Structural Origins of the Functional Divergence of Human Insulin-Like Growth Factor-I and Insulin[†,‡]

Andrzej M. Brzozowski,* Eleanor J. Dodson, Guy G. Dodson, Garib N. Murshudov, Chandra Verma, Johan P. Turkenburg, Frederik M. de Bree,[#] and Zbigniew Dauter[§]

Structural Biology Laboratory, Chemistry Department, University of York, Heslington, York YO10 5DD, United Kingdom

Received January 28, 2002; Revised Manuscript Received April 23, 2002

ABSTRACT: Human insulin-like growth factors I and II (hIGF-I, hIGF-II) are potent stimulators of cell and growth processes. They display high sequence similarity to both the A and B chains of insulin but contain an additional connecting C-domain, which reflects their secretion without specific packaging or precursor conversion. IGFs also have an extension at the C-terminus known as the D-domain. This paper describes four homologous hIGF-1 structures, obtained from crystals grown in the presence of the detergent SB12, which reveal additional detail in the C- and D-domains. Two different detergent binding modes observed in the crystals may reflect different hIGF-I biological properties such as the interaction with IGF binding proteins and self-aggregation. While the helical core of hIGF-I is very similar to that in insulin, there are distinct differences in the region of hIGF-I corresponding to the insulin B chain C-terminus, residues B25−B30. In hIGF-I, these residues (24−29) and the following C-domain form an extensive loop protruding 20 Å from the core, which results in a substantially different conformation for the receptor binding epitope in hIGF-I compared to insulin. One notable feature of the structures presented here is demonstration of peptide-bond cleavage between Ser35 and Arg36 resulting in an apparent gap between residues 35 and 39. The equivalent region of proinsulin is involved in hormone processing demanding a reassessment of the structural integrity of hIGF-I in relation to its biological function.

Human insulin-like growth factor I (hIGF-I) is a 70-amino acid single chain protein that mediates somatic growth. It has a high (45−52%) sequence similarity with the B and A chains of human insulin, and 67% sequence identity with human insulin-like growth factor-II (IGF-II) (see Figure 1) (*1*).

A number of NMR studies (*2*−*5*) and a few recent crystallographic analyses (*6*, *7*) have revealed the essentially identical nature of the core region and the organization of the three helical segments of insulin and hIGF-I. The short

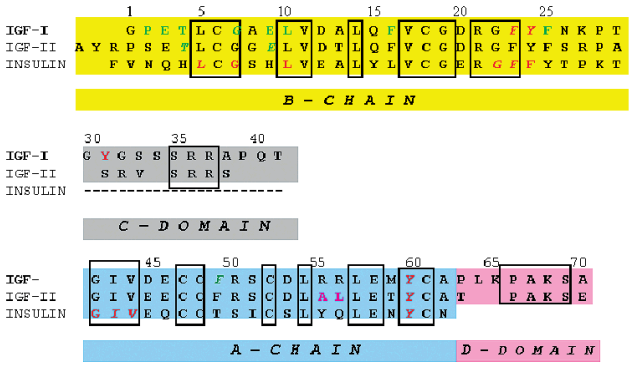

FIGURE 1: The sequence and chain organization of hIGF-I, hIGF-II, and insulin. The IGF-specific C- and D-domains are colored grey and pink, respectively; the B and A chains of insulin and their equivalents in hIGF-I/II are highlighted in yellow and blue, respectively. Residues important for the IGF-1R or IR binding are in red, with residues responsible for association with IGFBPs in green (mutations of the highlighted residues result in a minimum 90% drop in binding; residues for which substitution results in even higher impact on affinities toward receptors and IGFBPs are in italic (for details, see refs *8*−*11*).

N- and C-terminal extensions in hIGFs are directed away from the body of the molecule and are generally mobile. These regions are not detected by NMR and are only partially visible in the crystallographic analyses. Although the IGF-specific C-domain is covalently attached to the region that is the equivalent of the A and B chains of insulin, it is still relatively mobile. The presence of the C-domain does not affect the conformation of the A chain, which is insulin-like. It is, however, associated with structural differences in

[†] This work was partially supported by the European Community (Human Capital and Mobility Program, Contract No. CHRX-CT94-0556). The infrastructure of the Structural Biology Laboratory at York is supported by the BBSRC. We thank the European Union for support of the work carried out at EMBL Hamburg outstation through the HCMP access to large installations project (Contract No. CHGE-CT93−0040).

[‡] Atomic coordinates of the hIGF-I-esrf, hIGF-I-dares, hIGF-I-hamb-RT, hIGF-I-inh-RT crystal structures have been deposited in Protein Data Bank (accession codes: 1GZR, 1H02, 1GZZ, 1GZY).

* Corresponding author. E-mail: marek@ysbl.york.ac.uk; telephone: +44-1904-432570; fax: +44-1904-410519.

[#] Present address: Netherlands Institute for Brain Research, Academic Medical Center, Meibergdreef 33, 1105 AZ Amsterdam, The Netherlands.

[§] Present address: Synchrotron Radiation Research Section, NCI, Brookhaven National Laboratory, Bldg. 725A-X9, Upton, NY 11973, USA.

[1] Abbreviations: hIGF, insulin-like growth factor; IGFBP, IGF-binding protein; IGF-1(2)R, IGF-type 1 and type 2 receptor; IR, insulin receptor; ELISA, enzyme-linked immunosorbent assay; TRIS, (Tris-[hydroxymethyl]aminomethane); SB12, *N*-dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate; big deoxy CHAPS, *N*,*N*-bis(3-D-gluconoamidopropyl) deoxycholamide; 2K PEG MME, poly(ethylene glycol) monomethyl ether of 2000 M.W.

10.1021/bi020084j CCC: $22.00
Published on Web 07/04/2002

the conformation of the residues that are equivalent to insulin residues B25−B30. Both insulin and hIGF-I bind as monomers to their receptors and display an ability to bind to each other's receptors; this latter phenomenon is considered to be physiologically significant (*12*). By contrast, the two molecules have very different cellular origins; insulin is synthesized in the β-cell of the Islets of Langerhans, while hIGF is synthesized mainly in the liver, although other tissues are also involved. The two molecules undergo profoundly different mechanisms of processing, secretion, and circulation in the blood. It is the characteristic hexamer structure that provides the solution properties and stability needed for transport, processing, and storage of proinsulin in the β-cell. Proinsulin, contrary to IGF, is processed further by specific convertases to insulin, whose hexamers rapidly disassemble upon release to the bloodstream into the native monomer that binds to its receptor; this complex is then endocytosed and degraded. hIGF-I, in a contrasting process, circulates as an inactive species for lengthy periods protected by the numerous IGF binding protein (IGFBP) molecules (see for example refs *13*−*15*). hIGF-I exerts its pleiotropic actions by binding and activating a 440-kDa type 1 IGF receptor (IGF-1R), an $\alpha_2\beta_2$ tyrosine kinase heterodimer (50% homologous with the insulin receptor (IR)) (*15*, *16*).

