Proteins evolve in a fitness landscape encompassing a complex network of biological constraints. Due to the interrelation of folding, function, and regulation, the ground-state structure of a protein may be inactive. A model is provided by insulin, a vertebrate hormone central to the control of metabolism. Whereas native assembly mediates storage within pancreatic β-cells, the active conformation of insulin and its mode of receptor binding remain elusive. Here, functional surfaces of insulin are probed by photo-cross-linking of an extensive set of azido-derivatives constructed by chemical synthesis. Contacts are circumferential, suggesting that insulin is encaged within its receptor. Mapping of photo-products to the hormone-binding domains of the insulin receptor demonstrates alternating contacts by the B-chain β-strand (residues B24-B28). Whereas even-numbered probes (at positions B24 and B26) contact the N-terminal L1 domain of the α-subunit, odd-numbered probes (B25 and B27) contact its C-terminal insert domain. This alternation corresponds to the canonical structure of a β-strand – wherein successive residues project in opposite directions – and so suggests that the B-chain inserts between receptor domains. Detachment of a receptor-binding arm enables photo-engagement of surfaces otherwise hidden in the free hormone. The arm and associated surfaces contain sites also required for nascent folding and self-assembly of storage hexamers. The marked compression of structural information within a short polypeptide sequence rationalizes the diversity of diabetes-associated mutations in the insulin gene. Our studies demonstrate that photo-scanning mutagenesis can decode the active conformation of a protein and so illuminate cryptic constraints underlying its evolution.

INTRODUCTION

Insulin is a small globular protein containing two chains, A (21 residues) and B (30 residues). Although its unbound structure is well characterized (1-3), how the hormone binds to the insulin receptor (IR) is not well understood. In this article we investigate the receptor-binding surface of insulin by synthetic photo-scanning. An extensive set of photo-activatable insulin analogs was prepared, each containing a single para-azido-Phe (Pap) substitution (4) (Fig. 1A). The modular structure of the IR (5) permits assignment of contact points to the major hormone-binding elements of the receptor α-subunit (6). Analysis of photo-products provides evidence for the structural reorganization of insulin within the crux of the receptor ectodomain (7). An alternating pattern of contacts by the B-chain β-strand provides a signature of its insertion between receptor domains.

The IR is a disulfide-linked dimer (designated (αβ)2) in which extracellular α-subunits bind insulin and transmembrane β-subunits contain the cytoplasmic tyrosine-kinase (TK) domain (Fig. 1B) (5). Mutagenesis has established that the major hormone-binding regions are the N-terminal L1 β-helix and C-terminal insert-domain-derived tail (ID-N) of the α-subunit (8). Whereas the cysteine-rich domain (CR) and L2 β-helix contribute primarily to the architecture of the ectodomain, insulin may also contact the fibronectin homology domains of the α-subunit (Fn1 and part of Fn2) (9,10). The homologous type 1 IGF receptor (IGFR) exhibits a similar domain organization. Studies of chimeric receptors suggest that the IR L1 domain confers specificity for insulin (relative to insulin-like growth factor I; IGF-I) (11-13). Although co-

Copyright 2009 by The American Society for Biochemistry and Molecular Biology, Inc.
crystals of the insulin-IR complex have not been obtained to date, evidence for selected domain-specific contacts has been provided by photo-cross-linking (4,14-17). The modular structure of the α-subunit enables convenient assignment of photo-adducts to L1 or ID-N by limited proteolysis (6). Key tryptic and chymotryptic sites (Fig. 1B, left) are exposed at domain boundaries (supplemental Figs. S1 and S2), leading to photo-adducts of characteristic molecular mass (Table 1).

Insight into the receptor-binding surface of insulin has been obtained from analysis of species variants, chemical modification, and mutagenesis (Fig. 1C, D) (5,18). The classical binding surface spans the N-terminal A-chain α-helix (residues A1-A8), the central B-chain α-helix (B9-B19), and the C-terminal B-chain β-strand (B24-B28) (1,19-21). Mutagenesis defines a functional epitope (by analogy to growth hormone (22)) containing residues GlyA1, IleA2, ValA3, and in the C-terminal B-chain β-strand (residues B24-B28) (1,19-21). Mutagenesis defines a functional epitope (by analogy to growth hormone (22)) containing residues GlyA1, IleA2, ValA3, GlnA5, TyrA9, AsnA21, ValB12, GlyB13, PheB24, PheB25, and TyrB26 (light gray in Fig. 1C, D) (18,23); an extended structural epitope (i.e., contacts at which Ala substitutions have minor effect (24)) comprises the contiguous side chains of ThrA8 and TyrB16 (green in Fig. 1C, D) (25-27). In addition to the classical surface (site 1 in Fig. 1D), kinetic analysis of insulin analogs has revealed a second and distinct functional surface (SerA12, IleA13, GluA17, GluB13, and LeuB17; site 2 in Fig. 1D) (18).

By analogy to growth hormone (28), binding sites 1 and 2 are proposed to function in trans within the (αβ)2 receptor dimer, creating a single high-affinity complex (20). HisB10 (dark gray in Fig. 1C, D) may also contribute to receptor binding as to a limited extent (< 4-fold) substitutions can augment (29) or impair (23) receptor binding. Mutagenesis has been precluded at some sites by impaired foldability (23), and its interpretation at other sites can be confounded by structural perturbations (30).

In the preceding article in this issue we demonstrated that stereospecific detachment of the C-terminal β-strand of the B-chain enhances the activity of insulin (31). The present study provides evidence that the detached segment inserts between receptor domains. Our approach is based on photo-mapping. To this end, a set of >20 photo-activatable insulin derivatives (each containing residue-specific para-azido-Phe substitutions at single sites) was prepared (Fig. 2A). The photo-activatable insulin derivatives were found to exhibit a range of cross-linking efficiencies. Sites of photo-cross-linking within the IR were mapped to L1, ID-N, or intervening domains. Such mapping yielded a striking pattern of alternating N- and C-terminal contacts by the conserved B-chain β-strand (residues B24-B28) accompanied by C-terminal engagement of the underlining A-chain α-helix (residues A1-A8). These results illustrate the potential of synthetic photo-mapping to decode active but cryptic protein conformations.

Insulin provides a classical paradigm for studies of protein structure and evolution (1,19). Yet ground-state structures provide an incomplete account of functional relationships. Compressed within the short sequences of the A- and B-chains lies information encoding a complex conformational life cycle, extending from nascent folding to storage and induced fit (1,2,31). This life cycle both reflects and imposes an interlocking set of evolutionary constraints. Diabetes-associated mutations in the insulin gene highlight the inter-relation of foldability, structure, activity, and the overarching threat of toxic misfolding.

**MATERIALS AND METHODS**

**Preparation of Insulin Analogs.** A summary of insulin analogs is provided as supplemental Table S1. Chemical synthesis of variant A- and B-chains containing photo-stable precursor para-azido-Phe (Pmp) was performed by manual solid-phase synthesis (4). B-chains each contained three “DKP” substitutions to prevent self-association of insulin (HisB10→Asp, ProB28→Lys, and LysB29→Pro) (2,32). In selected cases human and porcine B-chains, obtained by sulfitolysis from human insulin and porcine insulin (kindly provided by Eli Lilly and Co.), were employed in combination with synthetic A-chain analogs. Pmp substitutions were introduced in the A-chain at positions A1-A4, A8, A13, A14, A19, and A21; and in the B-chain at positions B0 (an N-terminal extension), B5, B6, B8, B10, B16, B17, B24-B27, B29, and B31 (a C-terminal extension). Individual D- and L-Pmp stereoisomers were introduced at positions A1 and B24. At position B8 (occupied in native insulin by an invariant Gly with positive
φ angle; Ref 33) only the D isomer was prepared. The synthesis and characterization of proto-probes at positions A1, A3, A21, B0, B16, B24, and B25 have previously been described (6,17,25,26,34,35).

**Biotin Labeling.** To permit detection of photo-products by NeutrAvidin™ (NAv; Pierce Chemicals, Rockford, IL), an N-terminal biotin tag was included in each analog (see supplemental Table S1). Tags were generally introduced by means of a caproyl linker either through the Nα of PheB1 (in the case of B-chain Pmp derivatives) (25), or through the Nε of D-LysA1 (in the case of the A-chain Pmp derivatives) (34); A1 and B0 analogs were biotinylated through the Nα of the N-terminal Pmp residue.

**Chain Combination.** Insulin chain combination and protein purification were performed as described (25,34). Fidelity of synthesis was verified by matrix-assisted laser-desorption/time-of-flight (MALDI-TOF) mass spectrometry (MS).

**Receptor-Binding Assays.** Activities of insulin analogs were evaluated a competitive displacement assay using a human placental membrane preparation as described (36). The percentage of tracer bound in the absence of competing ligand was <15% to avoid ligand-depletion artifacts. A summary of receptor-binding affinities of the Pmp insulin analogs is provided in supplemental Table S1. Assays were performed in duplicate (positions B29 and B31) or triplicate (the remaining 29 analogs).

