The complete amino acid sequence of human insulin-like growth factor I (IGF-I), a polypeptide isolated from serum, has been determined. IGF-I is a single chain polypeptide of 70 amino acid residues cross-linked by three disulfide bridges. The calculated molecular weight is 7649. IGF-I displays obvious homology to preproinsulin; positions 1 to 29 are homologous to insulin B chain and positions 42 to 62 to insulin A chain. A shortened "connecting" peptide with 12 residues (positions 30 to 41) compared to 30 to 35 in proinsulins shows no homology to proinsulin C peptide. An octapeptide sequence at the COOH-terminal end is also a feature not found in proinsulins. The number of differences in amino acid positions between IGF-I and insulins suggests that duplication of the gene of the common ancestor of proinsulin and IGF occurred before the time of appearance of the vertebrates.

The survival and growth of most types of untransformed cells of vertebrates in vitro is dependent on serum (1). Despite well documented effects of serum on cellular growth (2), its physiological role remains uncertain (3, 4, 5). Several lines of evidence point to a physiological role of IGF as a growth-promoting substance (6, 7, 8, 9). Indeed, immunoreactivity of plasma IGF-I in man show a correlation with levels of circulating growth hormone (6, 7) which justifies the assignment of IGF to a new class of hormones called somatomedins (8, 9). We now describe the complete amino acid sequence of IGF-I.

**EXPERIMENTAL PROCEDURES**

IGF-I was obtained from a Cohn fraction of human plasma as described previously (6). The polypeptide was reduced with dithiothreitol in 6 M guanidine-HCl, 0.1 M Tris-HCl at pH 9.5 for 4 h and either 5-pyridylethylated with 4-vinylpyridine (19) or carboxymethylated with iodoacetate (20). For aminooxylation with ethylenediamine the protein was reduced in 8 M urea, 0.35 M Tris-HCl at pH 8.6 (21). a-Chymotrypsin and trypsin (treated with tosylation) were the most highly purified preparations available from Worthington. Protease from Staphylococcus aureus V8 was purchased from Miles, thermolysin from Serva, carboxypeptidase C and aminopeptidase M from Röhm/Roth, and Aeromonas aminopeptidase was a generous gift of Dr. Ken Wilson (Biochemisches Institut der Universität Zürich). Reagents for automatic Edman degradation were purchased from Beckman as Sequencial Grade, except dimethylbenzylation and propargylol which were from Pierce. Ion exchange resin M-72 was obtained from Beckman and Sephadex from Pharmacia.

For tryptic digestion of carboxymethylated IGF-I, 2% by weight of trypsin in water at pH 8.0 in a pH-stat (Radiometer, Copenhagen) was used for 4 h at 37°C. Tryptic digestions of aminooxyalted chymotryptic or staphylococcal protease peptides were carried out in 0.1 M ammonium bicarbonate, pH 8.5, using 1 to 2 mol% of trypsin for 2 h at 37°C. For chymotryptic digestions, 3% by weight and 1 to 5 mol% of chymotrypsin were used for aminooxyalted IGF-I and for staphylococcal protease peptides, respectively, for 2 h at 37°C in 0.1 M bicarbonate, pH 8.5. Digestions with staphylococcal protease of aminooxyalted IGF-I, of trypsin and of chymotryptic peptides were carried out in 0.1 M ammonium bicarbonate, pH 7.8, using 6% methionine.

The abbreviations used are: IGF-I and IGF-II, insulin-like growth factor I and II; NSILA, nonsuppressible insulin-like activity; Cys(Ae), aminoethyl-cysteine; Cys(Cm), carboxymethyl-cysteine; PTH, phenylthiohydantoin; Tricine, N-Tris(hydroxymethyl)-methylglycine; dansyl, 5-dimethylaminonaphthalene-1-sulfonyl.

1 The abbreviations used are: IGF-I and IGF-II, insulin-like growth factor I and II; NSILA, nonsuppressible insulin-like activity; Cys(Ae), aminoethyl-cysteine; Cys(Cm), carboxymethyl-cysteine; PTH, phenylthiohydantoin; Tricine, N-Tris(hydroxymethyl)-methylglycine; dansyl, 5-dimethylaminonaphthalene-1-sulfonyl.