Here we present four crystal structures of hIGF-1 analyzed at 100 K and room temperature, based on crystallographic data collected using X-ray sources at home, the SRS (Daresbury), the EMBL (Hamburg), and the ESRF (Grenoble). The crystals were grown from media containing a detergent SB12 that acts to reduce aggregation and, perhaps, conformational flexibility of the monomer. Although the axial parameters of the crystals are very similar, the quality of the data collected at 300 and 100 K varies considerably, reflecting overall mobility of the hIGF-I. The structures are generally similar to those reported in the earlier investigations, but there is more detail detected at the N- and C-terminal extensions and particularly in the B- and C-domain loop.

The corresponding chain numberings in insulin and hIGF-II are given respectively in [ ] and {} brackets. Comparisons of hIGF-I with insulin structures and hIGF-II result from structure alignments using helix Ala8[B9]{Gly11}−Cys18-[B19]{Cys22}.

RESULTS AND DISCUSSION

Structural Organization of hIGF-I. The description and discussion of hIGF-I are based mostly on the 2.0 Å X-ray data collected at the ESRF (referred to as hIGF-I-esrf), which has the best data statistics (Table 1). The other three hIGF-I structures are referred to a: hIGF-I-dares (data collected at the SRS at a temperature of 120 K), hIGF-I-hamb-RT (data collected in Hamburg at room temperature), and hIGF-I-inh-RT (data collected in-house at room temperature); for all data and refinement statistics see Table 1 of the Supporting Information.

The atomic structure of hIGF-I was determined almost completely in the ESRF analysis (hIGF-I-esrf is referred to here also as hIGF-I); only two N-terminal, four C-terminal, and three C-domain residues (Arg36−Arg37−Ala38) have not been detected satisfactorily (Figure 3c). The core of all four structures shows close overall similarity with the hIGF-I-esrf molecule, while the more profound structural differences are located around the C-domain and N- and C-termini (Figure 2, Figure 3c).

Table 1: X-ray Data and Refinement Statistics

structure name	IGF-I-esrf
Data Collection and Processing Statistics	
data collection site	ID14-4 ESRF
wavelength (Å)	0.93
space group	$C222_1$
unit cell dimensions (Å) (*a*, *b*, *c*)	30.78, 69.47, 65.0
resolution range (Å) (outer shell)	20−2.0 (2.07−2.0)
observations	17172
unique reflections	4839
$I/\sigma(I)$	18(12)
completeness (%)	98.3(95.8)
R_{merge} [a]	3.7(20.3)
Refinement Statistics	
resolution range (Å)	20−2.0
reflections used (R_{free} set)	4620(208)
R_{cryst}/R_{free} [b]	23.4/29.5
protein/ligand atoms/waters	475/20/35
r.m.s. bonds/angles (Å) [c]	0.015/2.6
r.m.s. main chain ΔB (Å^2) [d]	1.05
mean B-factor (Å^2) [e]	31.3/33.3/35.5/38.5
%A, B, L (a, b, l, p) [f]	89.8(10.2)

[a] $R_{merge} = 100\sum|I - \langle I\rangle|/\sum\langle I\rangle$. [b] $R_{cryst} = \sum|F_{obs} - F_{calc}|/\sum F_{obs}$; R_{free} is as R_{cryst} but calculated over 4.5% of data that were excluded from the refinement process. [c] Root-mean-square deviations in bond length and angle distances from Engh and Huber ideal values. [d] Root-mean-square deviations between $B_{factors}$ for bonded main chain atoms. [e] Mean temperature factor for whole molecule, main chain, side chain, ligand, and water atoms, respectively. [f] Percentage of residues located in the most favored (additional) regions of Ramachandran plot as determined by PROCHECK (*17*).

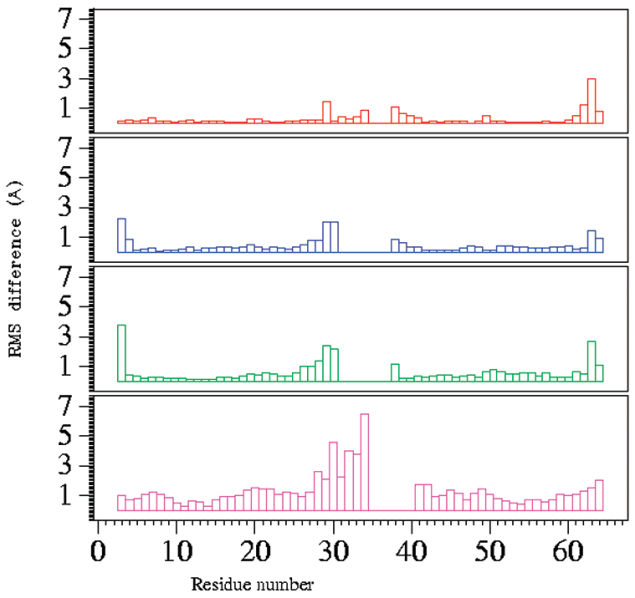

FIGURE 2: Comparison of rms deviations (in Å) between Cα atoms of hIGF-I-dares (red), hIGF-I-hamb-RT (blue), hIGF-I-inh-RT (green), hIGF-I described by Vajdos et al. (6) (magenta), and hIGF-I-esrf as a reference structure.

Comparison of the four hIGF-I structures reported here, the hIGF-I structure described by Vajdos et al. (*6*), and insulin reveals that the hIGF-I core (consisting of three helices) is very similar to its equivalent in insulin (the 1.5 Å resolution structure described by Baker et al. 1988 (*18*) has been used throughout as a reference structure for insulin).

