phosphorylated. The significance of those reagents lies in their ability to identify and quantify subsets of tyrosine kinases and substrates that differ in phosphorylation and hence functionality. Such anti-phosphatyrosinepeptide sera will greatly simplify cell-free analysis of tyrosine phosphorylation. That is, as to the specificity of an antipeptide antibody (or lack of it) for the desired region of the target, and affinity for it, these can be found only '*a posteriori*', a feature which is shared with the production of monoclonal antibodies.

On the other hand, when peptides are suitably selected from protein regions that do not share close sequence homologies, the antisera often display sufficient selectivity to be used in immunoprecipitation and western blotting (Marquardt *et al.*, 1990; Stricker *et al.*, 1995) or even for tissue staining (Creaney *et al.*, 1995; Harrison *et al.*, 1996). Monoclonal antipeptide antibodies, however, may be more suitable for those tests, since they are generally more specific.

The aspecific binding to proteins is well evidenced by testing an antipeptide antiserum on a complex protein mixture after separation of bands by SDS-PAGE, and transfer to nitrocellulose sheets (Western blotting). Incubation of strips with the antipeptide serum, followed by an antirabbit-peroxidase-labeled second antibody, and then a suitable substrate, reveals a very complex reactivity pattern. Although some proteins might be related to the antigen, being precursors or degradation products with similar constitutive features, there is no doubt that many crossreactivities are occasional and fortuitous. The reactivity pattern is remarkably less complex when affinity-purified antipeptide antibodies are used instead of antisera. Also in immunoprecipitation experiments, followed by SDS-PAGE and Western blotting, the reactivity pattern is generally quite complex, although the desired protein band can be generally identified because of its higher reactivity and/or its position. Clear-cut experiments can be considered as relatively uncommon.

Binding of Antipeptide Antibodies to Native Proteins

The first question to be answered, when selecting peptides for the preparation of immunogens, is whether the peptide has a surface location in the native protein. Since only antibodies to exposed segments have the possibility of reacting with the intact macromolecule, peptides should be preferably selected from protein surface regions. When the three-dimensional structure of the protein has not yet been determined (as, for instance, from X-rays parameters such as atomic mobility, surface exposure, solvent accessibility, and protrusion), protein surface domains and antigenic determinants can be sometimes predicted from the local hydrophilicity of protein segments (Hopp and Woods, 1981; Eisenberg *et al.*, 1984), although this parameter does not always correlate with the reac-

tivity of antipeptide antibodies with native proteins. A more reliable approach (Krchknak *et al.*, 1987) is based on the calculation of the probability (p) of forming β-turns of four consecutive amino acid residues in the protein, using the Chou-Fasman method for predicting potential surface determinants (Chou and Fasman, 1978). The validity of this approach was assessed by selecting from the literature a panel of oncopeptides (Tanaka *et al.*, 1985; Niman *et al.*, 1985) whose efficacy in generating protein-reactive antipeptide antibodies had been previously experimentally demonstrated: each fragment was analyzed for the tetrapeptide sequence with the highest probability of forming β-turns. A histogram was then drawn showing the relation between the highest probability of β-turn occurrence (p) and the frequency distribution of reactivity of antipeptide sera. By this method, the reactivity of 26 out of 27 (96%) experimentally positive antisera could be theoretically anticipated.

Peptide size is also an important factor: Fragments shorter than 10 residues seldom elicit protein-reactive antibodies. The reason may be statistical: the longer the peptide, the greater the probability that it contains the right number of amino acids forming a β-turn (Krchknak *et al.*, 1987).

The observation that synthetic peptides may elicit antibodies recognizing the segment in the native protein has led to several hypotheses about the structural basis for the generation of the immune response. The initial assumption was that peptides adopt in solution several different conformations, some of which might approximate its cognate structure in the intact protein molecule (Niman *et al.*, 1983). Thus, an antiserum to a peptide might be considered as a collection of antibodies to many different conformations, some of which might then be reactive with the folded protein.