2 IGF is the new designation for insulin-like activity not suppressible by antibodies against insulin (NSILA) (12, 13).
by weight of enzyme for 24 h at 37°C. Thermolytic digestions of chymotryptic and staphylococcal protease peptides were done under the same conditions as for trypptic digestion of these peptides. COOH- and NH₂-terminal sequences of the intact protein and of isolated peptides were determined using carboxypeptidase C, aminopeptidase M, and Aeromonas aminopeptidase in 0.05 M sodium acetate at pH 5.3, 0.1 M ammonium bicarbonate, pH 8.5, and 10 mM Tricine-HCl, pH 8.0, respectively.

Cyanogen bromide cleavage followed the procedure of Gross (22). Cleavage at aspartic acid residues in 0.25 M acetic acid was performed at 106°C in an evacuated sealed Pyrex tubes for 18 h (23).

Cyanogen bromide peptides were separated on Sephadex G-50 (0.9 x 150 cm) in 50% aqueous acetic acid and then further fractionated on Beckman M-72 ion exchange resin (0.9 x 20 cm) equilibrated in 0.05 M pyridine/acetate, pH 2.5, at 55°C and developed with a four-chamber gradient containing 200 ml of 0.05 M pyridine/acetate, pH 2.5, 200 ml of 0.2 M pyridine/acetate, pH 3.1, 200 ml of 0.5 M pyridine/acetate, pH 3.75, and 200 ml of 2.0 M pyridine/acetate, pH 5.0. The separation of peptides was monitored with a Technicon AutoAnalyzer equipped for automatic alkaline hydrolysis and ninhydrin reaction. Fractions were evaporated under reduced pressure and filtered over glass wool.

The thiazolinones were converted to the phenylthiohydantoins for sequencing and the miniprint supplement. The purity of the peptides was checked by high voltage paper electrophoresis at pH 6.5, or both, and by amino acid analysis.

The amino acid compositions of the 14 chymotryptic peptides and of the eight tryptic peptides are shown in Tables II and X. Digestion of reduced and aminomethylated IGF-I by chymotrypsin led unexpectedly to some hydrolysis of an Arg-Arg bond at position 36-37 (Fig. 1, and Table II in the miniprint section). Another atypical cleavage site was found with reduced and carboxymethylated IGF-I: trypsin cleaved a Tyr—Cys(Cm) bond at position 60-61.

The amino acid composition of the seven staphylococcal protease (SP) peptides is listed in Table VI. An unusual cleavage site was the Gly—Ser bond at position 32-33 (Fig. 1, and Tables IX and VI in the miniprint section). Besides these identified peptides there were some heterogeneous peptides in low yields which presumably are due to atypical and incomplete cleavage by this enzyme. Cleavage by CNBr led to two fragments due to the presence of a single methionine residue.

The amino acid composition of peptide CNBr-2 is shown in Table IX. Fig. 3 illustrates the peptide overlaps which allowed ordering of the different peptides.

NH₂-terminal Sequence (residues 1 to 31) — The amino acid sequence of residues 1 to 31 was determined by automatic Edman degradation of the S-aminomethylated and of the S-pyridylethylated derivative of IGF-I and has been reported in a preliminary communication (12). Full documentation of these data is given in Table XIV, b and c. As an example of repetitive yields and of background amino acids during Edman degradations, the data on the longest run in this work (31 cycles on pyridylethylated IGF-I) are presented in Table XIVc. The repetitive yields were between 90% and 96%. Amino acids originating from carry-over were always present in sizable amounts and gradually increased during the run (e.g. 14% at Step 3, 72% at Step 25). This amount of carry-over did not prevent positive identification of any of the 31 residues. In most other cases of Edman degradations on shorter peptides, carry-over was less pronounced, e.g. in peptide C-3, the carry-over at Step 2 was 6% (data not shown).

Confirmation of the sequence 1 to 31 was obtained by compositional analysis, by Edman degradation and by digestion with amino- and carboxypeptidases of Fragments T-1, T-2, C-1, C-2, C-1A, C-3, C-4, SP-1, and SP-2 (Fig. 1 and Tables II, VI, XIV, a and b, XV, a and b).

Amino Acid Sequence of Residues 32 to 50 — Although peptides T-2, SP-3, and CNBr-1 all cross residues 31-33, there is no Edman degradation or exopeptidase digestion step which actually cleaved this bond. The identification of this bond (31-32) is thus founded solely on the compositional overlaps of peptides T-2, SP-3, and CNBr-1 (Tables VI, IX, and X). Gly 32 was found by dansylation to be the NH₂-terminal residue of both peptides C-5 and C-5A. Residues 33 to 44 could be sequenced by Edman degradation of peptide SP-3A (Table XIVb). The sequence of residues 37 to 48 could be derived from Edman degradation (Table XIVa) and position 49-50 from carboxypeptidase digestion (Table XVb) of peptide T-3 (Table XI). In complete agreement with the structure given in Fig. 1 are the results obtained on the following peptides: C-5, C-5A, C-5B, C-5C, C-6 (Tables II, III, XVa), SP-4 (Table VII), and T-2 (Table XVb).