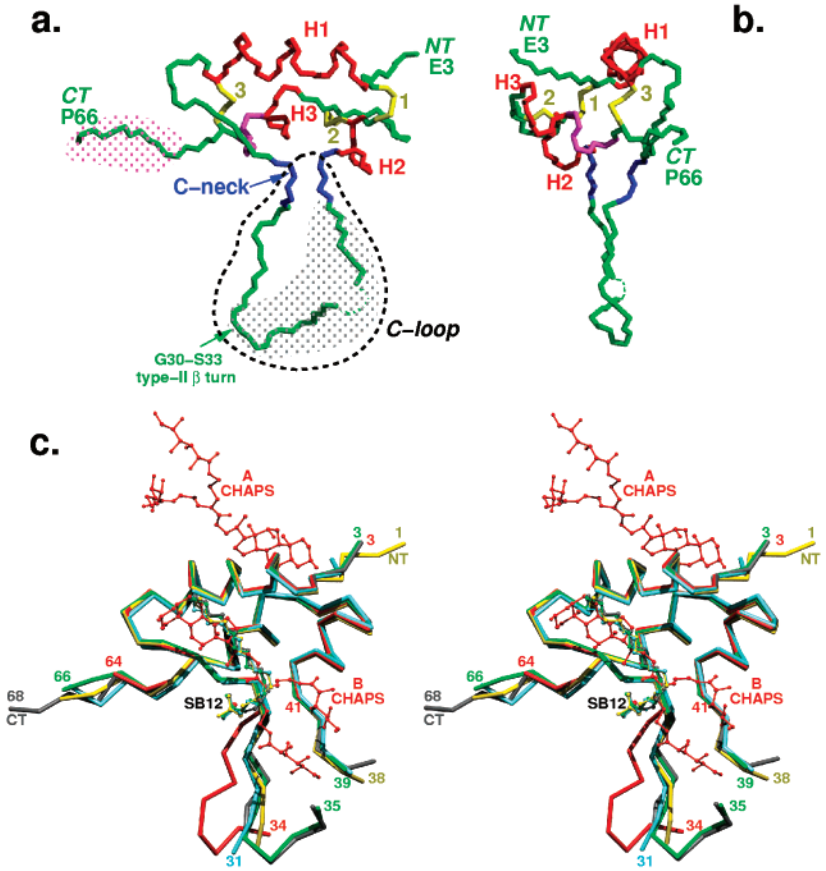

FIGURE 3: (a) and (b) Two orthogonal views of a backbone representation of the hIGF-1 structure. Helices are in red and are defined as helix 1 or H1 (Gly7−C18), helix 2 or H2 (Ile43−Cys47), helix 3 or H3 (Leu54−Leu57); the 3_{10} helix (Glu58−Tyr60) is in magenta; disulfides are in yellow with 1 corresponding to Cys6−Cys47, 2 to Cys46−Cys52, and 3 to Cys18−Cys61. The C-neck (Phe25−Asn26 and Gly42−Ile43) is in blue; NT and CT are the N- and C-termini; gray and magenta dotted surfaces indicate the areas of the C- and D-domains, respectively. (c) Overall Cα chain comparison of the four hIGF-I structures: hIGF-I-esrf (green), hIGF-I-dares (grey), hIGF-I-hamb-RT (yellow), hIGF-I-inh-RT (blue) and hIGF-I-CHAPS (*6*) (red), with detergents as ball-and-stick models; numbers indicate the N- and C-termini and the C-domain gaps observed in these structures. CHAPS A corresponds to its position described in (*6*); CHAPS B depicts a symmetry related orientation of CHAPS A (not discussed in ref 6) that is close to the SB12 location reported here (symmetry equivalent of SB12 in CHAPS A position is not shown for picture clarity) (divergent stereo).

The most striking departure from the insulin fold is at Tyr24 and Phe25 ([B25&B26]). The Gly22[B23]−Phe25-[B26] β strand does not extend as far as in insulin where it finishes at Thr29[B30]. In hIGF-I, it ends with a tight bend at Phe25[B26] leading to the IGF-specific C-loop, which includes residues Asn26−Thr29 and the C-domain peptide that links back to the A-domain. This C-loop extends such as a dagger ~20 Å away from the core of the molecule, giving hIGF-I an overall shape of a sharp wedge. Residues Gly30−Tyr31−Gly32−Ser33 of the C-domain form a classical type-II β-turn at the center of the C-loop, directing it to the A-domain. There is a gap in the electron density for the C-loop for residues Arg36−Arg37−Ala38 in hIGF-I. The C-domain polypeptide chain then returns back to the core of IGF-I via Gly42, with a sharp bend of the main chain at Ile43 [A2]. Two antiparallel hydrogen bonds between the beginning and end of the C-loop (25CO···NH43 and 27NH···OC41) create two short antiparallel β-strands, referred to here as the "C-neck" (Figure 3a,b). The definition of this new structural feature is reinforced by additional hydrogen bonds between 42CO···HN45 and 42NH···45OD1 of Asp45-[A4] from the contiguous α-helix 2 (Ile43[A2]−Cys47[A6]). The involvement of the Asp45[A4] side chain extends the neck, which consists of two, short β-strands, into a structural motif of three pseudo-β-strands. It is worth noting that a similar hydrogen bond motif also occurs in insulin, where [A4Glu] is salt bridged to the free amino group of [A1Gly] (*18*).

The detergent SB12 was an additive in our crystallization of hIGF-I, and one SB12 molecule is found to pack between protein molecules in the crystal. Its alkyl chain stretches along the hydrophobic patch formed by Val11[B12], Phe23-[B24], Phe25[B26] and bends to involve its SO_3^- group in hydrogen bonds with NH main chain atoms of Asn26 and Phe25 (see Figure 4 in Supporting Information). As the Val11[B12], Phe23[B24], Phe25[B26] surface corresponds to the dimer-forming interface of insulin, it seems likely that the SB12 detergent prevents nonspecific aggregation of hIGF-I observed during crystallization without detergent (*6*).

The fragments of hIGF-I reported as highly mobile in all NMR structures, namely, the N- and C-termini and the C-domain, occupy cavities in the crystal and are free of strong lattice contacts; hence, the conformations observed here may be physiologically relevant.

Structural Variation in hIGF-I. The comparison of all hIGF-I crystal structures solved during this study reveals the dynamic nature of some parts of this molecule (Figure 3c). The mobility of the C-loop is visibly increased in both room

temperature structures (hIGF-I-inh-RT and hIGF-I-hamb-RT), where the gaps in the electron density are widened from Tyr31 to Pro39 (in hIGF-I-inh-RT). The inherent flexibility of the C-loop is further corroborated by comparison with the recent hIGF-I structure (*6*) obtained, as described above, from crystals grown in the presence of detergent big CHAPS (referred to here as IGF-I-CHAPS). Despite different crystallization conditions and detergent used, this crystal form belongs to the same space group $C222_1$, with a maximum difference of ~3% in the unit cell dimensions. Although the crystal packing and overall structure of hIGF-I reported here (grown in the presence of SB12; the four structures are collectively referred to in this paragraph as hIGF-I-SB12 when discussing common features) and hIGF-I-CHAPS are similar, the inherent mobility of the C-loop and differences in the chemical nature of the detergents used, result in local structural variations and partially different detergent−protein interactions in these, otherwise relatively isomorphous, structures.

The most significant structural divergence between IGF-I-SB12 and IGF-I-CHAPS occurs in the conformation of the C-loop. In hIGF-I-CHAPS, the weak and disconnected electron density for part of the C-loop is longer than in hIGF-I-SB12 and includes the Ser35−Gln40 stretch of the C-domain. The conformation at Asn26−Ser34 in the "CHAPS" crystal form is also significantly different from that in hIGF-I-SB12. The C-loop starts to deviate from hIGF-I-SB12 at Asn27 (0.75 Å between their Cα atoms) to reach a separation of ~9 Å between the Cα atoms of Ser34; consequently, the side chains of Tyr31, crucial for IGF-1R binding, are ~8.6 Å apart (see Figure 2c). The structure of the terminal part of the C-loop (around Thr41−Gly42) is very alike in both structures. Additionally, these structural differences might also reflect slightly different detergent binding modes in both crystal structures. Although the deoxycholamide headgroup of the detergent (CHAPS B in Figure 3c, symmetry related to CHAPS A described in ref *6*) in hIGF-I-CHAPS occupies the position of the alkyl chain of SB12, its long hydrophilic moieties do not overlap with the short polar group of SB12. They would also sterically clash with the distal part of the C-loop in its hIGF-I-SB12-like conformation. Additionally, the Br^- ion found in hIGF-I-CHAPS mimics the position of the SO_3^- group of the SB12 detergent (with Br−S distance of 0.66 Å).