The percentage of antipeptide antibodies recognizing the native protein can be roughly estimated by inhibition techniques or by the use of immunoadsorbents. Briefly, one can measure with an ELISA the residual antibody activity for the peptide, after incubation with an excess of native protein, or, alternatively, first bind the intact protein to an inert matrix, and then calculate the percent of antibody that does not bind to the whole antigen, but still reacts with the free peptide (Chersi and Houghten, 1984). In most cases, the percentage of antipeptide antibodies reacting with the native protein can be estimated to be on the order of 15–25%. This percentage fits well with the hypothesis that short peptides (10–20 residues) are disordered in aqueous solutions, i.e. they can adopt a multiplicity of conformations. In the so-called stochastic hypothesis, the native-like conformation of the peptide is assumed to be rare; consequently, only a small number of peptide molecules elicit antibodies capable of recognizing the native protein. Nuclear magnetic resonance studies confirm that only few peptides exhibit secondary structure in water.

Monoclonal antibodies provide a more precise estimation of the frequency of reactivity of antipeptide antibodies with native proteins: This is given by the ratio of the number of clones producing protein-reactive antibodies to the total number of clones secreting antibodies reacting with the peptide. Taking into consideration only cases in which the number of clones responding to the immunogen (the peptide) is sufficiently large (5 or more), this frequency seems to be on the order of 30 to 50%. This percentage is higher than in the case of rabbit polyclonal antipeptide antibodies, but obviously a direct comparison could be made only by analyzing simultaneously the monoclonal and polyclonal response to the same peptide. It is also possible that in some experiments the antibodies were reacting in ELISA with native proteins because those molecules had undergone a partial denaturation in the immunoassay procedures.

It has been claimed that MAP antigens (i.e., a polymer consisting of 8 copies of peptide linked together by a branched polylysyl core), which can be injected directly without being previously coupled to a carrier protein, would have a higher probability of generating protein-reactive antibodies (Tam, 1988). This assumption seems to be contradicted by recent experiments, since there were no substantial differences in frequencies of protein-reactive antibodies in sera elicited by MAPs or by the more conventional KLH-peptide complexes (Briand, 1992).

Antibodies that do not react with the native macromolecule might be directed against a fragment that is hidden in the interior of the native molecule; alternatively, the antipeptide antibody might be prevented from reacting if the target fragment assumes, in the native structure, a rigid conformation. Flexibility of segments of proteins appears thus to play an important role in epitope recognition: It might be postulated that antipeptide antibodies bind to the native molecule only when the local flexibility of short segments of the protein mimics the flexibility of the peptide in solution (Westhof *et al.*, 1984).

As epitopes located in rigid segments of the protein might not be recognized by antipeptide antibodies, the prediction of chain flexibility of protein segments might be better suited for selecting peptides than methods based on hydrophilicity indexes. As an example, antibodies to N-terminal or C-terminal peptides generally react with the native protein, and this is due to the fact that the first 10, and the last 10, amino acid residues in almost every protein are characterized by high chain flexibility (Karplus and Schulz, 1985). As the rigidity of the polypeptide chain may prevent antipeptide antibodies from reacting, this might explain why antibodies elicited by peptides selected from extracellular segments of several membrane proteins bind poorly to intact cells (Di Modugno *et al.*, 1995). Apart from restriction in mobility exerted by other molecules, such proteins are anchored to the lipid bilayer of the membrane through multiple, short hydrophobic segments, and this feature prevents flexibility and thus the binding by the antibody.

In ELISA assays, the three-dimensional structure of a protein is often altered by the exposure to air and by the drying onto microtiter plates. This may result in conformational changes and in the exposure of hidden polypeptide

segments. Subsets of antipeptide antibodies that do not react originally with the macromolecule in its native form may now recognize the antigen. It is then predictable that the reactivity of the antipeptide antibody with the protein in ELISA may not be related to its ability to react with the native molecule in solution.

Most proteins extracted from cells by detergents are also generally subjected to a partial modification of their structure, and it should be assumed that ELISA tests performed on such material do not reflect either the binding to a native, or to a completely denatured protein.