Amino Acid Sequence of Residues 51 to 70 — The overlap 50-51 is provided by two peptides, C-7 and SP-4 (Tables IV and VII). Peptide C-7 has an NH₂-terminal Arg-Ser sequence as determined by aminopeptidase digestion (Table XVa). The same Arg-Ser sequence can be derived from the amino acid compositions of peptides SP-4 C-2, SP-4 T-2, and SP-4 T-4 (Table VII). Residues 51 to 55 and 56 to 60 were derived from analysis of peptides T-4 and T-5 (Tables XIVa, XVb, and XII). The overlap 56-58 (Arg-Arg) was obtained from amino peptidase digestion of peptide C-8 (Table XVa), from amino acid compositions of C-8-derived peptides (Table V) and of SP-5-
Amino Acid Sequence of Insulin-like Growth Factor I

FIG. 1. Amino acid sequence of human IGF-I and a schematic outline of the data supporting the sequence. Peptides are indicated by solid lines under the residues comprising the peptide. The letter code indicates the cleavage method used to produce the peptide: T, trypsin; C, chymotrypsin; SP, staphylococcal protease; CNBr, cyanogen bromide. Sequence data obtained by automatic Edman degradation of the intact protein (31 residues, Ref. 12) are not included in Fig. 1 but shown in Table XIVb. To indicate the method of sequential degradation and of identification of residues, the following symbols are used: →, amino acid analysis; →, digests with trypsin, chymotrypsin, and thermolysin; †, cyanogen bromide; ? digestion with aminopeptidase M or aminopeptidase from Aeromonas; ‡, dansylation and thin layer chromatography; • digestion with carboxypeptidase C and amino acid analysis. These symbols are also used in combination, for example, (●) denotes a residue that was identified after Edman degradation by carboxypeptidase digestion and amino acid analysis. Amides—The amides were determined by at least two of the following methods: electrophoretic mobility of the peptides at pH 6.5, direct identification of the PTH-derivatives by thin layer chromatography, and amino acid analysis following hydrolysis by amino- or carboxypeptidase. A complete listing of the methods used for the assignment of each of the three amides and the eight free acids is given in Table XVI.

Discussion

The sequence analysis of IGF-I was hampered by the fact that only small amounts of material were at our disposal. The 3 µmol of IGF used for the sequence represent the yield from about 3 tons of a Cohn fraction (precipitate B) from human plasma (6). Nevertheless, the position of all 70 residues could be determined in at least two sets of different peptides, the only exception being Val 11 and Asp 12. These 2 residues were determined only by Edman degradation (three runs) of IGF-I. The position of all amides and free acids was also deduced by at least two different methods. The amino acid composition of the sum of all consecutive tryptic peptides (T-1 through T-7, without peptide T-5A) is the same as the one deduced from the sequence. Likewise, the amino acid compositions of the sums of all consecutive chymotryptic peptides (C-1 through C-10) and the staphylococcal protease peptides SP-1 through SP-6) are identical. Based on earlier determinations (6), the 6 half-cystine residues must be present as 3 cystine residues. The assignment of the three disulfide cross-links is currently under investigation.

The amino acid sequence of IGF-I (Fig. 1) indicates a polypeptide chain of 70 amino acid residues for the monomer. The calculated molecular weight is thus 7649 (Table XIII) and not 5700 as determined earlier by molecular weight measurements (6). A new estimation of the molecular weight by Megyesi et al. (30) of 111 IGF on Sephadex G-50 in 0.05 M ammonium carbonate, pH 8.6, is compatible with this value. The reason for the discrepancy in the previously determined molecular weight by a factor of 1.33 is not clear. The relative amino acid composition given in the earlier publication is consistent with that from the sequence study except that estimations of cysteic acid and tyrosine are low by 1 residue and glutamic acid is high by 1 residue (Table XIII).
Amino Acid Sequence of Insulin-like Growth Factor I

The previously published NH₂-terminal amino acid sequence (31 residues) of IGF-I displayed a striking homology with insulin B chain (12). The complete amino acid sequence as presented now extends this homology further to the A chain part of proinsulin. As Fig. 2 illustrates, 25 out of 51 residues in the sequence of human insulin are identical in IGF-I. The 9 residues which are invariant in all insulin B chains so far sequenced (35) are conserved in IGF-I (B6-8, B11-12, B15, B19, B23-24). Out of the 10 invariant residues of insulin A chains (A1-2, A5-7, A11, A16, A19-21), only glutamine A5 is replaced by glutamic acid and asparagine A21 by alanine.