It is perhaps unexpected that despite the different chemical natures of the two detergents (nonionic big CHAPS compared to zwitterionic SB12), both compounds target similar surfaces of hIGF-I. Their perceived impact on the biologically relevant behavior of hIGF-I may, however, differ depending on whether the interactions represented by CHAPS-A or CHAPS-B are considered (see Figure 3c). The binding mode "A" corresponds roughly to the IGF-I/IGFBP-5 complex interface (*7*), and results in the interference with IGFBP-1 and IGFBP-3 association with the hormone (*6*). The hIGF-I displacement from the IGF-I/IGFBP-3 complex by the nonpeptide, isoquinoline compounds (*19*) also corroborates this hypothesis, as their poly-hydroxy-aromatic character is similar to the deoxycholamide functional group of CHAPS. In contrast to the binding mode "A", the main effect of detergent binding mode "B" would be to diminish dimerization of hIGF-I, as dimer formation has already been reported from sedimentation equilibrium data (*6*). Our MALDI-TOF analyses of different hIGF-samples also showed a significant population of hIGF-I in dimeric form (data in Supporting Information), while the dissolved crystal samples behaved differently. In these, the dimer population was significantly diminished. As a MALDI-TOF-characteristic, artificial sample dimerization occurs fairly frequently, these observations have to be considered with care. Whether SB12 (or CHAPS) interferes with IGF's insulin-like dimerization, the interactions with cognate binding proteins or destabilization of other types of dimers, remains unclear.

It is possible, however, that breaks in the C-domain electron density in the 31−39 region of hIGF-I result from cleavage at Arg36−Arg37, although there is no evidence up to now as to whether this occurs during or before structural experiments. The short (~3 Å) Ser35···Pro39 separation in the hIGF-I-esrf and hIGF-I-dares maps and the well-defined local electron density made it very difficult to close the gap (see Figure 4), suggesting cleavage. Moreover, there is continuous electron density between Ser35 and Pro39 that allows modeling of the -COOH group at Ser35 to lie adjacent to the main chain of Pro39. The presence of a weak (at 0.5σ level on $2F_o-F_c$ maps) electron density stretching out from Pro39 into the intermolecular space may indicate the presence of the "missing" Arg36−Arg37−Ala38- residues. The fulfilment of the stereochemical criteria for this "alternative", extended conformation of the Arg36−Pro39 chain would require repositioning of Pro39 from its current conformation; this rearrangement would fit the observed electron density. It should be noted that the high mobility of the C-loop, especially its 35−39 region, is correlated in the crystal with the disorder of the C-terminus, which is in close contact with some of the residues of the C-domain.

The SDS−PAGE electrophoresis of reduced samples from crystals used for X-ray data collection showed more than one major band, implying cleavage, a behavior exhibited also by the other hIGF-I samples used for crystallization (see Supporting Information). Additionally, the Electrospray and MALDI-TOF mass spectrometry showed that the mass of hIGF-I in the crystal was increased by 16 Da in comparison with a fresh solution of the hIGF-I sample. This would correspond to one oxygen atom taken up by peptide bond hydrolysis. An alternative explanation for the 16 Da mass increase could be the oxygenation of Met59; there is, however, no extra electron density at this residue. The N-terminal sequencing of the crystal-derived IGF-I sample also points to heterogeneity, as both the RA- sequence and the main GPETL- N-terminal motif are observed. All these data show the presence of hIGF-I cleavage in the crystals and that there is a need for careful reexamination and more rigorous assessment of the structural integrity of hIGF-I in vitro and in vivo. There is an additional relevance to this cleavage as the sequence -Arg36−Arg37−Ala38- of the "missing" region is similar to the proinsulin B chain−C chain junction processed during its maturation (*20*) and a single cleavage in the C-domain (before Arg37) in the plasma-derived hIGF-I has already been reported (*21*).

IGF-I and Insulin Structural Relationship. Although the overall architecture of the helices 1, 2, and 3 in hIGF-I is very like its counterpart in insulin, there are substantial structural dissimilarities between hIGF-I and insulin (Figure 5). The most striking difference results from the presence of the C-loop in hIGF-I, which alters completely the character

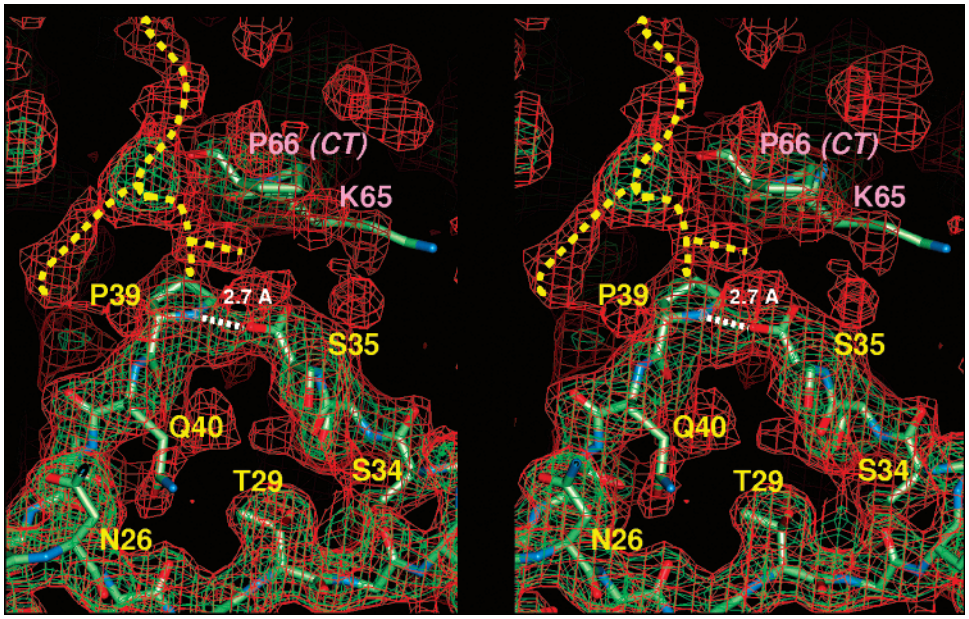

FIGURE 4: The electron density for the 35−39 gap region of the C-domain in the hIGF-esrf structure; green, 1σ level; red, 0.5σ level. Yellow dashed lines indicate unassigned low-contour electron density, next to Pro39, that may correspond to the missing Arg36−Arg37−Ala38 residues. The white dashed line represents the potential hydrogen bond between OH of the carboxyl group of Ser35 and the NH of Pro39 (the Ser35 carboxyl group is not included in the deposited PDB file as it was modeled only after additional cycles of REFMAC following completion of the model building and refinement; if the R36−R37-A38 are considered as present, this hydrogen bond could not be formed by Pro39 in its current conformation as this residue would have to be substantially remodeled to be joined with Ala38); pink labels are associated with a symmetry related C-terminus of a neighboring hIGF-I molecule (divergent stereo).