On the other hand, antipeptide sera are likely to recognize target proteins after SDS-PAGE in reducing conditions and Western Blotting, while conventional monoclonal or polyclonal antibodies to proteins, being mostly directed against conformational determinants, often fail to react. However, the reactivity pattern is often quite complex and does not provide definitive conclusions. Antipeptide antibodies purified by affinity chromatography may offer more significant results. Preincubation of the antipeptide antibody with the free peptide offers an additional criterium for identifying the target protein, since this treatment preferentially inhibits specific bindings. There are several reports concerning the successful utilization of polyclonal or monoclonal antipeptide antibodies in these assays (Niman *et al.*, 1983; Walker *et al.*, 1985; Marquardt *et al.*, 1990; Stricker *et al.*, 1995). In a few cases, rabbit antipeptide antibodies failed to react under such conditions. (Haspel *et al.*, 1988). Since the three-dimensional structure of the protein is modified in the process, with unfolding of the polypeptide chain, it is not clear why polyclonal antipeptide antibodies, directed to linear sequential epitopes, are unable to react with the random-coiled molecule. A possible explanation is that the protein is not expressed, or, alternatively, that the sequence of the segment in the protein, as deduced from DNA sequence, is not that reported in the literature. Finally, errors in peptide synthesis, or persistence of protective groups in the fragment, might produce antipeptide antibodies with quite different specificities from those required for reacting with the antigen.

The final point to be discussed is whether this technology is given too much credit with respect to the results it delivers. The answer is dependent on what the investigator expects from those reagents. The term 'predetermined specificity' does not imply that antipeptide antibodies are indeed and always truly specific. If should be noted, however, that the level of specificity required need not be very high for all applications: For preparative purposes, an affinity step with the antipeptide antibody suitably bound to an inert matrix, may result in a remarkable enrichment of the desired material; this step can be then eventually complemented by more conventional purification methods.

As far as concerns the reactivity of antipeptide antibodies with native proteins, while it was predictable from the very beginning, and thus accepted, that those reagents would not recognize protein segments buried in the interior of macromolecules, it was not so obvious that the reactivity might be so highly dependent on the conformation and flexibility of the polypeptide chain. This feature definitely contributed to reducing the usefulness of antipeptide antibodies in immunology and molecular biology.

Acknowledgements

The authors thank Mrs. Giuliana Falasca and Mr. Antonio Federico for their technical assistance, and Mr. Fabian Chersi for Computer programs.
This work was partly supported by a grant from Ministero della Sanita'. FDM was supported by a fellowship by Associazione Italiana Ricerca Cancro.

References

Anderer, F.A. (1963). Preparation and properties of an artificial antigen immunologically related to Tobacco Mosaic Virus. Biochim. Biophys. Acta *71*, 246–248.

Bangalore, L., Tanner, A.J., Laudano, A.P., and Stern, D.F. (1992). Antiserum raised against a synthetic phosphotyrosine-containingpeptide selectively recognizes p185 neu/erbB-2 and the epidermal growth factor receptor. Proc. Natl. Acad. Sci. USA *89*, 11637–11641.

Bittle, J.L., Houghten, R.A., Alexander, H., Shinnick, T.M., Sutcliffe, J.G., and Lerner, R.A. (1982). Protection against foot-and-mouth disease by immunization with a chemically synthesized peptide predicted from the viral nucleotide sequence. Nature *298*, 30–33.

Briand, J.P., Barin, C., Van Regenmortel, M.H.V., and Muller, S. (1992). Application and limitations of the Multiple Antigen Peptide (MAP) system in the production and evaluation of anti-peptide and anti-protein antibodies. J. Immunol. Meth. *156*, 255–265.

Chersi, A., and Houghten, R.A. (1984). Rabbit antipeptide antibodies against restricted domains of the histocompatibility complex. Z. Naturforsch. *39*, 646–651.

Chersi, A., Houghten, R.A., Chillemi, F., Zito, R., and Centis, D. (1986). Specificity of rabbit antibodies elicited by related synthetic peptides. Z. Naturforsch. *41*, 613–617.

Chersi, A., Di Modugno, F., and Falasca, G. (1995). Specificities of rabbit antisera to Multiple Antigen (MAP) peptides. Z. Naturforsch. *50*, 735–738.