**Photo-Cross-Linking Studies.** Conversion of Pmp-substituted insulin analogs to corresponding photo-activatable Pap derivatives was performed as described (4) and verified by MS. Photo-cross-linking of biotin-labeled Pap analogs to the purified receptor ectodomain (17) or IR (isoform B) (6,25) was induced by UV irradiation (4); photo-products were characterized by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot (6,17,25). Gels were probed by NAv to detect biotin-labeled photo-products and by a polyclonal antiserum recognizing the N-terminal segment of the IR α-subunit (designated IRα−N; Santa Cruz Biotech., CA). Hormone-receptor complexes were formed at a sufficiently high protein concentrations (ca. 200 nM receptor and photo-reactive insulin analog; >100-fold greater than the highest (weakest) dissociation constant) to ensure >95% receptor occupancy in each case. Assays were performed in duplicate.

**Specificity of Photo-Cross-Linking.** Control experiments verifying specificity were performed to demonstrate competition between binding of the Pap analogs and native ligands (human insulin and IGF-I). Additional control experiments demonstrated that no photo-cross-linking reactions occurred between Pap derivatives and heterologous proteins (lysozyme and immunoglobulin G) or between the photo-stable Pmp precursors and the ectodomain under same conditions (17).

**Mapping of Photo-Contacts.** Domain mapping in holoreceptor complexes was established by limited proteolysis with trypsin and chymotrypsin (6,25). Apparent molecular masses of proteolytic fragments (with and without N-linked glycosylation; Table 1) were inferred by SDS-PAGE in relation to standards. Insert-domain-specific photo-cross-linking by complementary analog PapB25-DKP-insulin (6) was mapped in parallel as a control. Key sites of proteolysis by trypsin and chymotrypsin was indicated in Figure 1B (brackets at left); structural rationales are provided in supplemental Figures S1 and S2.

**RESULTS**

Photo-probes were introduced through chemical synthesis of A- or B-chain analogs. Insulin chain combination was generally robust to substitution by the photo-stable precursor para-amino-Phe (Pmp) (yields > 25% relative to wild-type chain combination). Lower yields were obtained at A3, A21, B6, and B24; Pmp substitutions at B5 and B17 blocked disulfide pairing (gray filled circles in Fig. 2A). Substitution of ValA3 by Pmp yielded native and non-native disulfide isomers, resulting in a relative yield of 10-15%; products were distinguished by receptor-binding activity (35,37). Substitution of GlyB8 by D-Pmp enhanced the efficiency of disulfide pairing (gray filled circles in Fig. 2A). Substitution of ValA3 by Pmp yielded native and non-native disulfide isomers, resulting in a relative yield of 10-15%; products were distinguished by receptor-binding activity (35,37). Substitution of GlyB8 by D-Pmp enhanced the efficiency of disulfide pairing; substitution by L-Pmp was not attempted due to low L-specific yields (33). Pmp substitutions were likewise not attempted at sites (B12 and B15) previously associated with impaired chain combination or attenuated cellular expression of single-chain insulin analogs (23,38,39). Each photo-activatable analog
contained a biotin tag to permit identification of photo-adducts by an avidin-based reagent.

Receptor-binding affinities of Pmp analogs were highly variable, impaired in some cases and enhanced in others (range 0.2-240% relative to human insulin; see supplemental Table 1). 20 of 27 analogs exhibited relative affinities >25%. The highest affinities were conferred by PmpA8 and D-PmpB24 (> 100% relative to parent analogs), the lowest by PmpA2 and D-PmpB8 (< 1%). Such site-specific modulation of binding is in accord with trends in studies of mutant insulins (1,23,33,40), including stereospecific effects of D- and L-Pmp substitutions at B24 (see previous article in this issue (31)). The impaired binding of Glu A4 → Pmp and Tyr A14 → Pmp analogs was unanticipated by Ala scanning (23). Activities of corresponding Pap analogs were not determined. Because wild-type insulin binds with dissociation constant <0.1 nM, however, relative affinities >1% would correspond to dissociation constants <10 nM, sufficient at the protein concentrations employed (200 nM) to permit predominant formation of a photo-reactive hormone-receptor complex. Ultraviolet (UV) exposure for 20 seconds led in each case to essentially complete photolysis or cross-linking. No correlation was observed between relative affinities of Pmp analogs and extent of photo-cross-linking by corresponding Pap derivatives. Control experiments demonstrated that site-specific photo-cross-linking is competed by unmodified human insulin or (at higher protein concentration) by IGF-I.

**Photo-Cross-Linking Efficiencies**

Relative photo-cross-linking efficiencies are mapped in Figure 2B onto front and back surfaces of insulin in its classical conformation. High efficiency (>20%) was achieved by photo-probes at positions A1 (D-isomer), A3, A4, A8, A14, B16, B24 (D > L-isomer), B27, and B31 (red surface; B31 extension not shown). Of these derivatives, the highest efficiencies are observed at positions A3 (with biotin tag at the D-LysA1-ε-N), A8, and B24 (D-isomer) at which sites 30-40% of UV-irradiated complexes gave rise to a covalent photo-product (31,34,35). Inefficient photo-cross-linking (< 5%) was observed at positions A1 (L-isomer), A13, A19, A21, B0, B6, B8 (D-isomer) (black surface in Fig. 2B). Of these, the lowest efficiencies were exhibited by probes at A21 and B8 (D-isomer). Intermediate photo-cross-linking efficiencies were obtained in studies of Pap derivatives at positions A2, B10, B25, B26, and B29 (green surface in Fig. 2B). With the exception of the poorly binding PapA2 analog (see below), the extent of photo-cross-linking could not be appreciably increased by raising ligand concentrations, suggesting that relative efficiencies are intrinsic to the structure of the hormone-receptor complex rather than determined by initial binding affinities. Photo-cross-linking efficiency can in principle be influenced both by proximity of an interface and accessibility of nitrogenous groups (which favor covalent free-radical insertion) (41).

**Mapping of Photo-Products**

The modular IR structure enables mapping of photo-cross-links (i) to L1 by limited trypsin digestion and (ii) to ID-N by limited chymotrypsin digestion (see supplemental Fig. S1 and S2) (6). Following reduction by dithiothreitol (DTT), proteolytic fragments were resolved by SDS-PAGE. Bands containing a photo-adduct were identified by a biotin-specific reagent (Neutravidin; NAV). N-terminal fragments of the receptor α-subunit (and hence containing L1) were detected by Western blot using an anti-peptide antiserum (IRα-N); C-terminal fragments (containing ID-N) were characterized by B25 photo-cross-linking, previously shown to contact the C-terminal 15-residue peptide (4). The latter proteolytic signature has been confirmed using engineered midi-receptors (35).

**L1 Contact Sites.** Tryptic mapping yields a unique 31-kDa cross-linked fragment in L1-cross-linked B-chain complexes. The molecular mass of this fragment (excluding glycosylation and the tethered B-chain) indicates that it contains fewer than 170 amino acids and so consists of the L1 domain (residue 1-158) and flanking trypptic site. The structure of the ectodomain exhibits exposed basic side chains at the outer L1-CR junction (see supplemental Figure S1) as illustrated in the preceding paper in this issue (31). The following photo-probes cross-link to L1: B16, B24-L, B24-D, and B26. L1 photo-adducts were not observed in studies of PapA25 or A-chain derivatives.
Limited chymotryptic digestion yields distinct N- and C-terminal signatures. N-terminal contacts are characterized by a fragment spanning L1 and part of CR (apparent glycosylated molecular mass 50 kD and deglycosylated mass 31 kD, including insulin-chain adduct). Observation of this fragment corroborated tryptic assignment of the above L1 contacts; the outer surface contains solvent-exposed Phe side chains at an appropriate position (CR residue 256 and 258; see supplemental Figure S2). Further digestion of the B24-cross-linked complex yields a light 19 kD α band recognized by IRα-N; its estimated mass (<11 kD exclusive of the B-chain adduct and two N-linked carbohydrates) suggests that the site of cross-linking is likely to occur within the first 100 residues of the L1 domain in accord with sites of Ala substitution in L1 that markedly impair insulin binding (8,42).

**ID-N Contacts.** Photo-cross-linking of PapB25 has previously been shown to map to the extreme C-terminal peptide of the IR α subunit (4) and so provides a standard. Unlike L1-specific contacts, PapB25 photo-cross-linking is characterized by (a) a C-terminal fragment of apparent molecular mass 34 kD (glycosylated) and 23 kD (deglycosylated), which on further digestion yields (b) a C-terminal fragment of 20 kD (17 kD on deglycosylation) (6). The larger fragment encompasses Fn1-N and ID-N (residues 590-731 in the B-receptor isoform). Its difference in apparent mass on deglycosylation (11 kD) represents three N-linked carbohydrates (positions 606, 624, and 671) (43). The structure of the ectodomain suggests possible cleavage at exposed Trp side chains at positions 551 and 559 (see supplemental Figure S2). The smaller fragment contains a single N-linked carbohydrate and so begins after residue 624. Its mass is consistent with the sum of the B-chain adduct (3.7 kD) and a fragment containing ID-N (residues 638-731; glycosylated mass 16 kD) and at most a small number of residues from the Fn2a junction.