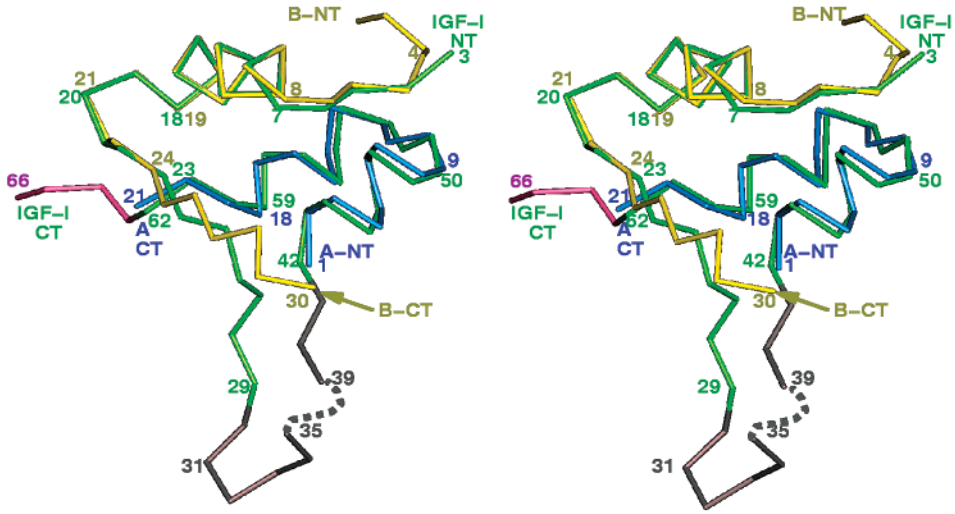

FIGURE 5: Comparisons of the Cα chains of hIGF-I (hIGF-I-esrf) and human insulin with domain color coding as in Figure 1: B and A chains of insulin in yellow and blue, respectively; hIGF-I C-domain in gray, D-domain in magenta (remaining parts of hIGF-I in green; divergent stereo).

of the surface that in insulin is responsible for its dimerization and receptor binding. The most visible consequence of this characteristic hIGF-I structural feature is a repositioning of the B chain C-terminal residues into an integral part of the C-loop. The neck at the C-loop brings the Gly22[B23]−Phe25[B26] β-strand much closer to the N-terminus of helix 2 [A1−A2] and to the core of the protein. A ∼25° angle between these strands in hIGF-I and insulin (with the Cα's of Phe23[B24] as the reference starting point) results in a ∼3.4 Å difference between the corresponding Cα atoms of Phe26 and [TyrB27]. This conformation prevents the formation of an antiparallel β-sheet structure in the hIGF-1 molecule and has not been observed in proinsulin or in any insulin molecule containing an engineered link between the B chain C-terminus and the A chain N-terminus (see for example refs *22*−*26*). The main chain bend at residues Gly22[B23]−Thr29[B30] (by ca. 105°) in hIGF-I at the Cα atom of Phe25 directs these residues outward into the C-domain (Figure 6). This positioning of the 26−29[B27−B30] polypeptide exposes residues 42−43[A1−A2]) that in insulin form the main IR binding epitope (the so-called site 1 (*27*)) (Figure 7). In insulin, these residues are buried by the C-terminal part of the B chain, and it has been postulated that its dislocation/rearrangement is a prerequisite for receptor

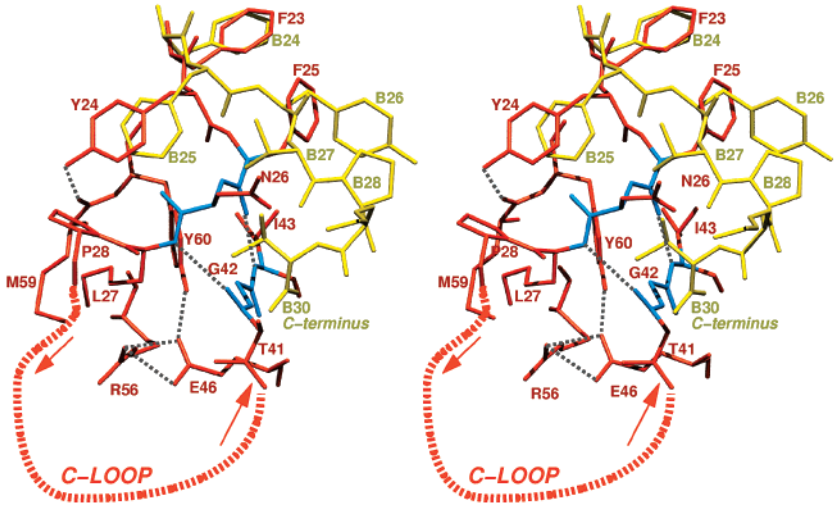

FIGURE 6: Comparison of the conformation of the hIGF-I 23−29 region (in red) and the corresponding C-terminal part of the insulin B chain (in yellow); the main chain of the C-neck part of hIGF-I in blue, hydrogen bonds are shown as dashed gray lines. The C-loop is depicted as a thick, red dashed half-circle with arrows indicating the direction of the polypeptide chain.

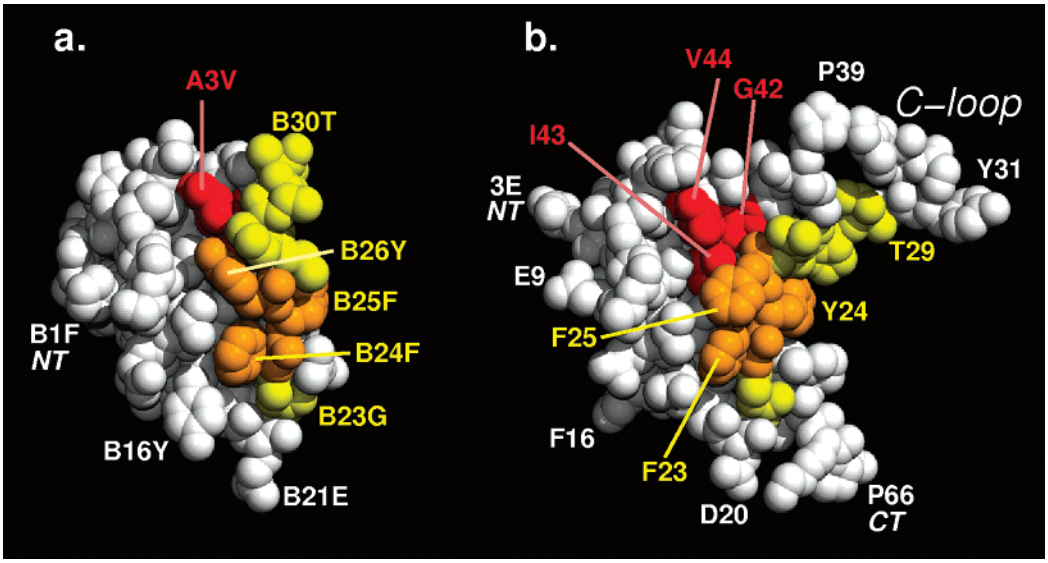

FIGURE 7: van der Waals representations of insulin (a) and hIGF-I (b) surfaces. Corresponding regions of the B chains implicated in interactions with receptors are marked in yellow and gold, with residues crucial for IR or IGF-1R binding are highlighted in gold. Residues [A1−A4] of the insulin A chain (and their 42−43 equivalents in hIGF-I) that are also important for interactions of insulin with its cognate receptor are marked in red. They are buried under the B chain in the insulin molecule, and are exposed in hIGF-I due to the presence of the C-loop and the tight main-chain bend at Phe25 in this growth factor.

binding (e.g., refs *23* and *28*). In contrast to insulin, the hIGF-I binding surface's exposure is achieved through the well-defined conformation at the C-loop.