Choppin, J., Metzger, J.J., Bouillot, M., Briand, J.P., Connan, F., Van Regenmortel, M.H.V., and Levy, J.P. (1986). Recognition of HLA molecules by antisera directed to synthetic peptides corresponding to different regions of the HLA-B7 heavy chain. J. Immunol. *136*, 1738–1744.

Chou, P.Y., and Fasman, G.D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzym. *45*, 46–148.

Creaney, J., Wijffels, G.L., Sexton, J.L., Sanderman, R.M., Spithill, T.W., and Parsons, J.C. (1995). Fasciola Hepatica: localisation of glutathione S-transferase isoenzymes in adult and juvenile liver fluke. Exp. Parasit. *81*, 106–116.

Del Giudice, G., Cooper, J.A., Merino, J., Verdhini, A.S., Pessi, A., Togna, A.P., Engers, H.D., Corradin, G., and Lambert, P.H. (1986). The antibody response in mice to carrier-free synthetic peptides of Plasmodium Falciparum Circumsporozoite repetitive epitope is I-A restricted. Possible implications for Malaria vaccines. J. Immunol. *137*, 2952–2955.

Di Modugno, F., Nista, A., Federico, A., and Falasca, G. (1995). Antipeptide antibodies to P-glycoprotein: now effective? J.E.C.C.R. *14*, 171–174.

Dibwad, A., Bogen, B., Natwig, J.B., Forre, O., and Sioud, M. (1995). Peptide phage libraries can be an efficient tool for identifying antibody ligands for polyclonal antisera. Clin. Exp. Immun. *102*, 438–442.

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface proteins sequences with a hydrophobic moment plot. J. Mol. Biol. *179*, 125–142.

Epstein, R.J., Druker, B.S., Roberts, T.M., and Stiles, C.D. (1992). Synthetic phosphopeptide immunogens yield activation-specific antibodies to the C-erb B-2 receptor. Proc. Natl. Acad. Sci. USA *89*, 10435–10439.

Epstein, R.J. (1995). Preferential detection of catalytically inactive c-erbB-2 by antibodies to unphosphorylated peptides mimicking receptor tyrosine autophosphorylation sites. Oncogene *11*, 315–323.

Fisher, P.M., Comis, A., and Howden, M.H.V. (1989). Direct immunization with synthetic peptidyl-polyamide resins. J. Immunol. Meth. *118*, 119–123.

Geysen, H.M., Rodda, S.J., and Mason, T.J. (1986). A priori delineation of a peptide which mimics a discontinuous antigenic determinant. Mol. Immunol. *23*, 709–715.

Green, N., Alexander, H., Olson, A., Alexander, S., Shinnick, T.M., Sutcliffe, J.G., and Lerner, R.A. (1982). Immunigenic structure of the Influenza Virus Hemagglutinin. Cell *28*, 477–487.

Harrison, J.B., Chen, H.H., Sattelle, E., Barker, P.J., Huskisson, N.S., Rauh, J.J., Bai, D., and Sattelle, D.B. (1996). Immunochemical mapping of a C-terminus antipeptide antibody to the GABA receptor subunit in the nervous system in Drosophila Melanogaster. Cell and Tissue Res. *284*, 269–278.

Haspel, H.C., Rosenfeld, M.G., and Rosen, O.M. (1988). Characterization of antisera to a synthetic carboxy-terminal peptide of the glucose trasporter protein. J. Biol. Chem. *263*, 398–403.

Hopp, T.P., and Woods, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA *78*, 3824–3828.

Karplus, P.A., and Schulz, G.E. (1985). Prediction of chain flexibility in proteins. Naturwiss. *12*, 212–213.

Krchknak, V., Mach, O., and Maly, A. (1987). Computer prediction of potential immunogenic determinants from protein amino acid sequences. Analyt. Biochem. *165*, 200–208.

Jacob, C.O., Sela, M., and Arnon, R. (1983). Antibodies against synthetic peptides of the B subunit of cholera toxin: crossreaction and neutralization of toxin. Proc. Natl. Acad. Sci. USA *80*, 7611–7615.