Mapping studies using limited chymotryptic digestion demonstrated that the following ID-N contacts: A1-D, A2, A3, A4, A8, A14, B25, and B27. Probes at B29 and B31 were observed to photo-cross-link to both L1 and ID-N (and possibly to intervening domains), presumably due to flexibility in the hormone-receptor complex. The B10 probe exhibits neither L1-CR nor Fn2a-ID-N signatures and so must have intervening contact sites. Adducts A1-L, A13, A19, A21, B0, B6, and B8 were not mapped due to their low photo-cross-linking efficiencies.

**Induced Fit of Insulin**

To investigate the active conformation of insulin, we focused in turn on the C-terminal β-strand of the B-chain (residues B24-B28) and N-terminal α-helix of the A-chain (A1-A8).

**B-Chain β-Strand.** Photo-cross-linking by photo-probes B24-30 (extended to B31) exhibited moderate to high efficiencies (top box in Fig. 3A); middle and lower boxes provide loading controls as probed by IRα-N (Fig. 3A, middle) and NAv (Fig. 3A, bottom). In this figure each site is represented by four lanes; the first three provide negative controls (respectively containing indicated Pap derivative either unirradiated in absence of receptor, irradiated in absence of receptor, or unirradiated in presence of receptor), and the fourth lane contains the products of the photo-cross-linking reaction. Covalent hormone-receptor complexes are thus observed only in lanes 4, 8, 12, 16, 20, 24, and 28 (Fig. 3A, top).

Limited chymotryptic digestion of photo-cross-linked B24-B27 products (Fig. 3B, left) yielded distinct NAv-detected N- and C-terminal patterns whereas detection by IRα-N yielded similar N-terminal patterns as a loading control and verification of digestion (Fig. 3B, right). PapB24 and PapB26 derivatives gave rise to an L1-CR* pattern (band α(N)-B in lanes (2,3) and (5,6), respectively, in Fig. 3B). By contrast, B25 and B27 photo-cross-links gave rise to similar C-terminal patterns (band α(C)-B in lanes (8,9) and (11,12), respectively, in Fig. 3B). Thus, the odd-numbered probes did not photo-cross-link to the 50 kD N-terminal chymotryptic fragment characteristic of even-numbered probes. Further chymotryptic digestion of B25 and B27 photo- adducts in each case yielded glycosylated 20 kD signature fragments (α(C)*-B band in lanes 14 and 17 in Fig. 3C) and deglycosylated 17 kD fragments (band ID-N*/B in lanes 15 and 18 in Fig. 3C). In these reactions distinct glycosylated and deglycosylated N-terminal fragments were detected by anti-serum IRα-N (Fig. 3D). The tryptic signature of PapB26-derived photo-adducts was similar to that of B24 (data not shown).
A-Chain α-Helix. Photo-probes at residues A1 (D chirality), A2, A3, A4, and A8 cross-link to the IR with varying efficiencies (Fig. 4A, top); negative and loading controls are as in Figure 3A. Sites of photo-cross-linking thus include both the polar (GluA4 and ThrA8) and nonpolar (IleA2 and ValA3) faces of the amphipathic A1-A8 α-helix. Chymotryptic mapping of covalent complexes A1 (D isomer), A2, A3, A4, and A8 each yielded 34 and 20 kD glycosylated fragments similar to those characteristic of the B25-cross-linked complex (Fig. 5). (Their slightly increased SDS-PAGE mobilities are consistent with the smaller size of the A-chain (21 residues) relative to the B-chain (30 residues).) None of the A-chain cross-linked complexes gave rise to the N-terminal 50 kD adduct characteristic of the PapB24-mediated photo-product between the B-chain and an L1-CT fragment (left-hand three lanes in each panel of Fig. 5). Fine mapping of PapA3-mediated photo-contacts to the same C-terminal peptide (αCT residues 703-719) as contacted by PapB25 has been demonstrated in engineered mini-receptors (35).

Photo-cross-linking by PapA2 is of particular interest in relation to classical packing of IleA2 within the hydrophobic core (see supplemental Fig. S3). Increasing the concentration of the PapA2 derivative from 200 to 600 nM (at a constant receptor concentration of 200 nM) enhanced the extent of photo-cross-linking (Fig. 4B), suggesting that its intrinsic photo-cross-linking efficiency is high but limited by partial receptor occupancy. The specificity of such photo-cross-linking was demonstrated by competition with unmodified insulin or IGF-I (Fig. 4C). Structural interpretation of A2 photo-cross-linking is uncertain. Due to the small size of insulin, molecular modeling suggests that analogous packing of PapA2 with the hydrophobic core could allow protrusion of the azido group from the protein surface. Direct engagement of IleA2 is nonetheless supported by mutagenesis (34,40,44).

**DISCUSSION**

Classical models of insulin binding (19,21,45) were based on crystal structures of zinc insulin hexamers (46,47). Such hexamers represent the storage form of the hormone within the glucose-regulated secretory vesicles of pancreatic β-cells (48). Because insulin functions in the bloodstream as a monomer, however, its active conformation has long been the subject of speculation. Although the conformation of an engineered insulin monomer in solution closely resembles a crystallographic protomer (2,3), its flexibility has left open the possibility of induced fit (49). In the previous article in this issue we have shown that the activity of insulin can be enhanced by chiral destabilization of the B-chain (31). Here, we have exploited chemical protein synthesis to obtain a photo-map of the receptor-binding surface. The pattern of photo-cross-linking provides a signature of the hormone’s active conformation.

**Induced Fit Extends Receptor-Binding Surface**

Evidence for induced fit within the B-chain has been provided by the very low activities of native-like analogs containing tethers between the C-terminus of the B-chain and N-terminus of the A-chain (50-53). Such impairment suggests that the tethered sites separate on receptor binding (Fig. 6A) (54). The detachment model thus posits that C-terminal β-strand of the B-chain detaches from the α-helical core of the hormone (54-56). This strand is anchored to the core by an invariant aromatic side chain (PheB24). Strikingly, as shown in the preceding article in this issue, D-amino-acid substitutions at this site enhance activity (55,57) by segmental unfolding (31). Together, these observations suggest that the C-terminal β-strand adopts an inhibitory conformation within classical structures of insulin. A conformational switch at B24 leading to detachment of the C-terminal segment would also rationalize the high activities of foreshortened analogs (58-61).

Ala scanning mutagenesis of the IR has demonstrated that the N- and C-terminal domains of the α-subunit (L1 and ID-N) contain the major hormone-binding sites (8,42,62). Remarkably, photo-probes at positions B24-B27 exhibit an alternating pattern of L1 and ID-N contacts. Because odd- and even-numbered side chains in a β-strand project in opposite directions (Fig. 6B), these observations provide evidence that this segment inserts as a β-strand between L1 and ID-N contacts. Because odd- and even-numbered side chains in a β-strand project in opposite directions (Fig. 6B), these observations provide evidence that this segment inserts as a β-strand between L1 and ID-N domains. Detachment of the β-strand from the hydrophobic core of insulin would in turn be
expected to expose the conserved inner surface of the A1-A8 α-helix (IleA2 and ValA3) (Fig. 6C, D).

Although largely inaccessible in the unbound hormone, the side chains of IleA2 and ValA3 are critical to biological activity (40,63). Direct contacts between these aliphatic side chains and the receptor would rationalize the marked effects of even subtle substitutions. An example is provided by inversion of Cβ chirality at A2 (yielding allo-IleA2 analogs): this modification is readily accommodated within a native-like core but impairs receptor binding by 50 fold (34,44). Similarly, substitution of ValA3 by Leu is structurally conservative (34) but associated with a 500-fold decrease in receptor binding (40,64).

The present study provides evidence that the A1-A8 α-helix contacts IN-D at multiple points, including at A2 and A3.