The new "IGF-like" fold of the 22−29[B23−B30] chain in the hIGF-I-SB12 crystal structure (typical also for the hIGF-I-CHAPS molecule) is additionally stabilized by a hydrogen bond between the hydroxyl group of Tyr24 and the carbonyl oxygen of Met59 [AsnA18] (see Figure 6). It is obvious, however, that the hydrogen bond interactions of Tyr24 alone are insufficient for defining the conformation in the 23−25 region since a hIGF-I mutant with the insulin-like sequence Phe23−Phe24−Tyr25 has the same affinity as the wild-type hormone towards the IGF-1, -2, and insulin receptors. On the other hand, the aromatic character of the side chain at position 24 is still required for the full biological activity (*29*). It is also possible that the hIGF-specific positioning of Tyr24 may be a part of a wider network of interactions, leading in consequence to further stabilization of Tyr60[TyrA19], a residue crucial for the hIGF-I, hIGF-II, and insulin activity (*30*). In hIGF-I, the hydroxyl group of Tyr60[TyrA19] forms a hydrogen bond with the OE1 atom of Glu46([GlnA5] (a contact preserved in insulin and hIGF-II (to {Glu45}) as well). This hydrogen bond seems specific to hIGF-I as the carboxyl of Glu46 is replaced in insulin by the amide group of [GlnA19]. Furthermore, the stability of Tyr60 and its neighboring C-neck region may also benefit from complete burial and protection of this residue under the side chains of Arg56, Lys27, and Met59. However, the side chains of Lys27 and Met59 in hIGF-I-SB12 are stabilized by van der Waals contacts only and may adopt alternative conformations upon receptor binding.

As a result of the hIGF-I-specific conformation at residues 22−29 that mimics the [B23−B30] chain displacement in insulin (e.g., refs *31* and *32*), the IGF-1R binding epitopes (Gly22[B23]−Phe25[B26] and Gly42−Asp45[A1−A4]) on the hIGF-I surface can be involved in binding of this

hormone to both receptors (IGF-1R and IR) without any major structural reorganization within this hormone molecule (Figure 7). Thus, the well-defined structure of the Gly22−29[B23−B30] chain and the C-loop neck region in hIGF-I (and in hIGF-II) imply that the postulated rearrangement of the C-terminus of the B chain in IGF-I and IGF-II upon IGF-1/2R (or IR) binding (*33*), as proposed for the insulin−IR association, is unlikely.

While the hIGF-I specific conformation of the Gly22−29[B23−B30] chain will maintain/impose a monomeric state of the molecule, other residues that differ from those in insulin (e.g., [HisB10] → Glu9, [TyrB16] → Gln15) can also be relevant. For example, the switch of the Pro [B28] and Lys [B29] amino acids in insulin to Lys27−Pro28 is characteristic of IGF-I. As this switch is known to prevent insulin dimer formation and has been successfully used to create a monomeric insulin (*34*) its presence in hIGF-I is likely to contribute to the monomeric state of this hormone.

Importance of the C-Loop. The role of the C-domain for IGF-1R binding is clearly indicated by the observations that its deletion or replacement by short-linkers abolish or diminish dramatically the affinity of hIGF-I for IGF-1R (*33*, *35*). The structure at the C loop (residues 25−41) may have a double impact on receptor binding affinities and specificities of hIGF-I and insulin. First, it enforces a particular, IGF-specific, conformation of the Gly22−Phe25 β-strand, preserving at the same time an 'insulin-like" fold across residues 42−43 ([A1−A2]).

Second, the C-domain is likely to be directly involved in specific IGF-1R interactions through some of its side chains, especially Tyr31, which is one of the key residues in IGF-1R affinity (*30*). The positioning of Tyr31 in the center of the structurally well-defined Gly30···Ser33 type-II β-turn underlines the importance of this interaction; its probable hydrophobic character is emulated by van der Waals contacts between Tyr31 and the alkyl moiety of a symmetry related SB12 detergent molecule in the crystal. The structural role of the C-loop is underlined further by the high IGF-like activity of an insulin analogue with its A and B chains linked by the C-domain of hIGF-I (*36*, *37*) and very low biological potency of both hormones with C-domain modifications (e.g., refs *22*−*26* and *35*). The C-domain therefore contributes directly to the hIGF-1 receptor binding, and is at least in part responsible for the hIGF-I specificity. By contrast, the C-domain of pro-insulin does not even prevent its aggregation (*38*, *39*), and may be considered as a polypeptide whose main role is an assurance of the efficient folding and maturation of an active hormone (*40*, *41*).

In summary, our X-ray studies of the hIGF-I shed light on a structural role of the C-loop of this hormone in its functional divergence from insulin and underlined the need for further studies of the chemical integrity of hIGF-I in vitro and in vivo. The likelihood of the cleavage occurring at Arg36−Arg37 and the resulting flexibility of the C-loop delay any final conclusions about the hIGF-I "active" conformation in complex with IGF-1R.

MATERIALS AND METHODS

Crystallization. Pure hIGF-I was kindly supplied by Pharmacia-Upjohn (Stockholm, Sweden). The protein was crystallized by the hanging drop method, in which drops were composed of various ratios of hIGF-I at 7 mg/mL (in H_2O) with reservoir solution consisting of 0.1 M Tris/HCl pH 7.5, 12−15% (w/v) PEG 2K MME and 5 mM SB12 detergent. Platelike orthorhombic crystals with unit cell dimensions $a = 30.78$ Å, $b = 69.47$ Å, $c = 65.0$ Å appeared within 2−3 weeks, with one molecule of the hormone per asymmetric unit.

Data Collection. Initially, a room temperature 2.5 Å data set was collected in-house using an RaxisIIc imaging plate detector mounted on a Rigaku RU200 rotating anode X-ray generator with MSC/Yale mirrors. Subsequently, the resolution was extended to 2.3 Å at station W7B at DESY (EMBL, Hamburg) and station 14.2 at SRS (Daresbury). The final 100 K 2.0 Å data set was collected at station ID14-1 at the ESRF (Grenoble, France). Prior to freezing, the crystal was cryoprotected by sequential, 5% step, soaks in the mother liquor containing glycerol with a final concentration of 30% (v/v). Data were recorded in one sweep and 1° oscillations using a MAR Research CCD detector placed at 150 mm from the crystal. All data were integrated and reduced using DENZO and SCALEPACK (*42*). X-ray data statistics for this data set are summarized in Table 1.