Jones, G.L., Edmundson, H.M., Spencer, L., Gall, J., and Sahl, A. (1989). The use of maleimidocaprolyloxy succinimide to prepare malarial peptide-carrier conjugate. J. Immunol. Meth. *123*, 211–216.

Langbeheim, H., Arnon, R., and Sela, R. (1976). Antiviral effect on MS-2 coliphage obtained with a synthetic antigen. Proc. Natl. Acad. Sci. USA *73*, 4636–4640.

Liu, F.T., Zinnacker, M., and Katz, D. (1979). New procedure for the preparation and isolation of conjugates of proteins and synthetic peptides of D amino acids, and immunochemical characterization of such conjugates. Biochemistry *18*, 690–695.

Marquardt, D., McCrone, S., and Center, M.S. (1990). Mechanism of multidrug resistance in HL60 cells: detection of resistance-associated proteins with antibodies against synthetic peptides that correspond to the deduced sequence of P-glycoproteins. Cancer Res. *50*, 1426–1430.

Nakamura, A., Yamamoto, Y., Tasaki, T., Sugimoto, C., Masuda, M., Kazusaka, A., and Fujita, S. (1995). Anti-peptide antibodies to the P4502D subfamily in rat, dog and man. Xenobiotica *25*, 1103–1109.

Niman, H.L., Houghten, R.A., Walker, L.E., Reisfeld, R.A., Wilson, I.A., Hogle, J.M., and Lerner, R.A. (1983). Generation of protein-reactive antibodies by short peptides is an event of high frequency: implications for the structural basis of immune recognition. Proc. Natl. Acad. Sci. USA *80*, 4949–4953.

Niman, H.L., Thompson, A.M.H., Yu, A., Markman, M., Willems, J.J., Herwig, K.R., Habib, N.A., Wood, C.B., Houghten, R.A., and Lerner, R.A. (1985). Antipeptide antibodies detect oncogene-related proteins in urine. Proc. Natl. Acad. Sci. USA *82*, 7924–7928.

Parry, N.R., Syred, A., Rowlands, D.J., and Brown, F: (1988). A high proportion of anti-peptide antibodies recognize foot-and-mouth disease virus particles. Immunology *64*, 567–572.

Posnett, D.N., McGrath, H., and Tam, J.P. (1988). A novel method for producing antipeptide antibodies. J. Biol. Chem. *263*, 1719–1725.

Smith, C., McIntosh, N., and James, K. (1993). Pitfalls in the use of ELISA to screen for MoAbs raised against small peptides. J. Immunol. Meth. *158*, 151–160.

Stricker, R., Kalbacher, H., Lottspeich, F., and Reiser, G. (1995). Characterization of a high-affinity Ins-P4 (inositol-phosphate) receptor from brain by an anti-peptide antiserum. FEBS Letters *370*, 236–240.

Tam, J.P. (1988). Synthetic peptide vaccine design: Synthesis and properties of a high density multiple antigen peptide system. Proc. Natl. Acad. Sci. USA *85*, 5409–5413.

Tanaka, T., Slamon, D.J., and Cline, M.J. (1985). Efficient generation of antibodies to oncoproteins by using synthetic peptide antigens. Proc. Natl. Acad. Sci. USA *82*, 3400–3404.

Walker, L.E., Ketler, T.A., Houghten, R.A., Schulz, G., Chersi, A., and Reisfeld, R.A. (1985). Human major histocompatibility complex class I antigens: residues 61–83 of HLA-B7 heavy chains specify an alloreactive site. Proc. Natl. Acad. Sci. USA *82*, 539–542.

Westhof, E., Altschuh, D., Moras, D., Bloomer, A.C., Mondragon, A., Klug, A., and Regenmortel, M.H.V. (1984). Correlation between segmental mobility and the location of antigenic determinants in proteins. Nature *311*, 123–126.

Zamorano, P., Wigdorovitz, A., Perez-Filgueira, M., Carrillo, C., Escribano, J.M., Sadir, A.M., and Borca, M.V. (1995). A ten-amino-acid linear sequence of VP1 of foot and mouth disease containing B- and T-cell epitopes induces protection in mice. Virology *212*, 614–621.