A classical model for induced fit has been provided by the allosteric transition among T6, T3Rf3, and R6 hexamers (Fig. 8A) (65). In this conformational equilibrium the N-terminal segment (residues B1-8) undergoes a dramatic change in secondary structure (66,67): extended in the T state (green) and α-helical in the R state (blue). The relationship between the TR transition and the mechanism of receptor binding is unclear as activity and allostery may be uncoupled by mutagenesis (68). In addition, the N-terminal five residues of the B-chain may be deleted without significant change in activity (69). Evidence for an R-state-related change in conformation at GlyB8 has nonetheless been provided by chiral mutagenesis. The glycine (red in Fig. 8B) exhibits a positive φ dihedral angle (like a D-amino acid) in T-states (including T-like monomers in solution) but a negative φ angle (like an L-amino acid) in R-states. D-Amino-acid substitutions at B8 (replacing the pro-D Hα atomic; arrow in Fig. 8B) stabilize the T-state but impair receptor binding (33,70). The low efficiency of photo-cross-linking by D-Pap at this site suggests that such loss of activity is due to chiral impairment of a B8-specific conformational change on receptor binding. By contrast, a frustrated L-SerB8 analog retains high affinity despite marked thermodynamic instability (70). We imagine that the B8-related conformational switch (red in Fig. 8C) alters the topography of cystine A7-B7 (gold in Fig. 8B) and (like the B24-related switch; tawny in Fig. 8C) extends the nonpolar receptor-binding surface of insulin. Molecular modeling suggests that displacement of the N-terminal arm of the B-chain from its T-state conformation would expose LeuB6 and part of the A-chain (IleA10 and LeuA13).

Limitations of Photo-Scanning

The above interpretations assume that photocross-linking provides a valid probe for native contacts within the hormone-receptor complex. A general limitation is nonetheless posed by possible structural perturbations associated with the photo-probe. Although Pap is among the smallest of available azido reagents, its substitution at sites containing smaller amino acids (such as Gly, Leu, Val, and Thr) or at sites containing formal charges (Glu and Lys) can in principle lead to local or nonlocal conformational changes in the hormone. Fortuitously, the sequence and structure of insulin facilitate photo-scanning. Key sites contain either aromatic side chains similar to shape and size to Pap (TyrA14, TyrA19, PheB1, TyrB16, PheB24, PheB25, and TyrB26) or non-aromatic side chains positioned such that a Pap substituent would be expected to project into solvent (D-PapA1, PapA4, PapA8, PapB0, D-PapB8, PapB28, and PapB29). The conclusions of the present study are thus likely to be robust to probe-induced distortions. Structural perturbations are nonetheless likely at A2 and A3. The side chain of IleA2 packs against the aromatic ring of TyrA19 within the core whereas ValA3 projects within an inter-chain crevice adjoining TyrB26. Crystal structures of Pmp or related photo-stable analogs have not been determined. Despite the uncertain structural effects of Pap at these sites, prior studies of insulin analogs support the direct engagement of IleA2 and ValA3 at the receptor interface (30,35,44,63).

Relationship to the Structure of the Receptor

The photo-active surface of insulin spans both its front and back surfaces. Such circumferential binding is consistent with its proposed mode of binding within the crux of the inverted-V-shaped dimeric ectodomain (Fig. 7A, B) (7). How insulin binds within the interior of the ectodomain is unclear. Modeling is limited by the absence of continuous electron density spanning the insert domain, including the C-terminal ID-N-derived
peptide contacted by Pap\textsubscript{B25} and Pap\textsuperscript{A3} (\(\alpha\)CT; residues 716-731 in the B isoform). An intriguing but presently uninterpretable feature of the crystal structure is a low-resolution tube of electron density adjoining the hormone-binding face of L1 (Fig. 7C). Alternation of B24-B27 photo-contacts suggests that this anomalous feature represents \(\alpha\)CT. It is possible that on binding of insulin this region of the receptor becomes better organized with changes in the relative orientation of L1 and \(\alpha\)CT.

Overall features of the ectodomain structure, supported by complementation analysis between mutations in L1 and ID-N (71), suggest that the tail of one subunit is near the head (L1) of the other (7). Such head-to-tail dimerization would imply that the B-chain \(\beta\)-strand inserts in \textit{trans} between \(\alpha\)-subunits. The proposed bivalency of insulin may regulate the orientation of subunits in the \(\alpha_2\beta_2\) holoreceptor and in turn initiate transmembrane signaling (20,72). Bridging L1/\(\alpha\)CT contacts are unrelated to binding site 2 as defined in kinetic studies of insulin analogs by De Meyts and colleagues (18).

Contacts between insulin and the IR are likely to extend beyond L1 and ID-N. Although the contribution of such contacts to binding affinity may be smaller than that of classical contacts (as defined by Ala scanning mutagenesis (23,24)), they may be integral to the mechanism of signal transduction. The present mapping studies indicate that the B10 photo-probe contacts neither L1-CR* nor Fn2-ID-N. In addition, previous studies of a B1 photo-probe by Brandenburg and coworkers identified a contact in a fragment spanning parts of L2 and Fn1 (residues 390-488) (16). Evidence for potential contacts in Fn1 is provided by a recent mutagenesis study (10); these contacts may relate to site 2 in insulin (residues A12, A13, B10, B14, and B17), which influences the kinetic properties of the insulin-receptor complex disproportionately to effects on affinity (18,73). The head-to-tail architecture of the ectodomain implies that any such contacts (like those to ID-N) would be in \textit{trans} with L1.

\textbf{Evolution of Insulin and Human Genetics}

Most of the photo-cross-linking sites identified in the present study are broadly conserved among vertebrate insulins and IGFs. Such conservation is likely to reflect independent evolutionary constraints simultaneously imposed by structural requirements of protein folding in the endoplasmic reticulum (ER), subcellular trafficking and prohormone processing leading to self-assembly in the secretor granule, and receptor binding at target tissues (74). A given residue (as illustrated by Phe\textsuperscript{B24} in the previous article in this issue (31)) may play distinct roles at each stage of biosynthesis and signaling. The presence of interlocking constraints implies that the fitness landscape of insulin is sharply peaked.

Advances in human genetics have identified several monogenic forms of diabetes mellitus (DM) due to mutations in the insulin gene. Disease-associated substitutions are predicted in each region of the hormone and its precursors (the signal peptide, A- and B-chains, dibasic processing sites, and connecting domain) (75-81). The classical insulinopathies (Leu\textsuperscript{A3}, Ser\textsuperscript{B24}, and Leu\textsuperscript{B25}; gray circles in Fig. 9A) occur at sites of efficient receptor photo-cross-linking (6,17); these substitutions leading to mutant hyperinsulinemia and adult-onset DM of variable penetrance (75). Substitution of His\textsuperscript{B10} by Asp by contrast enhances in vitro activity but within the \(\beta\)-cell leads to mistrafficking of the mutant proinsulin to a constitutive (i.e., non-glucose-regulated) granule, causing adult-onset DM with mutant hyperproinsulinemia. The majority of mutations mapping within the A- and B-domains are associated with permanent-onset neonatal DM, presumably due to misfolding of the variant proinsulin (78-81). Of these, most are due to the addition or removal of a cysteine, unbalancing disulfide pairing. One human mutation (Cys\textsuperscript{A7}→Tyr) corresponds to a rodent model of DM associated with endoreticular stress (the Akita mouse (82)).

Clinical mutations associated with neonatal-onset DM – and hence with a block to disulfide pairing – also occur at or near sites of induced fit: either within the N-terminal segment of the B-chain (residues B5, B6, and B8) or its C-terminal segment (B22 and B23). We speculate that at these sites kinetic determinants of foldability in the ER are at odds with conformational requirements of receptor binding, a conflict resolved by induced fit (31). Evidence for this conflict and its resolution have been provided by
comparative analysis of D- and L-amino-acid substitutions at conserved “hotspot-associated” glycines flanking both N- and C-terminal segments (Gly$^B_{B8}$ and Gly$^B_{B23}$; black spheres in B-chain in Fig. 9B). Chiral mutagenesis has demonstrated that native-like positive $\phi$ dihedral angles facilitate pairing of neighboring cysteines (B7-A7 and B20-A19, respectively) but are dispensable for receptor binding (33,70,83). Whereas D-substitutions at B8 enhance the efficiency of disulfide pairing in peptide models (33,70,83), the inactivity of such analogs (above) suggests that nascent structural relationships essential for foldability in the ER and prominent features of storage hexamers may be reorganized on receptor binding.

Impairment of disulfide pairing in nascent proinsulin can in principle be severe or mild, depending on the site of mutation and properties of the substituted side chain. Indeed, whereas neonatal-onset DM represents a block to folding, analogous mutations have been identified in patients presenting with auto-antibody negative Type I diabetes mellitus in the second decade of life (78-81) (maturity-onset diabetes of the young; MODY). One such mutation in Arg$^B_{B22} \rightarrow $Gln, which alters a solvent-exposed site in the B21-B24 $\beta$-turn (81). Although this mutation highlights the general importance of the C-terminal $\beta$-turn, Arg$^B_{B22}$ is not needed for receptor binding$^{23}$, and its specific contribution to folding is not apparent. The wild-type side chain is not well ordered in crystals or in solution (1-3); its guanidinium group does not form hydrogen bonds or salt bridges within the protomer.