Structure Determination and Refinement. The coordinates of the publicly available hIGF-I NMR structures (*2*−*4*) and its theoretical model (*43*) were initially used as search models in AmoRe (*44*) to solve the hIGF-I structure by molecular replacement. However, as these trials were unsuccessful subsequent molecular replacement trials were carried out with human insulin (PDB entry: 4INS, *18*) as the best search model. Careful analysis of the solutions (no single, clear solution was obtained) allowed determination of the position of the core of the hormone molecule. Initial electron density maps, calculated after rigid-body refinement in AmoRe, unambiguously indicated the position of all disulfides although the model building was rather difficult. The refinement with the use of an older version of REFMAC (*45*) was unsuccessful.

Refinement of hIGF-I was only possible using translation, libration, screw (TLS) (*46*), and individual atomic parameter (*45*) refinement. Each round of TLS refinement was followed by individual atomic (positional and thermal) refinement. Because of the high mobility of IGF, TLS refinement was crucial for the success of structure determination.

The successive rounds of refinement using all data between 15 and 2.0 Å (no sigma cutoffs), and manual rebuilding gave a final model with an R_{cryst} of 23.6 and an R_{free} of 29.5. The final model comprises 475 protein atoms, 20 ligand (SB12) atoms, and 35 water molecules. Poorly defined residues 1−2, 36−38, 67−70 were not included in the final structure, and occupancies of some side chain atoms of Gln40, Asp56, and Lys65 are set to zero because of their disorder. All model building was carried out using the molecular graphics package QUANTA (QUANTA98; Accelrys INC, San Diego, CA). Dictionaries for the SB12 ligand were derived from REFMAC. A summary of refinement statistics is given in Table 1.

Superpositions of different insulin and IGF models were carried out in QUANTA; after global superposition, the overlaps were fine-tuned using the "match closest residue" option in QUANTA. All NMR structures were minimized using CHARMM (*47*; details to be published elsewhere) prior to analysis. Figures 3c, 4, and 5 were produced using

QUANTA, and Figure 3a and 7 were produced with MOLVIEWER (M. Harsthorn, personal communication)

ACKNOWLEDGMENT

We would like to thank Par Gellerfors (Pharmacia-Upjohn, Stockholm) for protein samples and a very friendly collaboration. The technical assistance of the beam line managers at the ESRF, EMBL Hamburg, and SRS Daresbury during data collection is gratefully acknowledged. We also thank Steven Howell (NIMR, Mill Hill, London) and Arthur Moir (Department of Molecular Biology, University of Sheffield) for kind help with mass spectrometry and N-terminus sequencing analyses.

SUPPORTING INFORMATION AVAILABLE

ES-, MALDI-TOF spectroscopy, and SDS gel electrophoresis results; X-ray data and refinement statistics for other than hIGF esrf structures discussed in the text. This material is available free of charge via Internet at http://pubs.acs.org.