In human proinsulin, a peptide 35 residues long is intercalated between the A and B chain parts. Such a "connecting peptide" is also found in IGF-I, but it is only 12 residue long (position 30-41) and shows no obvious sequence similarity to proinsulin C peptide. Sets of 2 basic residues at the amino and carboxyl end of the C-peptide serve as cleavage sites in the conversion of proinsulin to insulin. In IGF-I, these basic residues have been substituted by two Thr-Gly bonds (positions 29-30 and 41-42). This may account for the fact that IGF, which apparently lacks this ability. As postulated by Steiner and by De Haen et al. (36, 37) for proinsulin, IGF might be a descendant of an old serine protease, assuming the function of stimulating cell and tissue growth after food intake. This function may have included some "insulin-like" activity which was then further refined when proinsulin emerged from IGF by gene duplication and subsequent mutation.

The only other two peptides for which a distant relationship to insulin has been discussed are nerve growth factor (38) and relaxin (39). A comparison of the sequence of these peptides with IGF-I shows that the sequence similarity between nerve growth factor and IGF-I is even smaller than the one between nerve growth factor and insulin. IGF can thus not serve as a link in the disputed descendence of nerve growth factor from an insulin ancestor. The degree of similarity between porcine relaxin and IGF-I is numerically the same as between relaxin and human insulin (12 identical amino acid positions). This fact gives support to the view that relaxin and insulin do have an evolutionary relationship.

A brief survey of the amino acid positions common to insulin and IGF-I shows the following main features: 1) the position of all half-cystine residues is conserved; 2) all but one of the nonpolar residues of the insulin B chain helix B9-19 which are buried inside the monomer are conserved, i.e. Leu B11, either proinsulin or IGF in phyla other than vertebrates, one can only speculate whether proinsulin is an evolutionary offshoot of IGF or vice versa. The distinctive feature of proinsulin of being activated by limited proteolysis suggests to us that proinsulin might be a more recent product of evolution than IGF, which apparently lacks this ability. As postulated by Steiner and by De Haen et al. (36, 37) for proinsulin, IGF might be a descendant of an old serine protease, assuming the function of stimulating cell and tissue growth after food intake. This function may have included some "insulin-like" activity which was then further refined when proinsulin emerged from IGF by gene duplication and subsequent mutation.

The only other two peptides for which a distant relationship to insulin has been discussed are nerve growth factor (38) and relaxin (39). A comparison of the sequence of these peptides with IGF-I shows that the sequence similarity between nerve growth factor and IGF-I is even smaller than the one between nerve growth factor and insulin. IGF can thus not serve as a link in the disputed descendence of nerve growth factor from an insulin ancestor. The degree of similarity between porcine relaxin and IGF-I is numerically the same as between relaxin and human insulin (12 identical amino acid positions). This fact gives support to the view that relaxin and insulin do have an evolutionary relationship.

A brief survey of the amino acid positions common to insulin and IGF-I shows the following main features: 1) the position of all half-cystine residues is conserved; 2) all but one of the nonpolar residues of the insulin B chain helix B9-19 which are buried inside the monomer are conserved, i.e. Leu B11,
Val B12, Ala B14, Leu B15, and Val B18, the one exception being position B17 which is phenylalanine in IGF-I instead of leucine; 3) the nonpolar interior residues of insulin A chain are likewise all conserved in IGF-I, i.e. Cys A6-11, Leu A16 and Ile A2; and 4) the 3 glycyl residues B8, B20, and B23, allowing sharp turns of the insulin B chain, are conserved in IGF-I. All these features are crucial for the maintenance of the three-dimensional structure of insulin (40) and may, therefore, indicate that IGF be folded in a way similar to the insulin monomer. On the other hand, the surface residues of the insulin dimer are mostly substituted in IGF-I, a fact which may well explain the lack of reactivity of IGF with antibodies against insulin (13). A full discussion of a model of IGF-I in relation to the three-dimensional structure of insulin will be published elsewhere.4

Several other growth-promoting factors obtained from serum or from serum-free conditioned medium have been described such as the somatomedins (5, 18) or multiplication-stimulating activity (41, 42). Since the primary structure of none of these polypeptides is as yet known, any speculations about possible relationships to IGF must be based on indirect evidence such as functional similarities or competition for binding sites on target cell receptors. Indeed, comparative studies of this sort have been carried out (43, 44) which suggest that at least one of the somatomedins and multiplication-stimulating activity might be related to IGF. The final proof, however, for such a family of insulin-related growth factors will have to wait until structural comparisons become feasible.