We speculate that globular proteins may contain cryptic determinants of foldability whose transient roles in folding are not apparent in the native state. It is possible, for example, that the conserved positive charge of Arg$^B_{B22}$ participates in stabilization of a thiolate intermediate in the mechanism of disulfide pairing and is dispensable once folding is achieved. Another example of a mild folding defect may be provided by the classical variant Ser$^B_{B24}$-insulin, which is expressed, processed, and secreted (75). As a seeming paradox, this mutant insulin (unlike Leu$^A_3$- and Leu$^B_{B25}$-insulins) exhibits reduced but nonnegligible receptor-binding activity, sufficient to avoid a diabetic phenotype (77). Our findings that B24 substitutions partially impair insulin chain combination and predispose to aggregation-coupled misfolding (preceding paper in this issue) (31) suggests that mild chronic ER stress may contribute to this associated adult-onset syndrome: its variable genetic penetrance may be regarded as the hallmark of a gene that modifies the survival of $\beta$-cells as a multigenic trait.

Concluding Remarks. The present photo-map of insulin provides evidence that its receptor-binding surface is extended by conformational changes in the B-chain. Photo-cross-linking sites span the nonpolar surfaces of both A- and B-chains. The C-terminal $\beta$-strand of the B-chain exhibits a striking alternation of photo-contacts between the C-terminal L1 domain of the receptor $\alpha$-subunit (contacted by probes at B24 and B26) and its N-terminal ID-N domains (contacted by probes at B25 and B27). This even-odd alternation corresponds to the canonical structure of a $\beta$-strand in which successive side chains project in opposite directions. We propose that insertion of the C-terminal segment of the B-chain between these receptor domains (likely to occur in trans) represents the first step in a complex choreography of conformational changes in the holoreceptor leading to transmembrane activation of the tyrosine kinase.

The classical structure of insulin represents an inactive conformation that mediates native self-assembly in the $\beta$-cell. Whereas its structural relationships predict nascent interactions pertinent to the mechanism of disulfide pairing in biosynthesis, reorganization of the B-chain on receptor binding is likely to involve both its N- and C-terminal segments. We envisage that conserved side chains play distinct roles in folding, assembly, and receptor binding. It would thus be of future interest to investigate the evolution of the insulin gene in relation to such interlocking structural constraints. Such investigation will require structural elucidation of the conformational life cycle of insulin from folding to receptor binding.

Supplemental information is available on the J. Biol. Chem. web site: three figures illustrating photo-mapping rationale and table of receptor-binding affinities.
**FOOTNOTES**

*Acknowledgements. We thank Q. X. Hua for collaborative interactions, W. Jia for assistance with the figures, B. Li for assistance with receptor purification, N. F. Phillips for advice, and P. Arvan, S. Chan, P. DeMeyts, G. G. Dodson, M. Lawrence, D. F. Steiner, and C. Ward for discussion. We are grateful to S. Chan and D. F. Steiner for communication of results prior to publication. This work was supported by National Institutes of Health grants DK065980 (to J.W.), DK56673 (P.G.K.), and DK040949 (M.A.W.).

1Abbreviations. αCT, C-terminal peptide of IR α-subunit comprising residues 704-718; DKP-insulin, insulin analog containing three substitutions in B chain (Asp^{B10}, Lys^{B28} and Pro^{B29}); DTT, dithiothreitol; ER, endoplasmic reticulum; ID, insert domain of IR; IR, insulin receptor; kDa, kilo-dalton of mass; L1, large β-helical domain 1 of IR (residues 1-158); MS, mass spectrometry; NAv, NeutrAvidin (reagent); PAGE, polyacrylamide gel electrophoresis; Pap, para-azido-Phe; Pmp, para-amino-Phe; SDS, sodium dodecyl sulfate. rp-HPLC, reverse-phase HPLC; and UV, ultra-violet; Amino acids are designated by standard one- and three-letter codes.

2Pap is small, rigid, and readily accommodated at diverse sites. Its systematic incorporation by total chemical synthesis thus circumvents two key limitation of conventional azido modification: (a) non-specific cross-linking by flexible swinging arms and (b) restricted potential sites of derivitization.

3Native-like crystal structures of L-Trp^{A1} and D-Trp^{A1} insulin analogs (84) suggest that L- and D-Pap^{A1} derivatives can be accommodated without transmitted perturbations. That Pap^{A3} may likewise be accommodated is suggested by the ready crystallization of Tyr^{A3}-insulin as a canonical zinc hexamer under conditions similar to those used in crystallization of wild-type insulin (Z.-L. Wan and M.A.W., unpublished results).
REFERENCES

1. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., and Reynolds, C. D. (1988) Philos. Trans. R. Soc. Lond. B Biol. Sci. 319, 369-456

2. Hua, Q. X., Hu, S. Q., Frank, B. H., Jia, W., Chu, Y. C., Wang, S. H., Burke, G. T., Katsoyannis, P. G., and Weiss, M. A. (1996) J. Mol. Biol. 264, 390-403

3. Olsen, H. B., Ludvigsen, S., and Kaarsholm, N. C. (1996) Biochemistry 35, 8836-8845

4. Kurose, T., Pashmforoush, M., Yoshimasa, Y., Carroll, R., Schwartz, G. P., Burke, G. T., Katsoyannis, P. G., and Steiner, D. F. (1994) J. Biol. Chem. 269, 29190-29197

5. De Meyts, P., and Whittaker, J. (2002) Nat. Rev. Drug Discov. 1, 769-783

6. Xu, B., Hu, S. Q., Chu, Y. C., Huang, K., Nakagawa, S. H., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2004) Biochemistry 43, 8356-8372

7. McKern, N. M., Lawrence, M. C., Streltsov, V. A., Lou, M. Z., Adams, T. E., Lovrecz, G. O., Elleman, T. C., Richards, K. M., Bentley, J. D., Pilling, P. A., Hoyne, P. A., Cartledge, K. A., Pham, T. M., Lewis, J. L., Sankovich, S. E., Stoichevska, V., Da Silva, E., Robinson, C. P., Frenkel, M. J., Sparrow, L. G., Fernley, R. T., Epst, V. C., and Ward, C. W. (2006) Nature 443, 218-221

8. Whittaker, J., Sorensen, H., Gadsboll, V. L., and Hinrichsen, J. (2002) J. Biol. Chem. 277, 47380-47384

9. Ward, C., Lawrence, M. C., Streltsov, V., Garrett, T., McKern, N., Lou, M. Z., Lovrecz, G., and Adams, T. (2008) Acta Physiol. (Oxf.) 192, 3-9

10. Whittaker, L. J., Hao, C., Fu, W., and Whittaker, J. (2008) Biochemistry Epub ahead of print

11. Andersen, A. S., Kjeldsen, T., Wiberg, F. C., Vissing, H., Schaffer, L., Rasmussen, J. S., De Meyts, P., and Moller, N. P. (1992) J. Biol. Chem. 267, 13681-13686

12. Kjeldsen, T., Andersen, A. S., Wiberg, F. C., Rasmussen, J. S., Schaffer, L., Balschmidt, P., Moller, K. B., and Moller, N. P. (1991) Proc. Natl. Acad. Sci. USA 88, 4404-4408

13. Schumacher, R., Soos, M. A., Schlessinger, J., Brandenburg, D., Siddle, K., and Ullrich, A. (1993) J. Biol. Chem. 268, 1087-1094

14. Yip, C. C., Hsu, H., Patel, R. G., Hawley, D. M., Maddux, B. A., and Goldfine, I. D. (1988) Biochem. Biophys. Res. Commun. 157, 321-329

15. Wedekind, F., Baer-Pontzen, K., Bala-Mohan, S., Choli, D., Zahn, H., and Brandenburg, D. (1989) Biol. Chem. 370, 251-258

16. Fabry, M., Schaefer, E., Ellis, L., Kojo, E., Fahrenholz, F., and Brandenburg, D. (1992) J. Biol. Chem. 267, 8950-8956

17. Xu, B., Hu, S. Q., Chu, Y. C., Wang, S., Wang, R. Y., Nakagawa, S. H., Katsoyannis, P. G., and Weiss, M. A. (2004) Diabetes 53, 1599-1602

18. De Meyts, P. (2008) Trends Biochem. Sci. 33, 376-384

19. Pullen, R. A., Lindsay, D. G., Wood, S. P., Tickle, I. J., Blundell, T. L., Wollmer, A., Krail, G., Brandenburg, D., Zahn, H., Gliemann, J., and Gammeltoft, S. (1976) Nature 259, 369-373

20. De Meyts, P. (1994) Diabetologia 37, S135-S148

21. Liang, D. C., Chang, W. R., and Wan, Z. L. (1994) Biophys. Chem. 50, 63-71

22. Cunningham, B. C., and Wells, J. A. (1993) J. Mol. Biol. 234, 554-563

23. Kristensen, C., Kjeldsen, T., Wiberg, F. C., Schaffer, L., Hach, M., Havelund, S., Bass, J., Steiner, D. F., and Andersen, A. S. (1997) J. Biol. Chem. 272, 12978-12983