REFERENCES

1. Rinderknecht, E., and Humbel, R. E. (1978) The amino acid sequence of human insulin-like growth factor I and its structural homology with pro-insulin. *J. Biol. Chem. 253*, 2769−2776.
2. Cooke, R. M., Harvey, T. S., and Campbell, I. D. (1991) Solution structure of human insulin-like growth factor 1: a nuclear resonance and restrained molecular dynamics study. *Biochemistry 30*, 5484−5491.
3. Sato, A., Nishimura, S., Ohukuba, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Ysuda, T., and Kobayashi, Y. (1993) Three-dimensional structure of human insulin-like growth factor-1 (IGF-1) determined by ^{1}H NMR and distance geometry. *Int. J. Pept. Protein Res. 41*, 433−440.
4. Torres, A. M., Forbes, B. E., Aplin, S. E., Wallace, J. C., Francis, G. L., and Norton, R. S. (1995) Solution structure of human insulin-like growth factor II. Relationship to receptor and binding protein interactions. *J. Mol. Biol. 248*, 385−401.
5. Tersawa, H., Kohda, D., Hatanaka, H., Nagata, K., Higashihashi, N., Fujiwara, H., Sakano, K., and Inagaki, F. (1994) Solution structure of human insulin-like growth factor II.; recognition sites for receptors and binding proteins. *EMBO J. 13*, 5590−5597.
6. Vajdos, F. F., Ultsch, M., Schaffer, M. L., Deshayes, K. D., Liu, J., Skelton, J., and de Vos A. M. (2001) Crystal structure of human insulin-like growth factor-1: detergent binding inhibits Binding Protein interactions. *Biochemistry 40*, 11022−11029.
7. Żesławski, W., Beisel, H.-G., Kamionka, M., Kalus, W., Engh, R. A., Huber, R., Lang, K., and Holak, T. A. (2001) The interaction of insulin-like growth factor with N-terminal domain of IGFBP-5. *EMBO J. 20*, 3638−3644.
8. Dubaquie, Y., and Lowman, H. B. (1999) Total alanine-scanning mutagenesis of insulin-like growth factor I (IGF−I) identifies differential binding epitopes for IGFBP-1 and IGFBP-3. *Biochemistry 38*, 6386−6396.
9. Jansson, M., Uhlen, M., and Nilsson, B. (1997) Structural changes in insulin-like growth factor (IGF) I mutant proteins affecting binding kinetic rates to IGF binding protein 1 and IGF-I receptor. *Biochemistry 36*, 4108−4117.
10. Jansson, M., Andersson, Uhlén, M., Nilsson, B., and Kördel, J. (1998) The insulin-like growth factor (IGF) binding protein 1 binding epitope on IGF-I probed by heteronuclear NMR spectroscopy and mutational analysis. *J. Biol. Chem. 273*, 24701−24707.
11. Magee, B. A., Shooter, G. K., Wallace, J. C, and Francis, G. L. (1999) Insulin-like growth factor I and its binding proteins: a study of the binding interface using B-domain analogues. *Biochemistry 38*, 15863−15870.
12. Massague, J., and Czech, M. P. (1982) The subunit structures of two distinct receptors for insulin-like growth factors I and II and their relationship to the insulin receptor. *J. Biol. Chem. 257*, 5038−5045.
13. Collett-Sorberg, P. F., and Cohen, P. (2000) Genetics, chemistry, and function of the IGF/IGFBP system. *Endocrine 12*, 121−136.
14. Jones, J. I., and Clemmons, D. R. (1995) Insulin-like growth factors and their binding proteins: biological actions. *Endocrine Rev. 16*, 3−34.
15. Adams, T. E., Epa, V. C., Garret, T. P. J., and Ward, C. W. (2000) Structure and function of type 1 insulin-like growth factor receptor. *Cell. Mol. Life Sci. 57*, 1050−1093.
16. Roberts, C. T., Lasky, S. R., Lowe, W. L., Seaman, W. T., and LeRoith, D. (1987) Molecular cloning of rat insulin-like growth factor I complementary deoxyribonucleic acids: differential messenger ribonucleic acid processing and regulation by growth hormone in extrahepatic tissues. *Mol. Endocrinol. 1*, 243−248.
17. Laskowski, R. A., MacArthur, M. W., Moss, D. M., and Thornton, J. M. (1993) PROCHECK − a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr. 26*, 283−291.
18. Baker, E. N., Blundell, T. N., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Crowfoot Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N., and Vijayan, N. M. (1988) The structure of 2Zn pig insulin crystals at 1.5 Å resolution. *Philos. Trans. R. Soc. (London) 319*, 369−456.
19. Liu, X.-J., Xie, Q., Zhu, Y.-F., Chen, C., and Ling, N. (2001) Identification of a nonpeptide ligand that releases bioactive insulin-like growth factor-I from its Binding Protein complex. *J. Biol. Chem. 276*, 32419−32422.
20. Steiner, D. F., Cunningham, D., Spigelman, L., and Aten, B. (1967) Insulin biosynthesis: evidence for a precursor. *Science 157*, 697−700.
21. Jansen, J., Van Buul-Offers, S. C., Hoogerbrugge, C. M., and Van Den Brande, J. L. (1990) Effects of a single cleavage in insulin-like growth factors I and II on binding to receptors, carrier proteins and antibodies. *Biochem. J. 266*, 513−520.
22. Cutfield, J., Cutfield, S., Dodson, E. J., Dodson, G. G., Hodgkin, D., and Reynolds, C. D. (1981) Evidence concerning insulin activity from the structure of a cross-linked derivative. *Hoppe-Seyler's Z. Physiol. Chem. 362*, 755−761.
23. Derewenda, U., Derwenda, Z., Dodson, E. J., Dodson, G. G., Bing, X., and Markussen, J. (1991) X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. *J. Mol. Biol. 220*, 425−433.
24. Derewenda, U., Derwenda, Z., Dodson, E. J., Dodson, G. G., Reynolds, C. D., Smith, G. D., Sparks, C., and Swenson, D. (1989) Phenol stabilizes more helix in new symmetrical zinc insulin hexamer. *Nature 338*, 594−596.
25. Whittingham, J. L., Edwards, D. J., Antson, A. A., Clarkson, J., and Dodson, G. G. (1998) Interactions of phenol and *m*-cresol in the insulin hexamer, and their effects on the association properties of B28Pro → Asp insulin analogues. *Biochemistry 37*, 11516−11523.
26. Cho, Y. S., Chang, S. G., Choi, K. D., Shin, H. C., Ahn, B. Y., and Kim, K. S. (2000) Solution structure of an active mini-proinsulin, M2PI: interchain flexibility is crucial for insulin activity. *J. Biochem. Mol. Biol. 33*, 120−125.
27. Schäffer, L. (1994) A model for insulin binding to the insulin receptor, *Eur. J. Biochem. 221*, 1127−1132.
28. Dodson, E. J., Dodson, G. G., Hubbard, R. E., and Reynolds, C. D. (1983) Insulin's structural behaviour and its relation to activity. *Biopolymers 22*, 281−291.
29. Cascieri, M. A., Chicchi, G. G., Applebaum, J., Hayes, N. S., Green, B. G., and Bayne, M. L. (1988) Mutants of human insulin-like growth factor I with reduced affinity for the type 1 insulin-like growth factor receptor. *Biochemistry 27*, 3229−3233.
30. Bayne, M. L., Applebaum, J., Chicchi, G. G., Miller, R. E., and Cascieri, M. A. (1990) Role of tyrosines 24, 31 and 60 in the high affinity binding of insulin-like growth factor-1 to the type 1 insulin-like growth factor receptor. *J. Biol. Chem. 265*, 15648−15652.
31. Ludwigsen, S., Olsen, H. B., and Kaarlshom, N. C. (1998) A structural switch in mutant insulin exposes key residues for receptor binding. *J. Mol. Biol. 279*, 1−7.
32. Hua, X. Q., Shoelson, S. E., Kochoyan, M., and Weiss, M. A. (1991) Receptor binding redefined by a structural switch in a mutant human insulin. *Nature 354*, 238−241.
33. Gill, R., Wallach, B., Verma, C., Ursø, B., De Wolf, E., Grötzinger, J., Murray-Rust, J., Pitts, J., Wollmer, A., De Meyts, P., and Wood, S. (1996) Engineering the C-region of human insulin-like growth factor-1: Implications for receptor binding. *Protein Eng. 9*, 1011−1019.

34. Ciszak, E., Beals, J. M., Frank, B. H., Baker, J. C., Carter, N. D., and Smith, G. D. (1995) Role of C-terminal B chain residues in insulin assembly: the structure of hexameric $Lys^{B28}Pro^{B29}$-human insulin. *Structure 3*, 615−622.
35. Bayne, M. L., Applebaum, J., Underwood, D., Chicchi, G. G., Green, B. G., Hayes, N. S., and Cascieri, M. A (1989) The C region of human insulin-like growth factor (IGF) I is required for high affinity binding to the type 1 IGF receptor. *J. Biol. Chem. 264*, 11004−11008.
36. Kristensen, J., Andersen, A. S., Hach, M., Wiberg, F. C., Schäffer, L., and Kjeldsen, T. (1995) The single chain insulin-like growth factor I/insulin hybdrid binds with high affinity to the insulin receptor. *Biochem. J. 305*, 981−986.
37. Cara, J. F., Mirmira, R. G., Nakagawa, S. H., and Tager, H. S. (1990) An insulin-like growth factor I/insulin hybrid exhibiting high potency for interactions with type I insulin-like growth factor and insulin receptors of placental plasma membranes. *J. Biol. Chem. 265*, 17820−17825.
38. Steiner, D. F. (1973) Cocrystallisation of proinsulin with insulin. *Nature 243*, 528−530.
39. Pekar, A. H., and Frank, B. H. (1972) Conformation of proinsulin. A comparison of insulin and proinsulin self-association at neutral pH. *Biochemistry 11*, 4013−4016.
40. Lipkind, G., and Steiner, D. F. (1999) Predicted structural alternations in proinsulin during its interactions with prohormone convertases. *Biochemistry 38*, 890−896.
41. Qiao, Z.-S., Guo, Z.-Y., and Feng, Y.-M. (2001) Putative disulfide-forming pathway of porcine insulin precursor during its refolding in vitro. *Biochemistry 40*, 2662−2668.
42. Otwinowski, Z., and Minor, W. (1997) Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol. A 276*, 307−326.
43. Blundell, T. L., and Humbel, R. E. (1980) Hormone families: pancreatic hormones and homologous growth factors. *Nature 287*, 781−787.
44. Collaborative Computational Project No. 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D50*, 760−763.
45. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallogr. D53*, 240−255.
46. Winn, M. D., Isupov, M. N, and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacement in macromolecular refinement. *Acta Crystallogr. D57*, 122−133.
47. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM − a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187−217.

BI020084J