The mode of biosynthesis of IGF is an open question and the site of its biosynthesis a riddle which hopefully can be investigated now. The availability of a serum growth factor of defined primary structure should furthermore facilitate studies aimed at the elucidation of the biochemical mechanisms of growth control in animal cells. Last but not least, the known primary structure of IGF-I offers the possibility of producing it synthetically in sufficient quantities for long term in vitro studies in growth-deficient states such as pituitary dwarfism or during repair processes such as wound healing.

Acknowledgments—We thank Lotte Braun, Edwin Schwander, and Christina Kasper for technical assistance. Dr. Kenneth J. Wilson for the supply of Aeromonas aminopeptidase, Professor Tom Blundell for discussions, the staff of this institute for critically reviewing the manuscript, and Christine Berger, Silvia Kocher, and Anna-Maria Mosca for the preparation of the illustrations and of the manuscript.

REFERENCES

34. IUPAC-IUB Commission on Biochemical Nomenclature (1968) J. Biol. Chem. 243, 3557-3559
## Amino Acid Sequence of Insulin-like Growth Factor I

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Residue</th>
<th>Location</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-1</td>
<td>Met</td>
<td>1-10</td>
<td>Met [22, 23, 24]</td>
</tr>
<tr>
<td>C-2</td>
<td>Phe</td>
<td>11-19</td>
<td>Phe [25, 26, 27]</td>
</tr>
<tr>
<td>C-3</td>
<td>Asn</td>
<td>20-29</td>
<td>Asn [28, 29, 30]</td>
</tr>
<tr>
<td>C-4</td>
<td>Glu</td>
<td>30-39</td>
<td>Glu [31, 32, 33]</td>
</tr>
</tbody>
</table>

*The sequence was determined by Edman degradation and mass spectrometry.*

---

### Table II: Qualitative data on sequence comparisons

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Residue</th>
<th>Location</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIPC</td>
<td>Met</td>
<td>1-10</td>
<td>Met [22, 23, 24]</td>
</tr>
<tr>
<td>CIPC</td>
<td>Phe</td>
<td>11-19</td>
<td>Phe [25, 26, 27]</td>
</tr>
<tr>
<td>CIPC</td>
<td>Asn</td>
<td>20-29</td>
<td>Asn [28, 29, 30]</td>
</tr>
<tr>
<td>CIPC</td>
<td>Glu</td>
<td>30-39</td>
<td>Glu [31, 32, 33]</td>
</tr>
</tbody>
</table>

*The sequence was determined by Edman degradation and mass spectrometry.*

---

### Table III: Quantitative data on sequence comparisons

- **Table II:** Table II provides qualitative data on sequence comparisons for the insulin-like growth factor I (IGF-I) sequence. The data includes the location and sequence for each residue, as determined by Edman degradation and mass spectrometry.

- **Table III:** Table III offers quantitative data on sequence comparisons. This table includes columns for peptide, residue, location, and sequence, providing a more detailed analysis of the sequence data.

---

### Table IV: Peptide Assignment for IGF-I

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIPC</td>
<td>Met [22, 23, 24]</td>
</tr>
<tr>
<td>CIPC</td>
<td>Phe [25, 26, 27]</td>
</tr>
<tr>
<td>CIPC</td>
<td>Asn [28, 29, 30]</td>
</tr>
<tr>
<td>CIPC</td>
<td>Glu [31, 32, 33]</td>
</tr>
</tbody>
</table>

*The assignments were made based on Edman degradation and mass spectrometry.*

---

**Fig. 1:** A representative summary of overlapping fragments of IGF-I.
The amino acid sequence of human insulin-like growth factor I and its structural homology with proinsulin.
E Rinderknecht and R E Humbel


Access the most updated version of this article at http://www.jbc.org/content/253/8/2769

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

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/253/8/2769.full.html#ref-list-1