24. Clackson, T., and Wells, J. A. (1995) Science 267, 383-386

25. Huang, K., Xu, B., Hu, S. Q., Chu, Y. C., Hua, Q. X., Qu, Y., Li, B., Wang, S., Wang, R. Y., Nakagawa, S. H., Theede, A. M., Whittaker, J., De Meyts, P., Katsoyannis, P. G., and Weiss, M. A. (2004) J. Mol. Biol. 341, 529-550

26. Wan, Z., Xu, B., Huang, K., Chu, Y. C., Li, B., Nakagawa, S. H., Qu, Y., Hu, S. Q., Katsoyannis, P. G., and Weiss, M. A. (2004) Biochemistry 43, 16119-16133
27. Glendorf, T., Sorensen, A. R., Nishimura, E., Pettersson, I., and Kjeldsen, T. (2008) *Biochemistry* **47**, 4743-4751
28. Cunningham, B. C., Ultsch, M., De Vos, A. M., Mulkerrin, M. G., Clauser, K. R., and Wells, J. A. (1991) *Science* **254**, 821-825
29. Schwartz, G. P., Burke, G. T., and Katsoyannis, P. G. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6408-6411
30. Nakagawa, S. H., and Tager, H. S. (1991) *J. Biol. Chem.* **266**, 11502-11509
31. Hua, Q. X., Xu, B., Huang, K., Hu, S. Q., Nakagawa, S., Jia, W., Wang, S., Whittaker, J., Katsoyannis, P. G., and Weiss, M. A. (2008). *J. Biol. Chem.* submitted
32. Weiss, M. A., Hua, Q. X., Lynch, C. S., Frank, B. H., and Shoelson, S. E. (1991) *Biochemistry* **30**, 7373-7389
33. Nakagawa, S. H., Zhao, M., Hua, Q. X., Hu, S. Q., Wan, Z. L., Jia, W., and Weiss, M. A. (2005) *Biochemistry* **44**, 4984-4999
34. Wan, Z. L., Huang, K., Xu, B., Hu, S. Q., Wang, S., Chu, Y. C., Katsoyannis, P. G., and Weiss, M. A. (2005) *Biochemistry* **44**, 5000-5016
35. Huang, K., Chan, S. J., Hua, Q. X., Chu, Y. C., Wang, R., Klaproth, B., Jia, W., Whittaker, J., De Meyts, P., Nakagawa, S. H., Steiner, D. F., Katsoyannis, P. G., and Weiss, M. A. (2007) *J. Biol. Chem.* **282**, 35337-35349
36. Weiss, M. A., Hua, Q. X., Jia, W., Nakagawa, S. H., Chu, Y. C., Hu, S. Q., and Katsoyannis, P. G. (2001) *J. Biol. Chem.* **276**, 40018-40024
37. Hua, Q. X., Gozani, S. N., Chance, R. E., Hoffmann, J. A., Frank, B. H., and Weiss, M. A. (1995) *Nat. Struct. Biol.* **2**, 129-138
38. Hu, S. Q., Burke, G. T., Schwartz, G. P., Ferderigos, N., Ross, J. B., and Katsoyannis, P. G. (1993) *Biochemistry* **32**, 2631-2635
39. Hua, Q. X., Liu, M., Hu, S. Q., Jia, W., Arvan, P., and Weiss, M. A. (2006) *J. Biol. Chem.* **281**, 24889-24899
40. Nakagawa, S. H., and Tager, H. S. (1992) *Biochemistry* **31**, 3204-3214
41. Eberle, A. N., and de Graan, P. N. E. (1985) *Methods Enzymol.* **109**, 129-157
42. Mynarcik, D. C., Yu, G. Q., and Whittaker, J. (1996) *J. Biol. Chem.* **271**, 2439-2442
43. Sparrow, L. G., Lawrence, M. C., Gorman, J. J., Strike, P. M., Robinson, C. P., McKern, N., and Ward, C. (2008) *Proteins* **71**, 426-439
44. Xu, B., Hua, Q. X., Nakagawa, S. H., Jia, W., Chu, Y. C., Katsoyannis, P. G., and Weiss, M. A. (2002) *J. Mol. Biol.* **316**, 435-441
45. Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., and Mercola, D. A. (1972) *Diabetes* **21**, 492-505
46. Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., Mercola, D. A., and Vijayan, M. (1971) *Nature* **231**, 506-511
47. The Peking Insulin Structure Group. (1971) *Peking Rev.* **40**, 11-16
48. Dodson, G., and Steiner, D. (1998) *Curr. Opin. Struct. Biol.* **8**, 189-194
49. Hua, Q. X., Ladbury, J. E., and Weiss, M. A. (1993) *Biochemistry* **32**, 189-194
50. Markussen, J., Jorgensen, K. H., Sorensen, A. R., and Thim, L. (1985) *Int. J. Pept. Protein Res.* **26**, 70-77
51. Nakagawa, S. H., and Tager, H. S. (1989) *J. Biol. Chem.* **264**, 272-279
52. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Bing, X., and Markussen, J. (1991) *J. Mol. Biol.* **220**, 425-433
53. Huang, Y., Liang, Z., and Feng, Y. (2001) *Sci. China* **44**, 593-600
54. Hua, Q. X., Shoelson, S. E., Kochoyan, M., and Weiss, M. A. (1991) *Nature* **354**, 238-241
55. Kobayashi, M., Ohgaku, S., Iwasaki, M., Maegawa, H., Shigeta, Y., and Inouye, K. (1982) *Biochem. Biophys. Res. Commun.* **107**, 329-336
56. Mirmira, R. G., Nakagawa, S. H., and Tager, H. S. (1991) *J. Biol. Chem.* **266**, 1428-1436
57. Mirmira, R. G., and Tager, H. S. (1989) *J. Biol. Chem.* **264**, 6349-6354
58. Schwartz, G. P., Burke, G. T., and Katsoyannis, P. G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 458-461
59. Lenz, V., Gattner, H. G., Sievert, D., Wollmer, A., Engels, M., and Hocker, H. (1991) *Biol. Chem. Hoppe Seyler* **372**, 495-504

60. Kurapkat, G., Siedentop, M., Gattner, H. G., Hagelstein, M., Brandenburg, D., Grotzinger, J., and Wollmer, A. (1999) *Protein Sci.* **8**, 499-508

61. Zakova, L., Barth, T., Jiracek, J., Barthova, J., and Zorad, S. (2004) *Biochemistry* **43**, 2323-2331

62. Williams, P. F., Mynarcik, D. C., Yu, G. Q., and Whittaker, J. (1995) *J. Biol. Chem.* **270**, 3012-3016

63. Kitagawa, K., Ogawa, H., Burke, G. T., Chanley, J. D., and Katsoyannis, P. G. (1984) *Biochemistry* **23**, 1405-1413

64. Kobayashi, M., Takata, Y., Ishibashi, O., Sasaoka, T., Iwasaki, T. M., Shigeta, Y., and Inouye, K. (1986) *Biochem. Biophys. Res. Commun.* **137**, 250-257

65. Brader, M. L., and Dunn, M. F. (1991) *Trends Biochem. Sci.* **16**, 341-345

66. Bentley, G., Dodson, E., Dodson, G., Hodgkin, D., and Merciola, D. (1976) *Nature* **261**, 166-168

67. Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, G. D., Sparks, C., and Swenson, D. (1989) *Nature* **338**, 594-596

68. Wan, Z., Huang, K., Whittaker, J., and Weiss, M. A. (2008) *J. Biol. Chem.* **283**, 21198-21210

69. Schwartz, G., and Katsoyannis, P. G. (1978) *Biochemistry* **17**, 4550-4556

70. Hua, Q. X., Nakagawa, S. H., Hu, S. Q., Jia, W., Wang, S., and Weiss, M. A. (2006) *J. Biol. Chem.* **281**, 24900-24909

71. Chan, S. J., Nakagawa, S., and Steiner, D. F. (2007) *J. Biol. Chem.* **282**, 13754-13758

72. Schäffer, L. (1994) *Eur. J. Biochem.* **221**, 1127-1132

73. De Meyts, P. (2004) *Bioessays* **26**, 1351-1362

74. Steiner, D. F., and Chan, S. J. (1988) *Horm. Metab. Res.* **20**, 443-444

75. Shoelson, S., Haneda, M., Blix, P., Nanjo, A., Sanke, T., Inouye, K., Steiner, D., Rubenstein, A., and Tager, H. (1983) *Nature* **302**, 540-543

76. Carroll, R. J., Hammer, R. E., Chan, S. J., Swift, H. H., Rubenstein, A. H., and Steiner, D. F. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8943-8947

77. Steiner, D. F., Tager, H. S., Chan, S. J., Nanjo, K., Sanke, T., and Rubenstein, A. H. (1990) *Diabetes Care* **13**, 600-609

78. Stoy, J., Edghill, E. L., Flanagan, S. E., Ye, H., Paz, V. P., Pluzhnikov, A., Below, J. E., Hayes, M. G., Cox, N. J., Lipkind, G. M., Lipton, R. B., Greeley, S. A., Patch, A. M., Ellard, S., Steiner, D. F., Hattersley, A. T., Philipson, L. H., and Bell, G. I. (2007) *Proc. Natl. Acad. Sci. USA* **104**, 15040-15044

79. Colombo, C., Porzio, O., Liu, M., Massa, O., Vasta, M., Salardi, S., Beccaria, L., Monciotti, C., Toni, S., Pedersen, O., Hansen, T., Federici, L., Pesavento, R., Cadario, F., Federici, G., Ghirri, P., Arvan, P., Iafusco, D., Barbetti, F., and Early Onset Diabetes Study Group of the Italian Society of Pediatric Endocrinology and Diabetes. (2008) *J. Clin. Invest.* **109**, 2148-2156

80. Edghill, E. L., Flanagan, S. E., Patch, A. M., Boustred, C., Parrish, A., Shields, B., Shepherd, M. H., Hussain, K., Kapoor, R. R., Malecki, M., MacDonald, M. J., Stoy, J., Steiner, D. F., Philipson, L. H., Bell, G. I., Neonatal Diabetes International Collaborative Group, Hattersley, A. T., and Ellard, S. (2008) *Diabetes* **57**, 1034-1042

81. Molven, A., Ringdal, M., Nordbo, A. M., Raeder, H., Stoy, J., Lipkind, G. M., Steiner, D. F., Philipson, L. H., Bergmann, I., Aarskog, D., Undlien, D. E., Joner, G., Sovik, O., Norwegian Childhood Diabetes Study Group, Bell, G. I., and Njolstad, P. R. (2008) *Diabetes* **57**, 1131-1135

82. Oyadomari, S., Koizumi, A., Takeda, K., Gotoh, T., Akira, S., Araki, E., and Mori, M. (2002) *J. Clin. Invest.* **109**, 525-532

83. Nakagawa, S. H., Hua, Q. X., Hu, S. Q., Jia, W., Wang, S., Katsoyannis, P. G., and Weiss, M. A. (2006) *J. Biol. Chem.* **281**, 22386-22396

84. Wan, Z.-L., and Liang, D. (1988) *Sientia Sinica* **31**, 1426-38
| protease       | photo-adduct mass (kD)\(^a\) | assignment               | reference |
|---------------|-------------------------------|--------------------------|-----------|
|               | glycosyl                      | deglycosyl               |           |
| chymotrypsin  | 50 (47)                       | (30)                     | L1/ part of CR\(^b\) | 6         |
| chymotrypsin  | 34                            | 23                       | C-term. fragment\(^c\) | 34        |
| chymotrypsin  | 20                            | 17                       | ID-N\(^d\)            | 34        |
| trypsin\(^3\) | 31 (28)                       | 20 (17)                  | L1\(^e\)              | 6         |

\(^a\)Molecular masses in parentheses are exclusive of respective A- or B-chain adduct.

\(^b\)Fragment contains ca. 250-260 residues (supplemental Fig. S2).

\(^c\)Fragment contains (from C- to N) ID-N, Fn2a, and part of Fn1.

\(^d\)Fragment is likely to begin within Fn2a.

\(^e\)Fragment likely to contain initial residues of CR (supplemental Fig. S1).
**FIGURE LEGENDS**

**Figure 1. Structure of insulin and domain organization of the insulin receptor.** (A) Ribbon model of insulin (as T-state protomer; Protein Databank identifier 4INS) with A-chain in red and B-chain in blue. Selected side chains are shown: Glu\textsubscript{A4} and Tyr\textsubscript{A14} (red); Gln\textsubscript{A5} and Thr\textsubscript{A8} (green); Leu\textsubscript{A13}, Leu\textsubscript{A17} and Val\textsubscript{B17} (magenta); His\textsubscript{B10} (black); and Phe\textsubscript{B24} and Phe\textsubscript{B25} (gray). (B) Domain organization of IR as (αβ)\textsubscript{2} dimer. Color-coded segments indicate structural domains; at left are shown selected sites of limited proteolysis of photo-cross-linked complexes: trypsin (tr, bracket at L1/CR junction) and chymotrypsin (ch, arrow within CR), and asterisk (tryptic and chymotryptic sites at Fn2a/ID-N junction). Beige arrow heads indicate sites of N-linked glycosylation in extracellular portion of IR. Figure is adapted from ref. 5 with permission of the authors. Dashed lines outline domains (light gray) not present in crystal structure of IR ectodomain (Fig. 7); these span the transmembrane α-helix (TM), juxtamembrane segment (JM), tyrosine kinase (TK), and C-terminal tail of β-subunit (βCT). (C) Space-filling model of insulin protomer (with residues B27-B30 removed) showing functional epitope (gray) and its extended structural epitope (green). (D) View of insulin protomer rotated by 90° about vertical axis with classification of binding sites 1 (gray) and 2 (magenta) as proposed by De Meyts (18,73). Residues B27-B30 were deleted from coordinate file to enable better visualization of Ile\textsuperscript{A2} and Val\textsuperscript{A3} in accord with detachment model (preceeding article in this issue (31)).

**Figure 2. Sites of photo-probes in insulin and summary of cross-linking efficiencies.** (A) Sequence of human insulin showing A-chain (top) and B-chain (bottom); in each case the N-terminal residue is at left and C-terminal residue at right. The three disulfide bridges (cystines A6-A11, A7-B7, and A20-B19) are shown as black lines. Sites of Pmp derivatives are highlighted in light blue (within A- or B-chains) or magenta (B-chain extensions B0 and B31). Gray shading (residues B5 and B17) indicates failed syntheses due to impaired chain combination. Pmp\textsubscript{B8} analog contained D isomer; both L- and D Pmp derivatives were individually prepared at positions A1 and B24. Not shown: “DKP” substitutions in B-chain to yield a monomeric template (Asp\textsuperscript{B10}, Lys\textsuperscript{B28}, and Pro\textsuperscript{B29}) and biotin tags at possible attachment sites (B0, B1, A1, or via the ε-NH\textsubscript{2} moiety of a D-Lys\textsuperscript{A1} substituent; see Methods and supplemental Table S1). (B) Front- and back surfaces of insulin color-coded by efficiency of photo-cross-linking. Front surface is predominantly comprised of B-chain residues; the back surface, A-chain residues. Sites of high, medium, and low photo-cross-linking efficiency are shown in red, green, and black respectively; sites not tested are shown in gray. The structure shown is based on a T\textsubscript{6} crystallographic protomer (2-Zn molecule 1; Protein Databank identifier 4INS).

**Figure 3. Photo-cross-linking studies of B-chain derivatives.** (A) Western blots showing subunits of the holoreceptor (α\textsubscript{2}β\textsubscript{2}, 460 kDa; α, 135 kDa; β, 95 kDa) cross-linked with biotin-labeled insulin derivatives containing Pap at positions B24-D (lanes 1-4), B24-L (lanes 5-8), B25 (lanes 9-12), B26 (lanes 13-16), B27 (lanes 17-20), B29 (lanes 21-24), or at extended site B31 (lanes 25-28). Samples were treated without (odd-numbered lanes) or with (even-numbered lanes) UV irradiation. After cross-linking and reduction with DTT, samples were resolved by SDS-PAGE and detected by either with alkaline-phosphatase-conjugated NeutrAvidin (NAv; top panel) or by an anti-receptor antibody (IRα\textsubscript{N}; middle panel); the latter demonstrates equal amounts of IR. Control blots probed with NAv (bottom panel; without DTT reduction) demonstrate equal amounts of insulin analog. Control lanes: in each case photo-cross-linking was not observed in absence of IR and UV irradiation (lanes 1, 5, 9, 13, 17, 21, and 25), in absence of IR and presence of UV irradiation (lanes 2, 6, 10, 14, 18, 22, and 26), or in presence of IR and absence of UV irradiation (lanes 3, 7, 11, 15, 19, 23, and 27). (B) Limited
chymotryptic digestion of Pap\textsuperscript{B24-}, Pap\textsuperscript{B25-}, Pap\textsuperscript{B26-}, and Pap\textsuperscript{B27} photo-cross-linked hormone-receptor complexes (6). At indicated time points (top of gels), aliquots were mixed with equal volume of Laemmli sample buffer containing 100 mM DTT, heated at 95 °C for 5 min, and resolved by SDS-PAGE. Photo-products and their fragments were blotted onto nitrocellulose membrane, and probed with N\textalpha{}-V (left) or IR\textalpha{}-N (right panel). Apparent molecular masses: (\textalpha{}(N)-B) 47 kD, (\textalpha{}(C)-B) 34 kD, and (\textalpha{}(N)/\textalpha{}(N)-B') 47/50 kD (Table 1). (C) Deglycosylation of B25 and B27 chymotryptic fragments as probed by N\textalpha{} (left) or IR\textalpha{}-N (right). IR was cross-linked to Pap\textsuperscript{B25-} or Pap\textsuperscript{B27} derivatives (lanes 13 and 16; dash at top (–) indicates absence of protease), digested with chymotrypsin (lanes 14, 15, 17, and 18; +), and then subjected to enzymatic deglycosylation (lanes 15 and 18; labeled as “d”) (25). Apparent molecular masses: (\textalpha{}(C)*-B) 23 kD and (ID-N*/B) 17 kD (see Table 1).

Figure 4. Photo-cross-linking studies of A-chain derivatives. (A) IR cross-linking of insulin derivatives containing Pap at positions A1-D (lanes 7-10), A2 (lanes 11-14), A3 (lanes 15-18), A4 (lanes 19-22), and A8 (lanes 23-26) in relation to B-chain derivatives B25 (lanes 1-2) and B24 (lanes 3-6). Samples were resolved by SDS-PAGE prior to (odd-numbered lanes) or following (even-numbered lanes) UV irradiation. A-chain Pap derivatives (except A1) have biotin tag attached to the \textepsilon{}-NH\textsubscript{2} moiety of a D-Lys\textsuperscript{A1} substituent; D- and L A1 Pap derivatives were labeled with biotin at N\textalpha{}. The format is as in Figure 3A. Cross-linked adducts were detected after DTT reduction by N\textalpha{} (top panel). Control blots probed by IR\textalpha{}-N (middle panel) demonstrate equal amounts of IR. Control blots probed with N\textalpha{} (bottom panel; without DTT) demonstrate equal amounts of insulin derivative. Control lanes: in each case photo-cross-linking band was not detected in absence of IR and UV irradiation (lanes 3, 7, 11, 15, 19, and 23), in absence of IR and presence of UV irradiation (lanes 4, 8, 12, 16, 20, and 24), or in presence of the IR and absence of UV irradiation (lanes 1, 5, 9, 13, 17, 21, and 25). (B) Western blots showing A2-specific IR photo-cross-linking at successive concentrations of Pap\textsuperscript{A2} derivative (lanes 3-7, corresponding to ligand concentrations of 200, 240, 320, 400, and 600 nM). Pap\textsuperscript{B25} specific photo-cross-linking is also shown (lane 1; lane 2 is empty). (C) Competition experiment. A2-specific photo-cross-linking could be competed by unmodified insulin or IGF-I, added prior to UV irradiation. Successive concentrations of insulin (lanes 1-4) or IGF-I (lanes 5-8) were successively 0X, 3X, 30X, and 300X times that of Pap\textsuperscript{A2} (600 nM).

Figure 5. Chymotryptic mapping of A-chain-specific photo-cross-linked complexes. Photo-adducts were obtained at indicated positions and digested with chymotrypsin as described (6). At successive time points (bottom of gels), aliquots were reduced by treatment with DTT, heat-denatured, and resolved by SDS-PAGE. Fragments were blotted onto nitrocellulose membrane and detected by N\textalpha{} (A and C) or antiserum IR\textalpha{}-N (B and D). Dashed boxes in panels A and B indicate 20 kD N-terminal L1-derived tryptic fragment estimated to contain ca. 100 residues (and three N-linked carbohydrates). Molecular masses and domain assignments are otherwise as indicated.

Figure 6. Mapping of photo-contacts and relationship to induced fit. (A) Schematic model of insulin fit. The insulin monomer is proposed to undergo a change in conformation from its closed unbound state (left) to a more open state (right) in which detachment of the C-terminal B-chain \textbeta{}-strand (residues B24-B28) exposes conserved aliphatic side chains Ile\textsuperscript{A2} and Val\textsuperscript{A3}. (B) Alternating pattern of photo-contacts by the C-terminal B-chain \textbeta{}-strand in which its even-numbered face (Phe\textsuperscript{B24} and Tyr\textsuperscript{B26}) contact the L1 domain of the IR (\textalpha{}-subunit residues 1-158) whereas the odd-numbered face (Phe\textsuperscript{B25} and Thr\textsuperscript{B27}) contact the C-terminal ID-derived tail of the \textalpha{}-subunit. (C) Space-filling model of insulin depicting its front and back surfaces (left and right). Putative L1-contact residues Tyr\textsuperscript{B16},
Phe$^\text{B24}$ and Tyr$^\text{B26}$ are shown in blue, and ID-contact residues (Gly$^\text{A1}$, Ile$^\text{A2}$, Val$^\text{A3}$, Glu$^\text{A4}$, Thr$^\text{A8}$, Tyr$^\text{A14}$, Phe$^\text{B25}$, and Thr$^\text{B27}$) are shown in yellow (A2 and A3) or gold. The A-chain is otherwise shown in light gray, and the B-chain in dark gray. (D) Corresponding molecular surfaces following removal of residues B26-B30 to simulate exposure of Ile$^\text{A2}$ and Val$^\text{A3}$ on detachment of the C-terminal B-chain \(\beta\)-strand in the hormone-receptor complex. The color code is the same as in panel C.

**Figure 7. Structure of IR ectodomain and possible proximity of L1 and ID-N.** (A) Ribbon model of component protomer (PDB code 2DTG). Individual domains L1, CR, L2, Fn1, Fn2, and Fn3 are shown in gray, black, light blue, red, purple, and dark blue, respectively. The insert domain (ID) exhibits missing or discontinuous electron density (residues 655-755; IR isoform A); only respective N- and C-terminal subsegments of ID-N and ID-C are well defined (orange). Color code is in accord with schematic model in Figure 1B. (B) Inverted-V-shaped dimer. The position of the plasma membrane is indicated at bottom in schematic lipid bilayer. (C) Stereo model showing L1 domain (gray ribbon) in relation to unassigned and discontinuous electron density (green), potentially from \(\alpha\)CT; density may represent ~20 residues of insert domain. \(\alpha\)CT and ID-N-related density may reorganize on insulin binding. Residues in L1 critical to hormone binding (as inferred from Ala scanning mutagenesis (62) are shown as red sticks. Panel C was kindly provided by C. Ward.

**Figure 8. Classical Insulin Allostery and Induced Fit.** (A) Conformational equilibria among three families of zinc insulin hexamers: T\(_6\) (left), T\(_3\)R\(_3\) (middle), and R\(_6\) (left). In each structure the A-chains are shown in light gray; residues B9-B30 in dark gray; side chain of Phe$^\text{B24}$ in tawny; and axial zinc ions (overlaid) in magenta. The adjustable conformations of residues B1-B8 in T- and R-states are highlighted in green and blue, respectively. Coordinates were obtained from PDB entries 4INS, 1TRZ, 1TNJ. (B) T-state specific B7-B10 \(\beta\)-turn and adjoining B7-A7 disulfide bridge (gold). Substitution of invariant Gly$^\text{B8}$ (\(\alpha\)C in red) by D-amino-acids stabilizes the T-state but markedly impairs receptor binding (33,70). Corresponding L-amino-acid substitutions destabilize the T-state but can be highly active. Arrow indicates \textit{pro-D} H\(_\alpha\) of Gly$^\text{B8}$. A- and B-chain residues (B6-B10 and A6-A8) are otherwise shown in light and dark gray, respectively; selected residue numbers are provided. (C) Schematic model of insulin T state showing proposed sites of conformational change on receptor binding: an R-state-related B8-related switch (red, \textit{right}) within the N-terminal segment of the B-chain (green) and a B24-related switch (tawny, \textit{right}) as revealed by stereospecific unfolding of the C-terminal segment of the B-chain (see preceding article in this issue) (31). The A-chain is represented as a gray rectangle, the central B-chain \(\alpha\)-helix (cylinder) and C-terminal \(\beta\)-strand (arrow) are shown in black.

**Figure 9. Sites of diabetes-associated mutations in human insulin.** (A) Sequence of human insulin with A-chain (top) in red and B-chain (bottom) in blue; disulfide bridges are shown in orange. Filled black circles represent mutations associated with neonatal diabetes mellitus due to presumed folding defects in proinsulin (78-81); gray shaded circles represent mutations that permit disulfide pairing in ER with subsequent circulation of mutant proteins. Whereas substitutions at positions A3, B24, and B25 lead to variant hyperinsulinemias, the B10 substitution interferes with protein trafficking, leading to mutant hyperproinsulinemia (77). (B) Positions of mutation sites in the T-state of insulin (crystallographic protomer 1 of 2-Zn insulin; Protein Databank accession code 4INS) (1). Black spheres indicate C\(_\alpha\) atoms of Gly at A1, B8, and B23.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 6
Figure 8
Decoding the cryptic active conformation of a protein by synthetic photo-scanning. Insulin inserts a detachable arm between receptor domains

Bin Xu, Kun Huang, Ying-Chi Chu, Shi-Quan Hu, Satoe Nakagawa, Shuhua Wang, Run-Ying Wang, Jonathan Whittaker, Panayotis G. Katsoyannis and Michael A. Weiss

J. Biol. Chem. published online March 25, 2009

Access the most updated version of this article at doi: 10.1074/jbc.M900087200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2009/03/27/M900087200.DC1