se produirait, en fait, au niveau d'un précurseur tétracyclique (du type dammarenediol, par exemple), c'est-à-dire avant la formation du cycle E.[6].

[6] C. DJERASSI (*Cactus Triterpenes*, Festschrift ARTHUR STOLL 1957, p. 330) avait déjà considéré l'hypothèse d'une relation intime entre l'oxygénation et le processus de formation du squelette.

Summary. Comparison of the known tetracyclic and β-amyrenic triterpenes shows that natural oxygenation of the skeleton occurs preferentially at certain carbon atoms that are biogenetically equivalent in the 2 series.

B. TURSCH et J. LECLERCQ

Service de Chimie Organique E.P., Université Libre de Bruxelles (Belgique), 10 novembre 1967.

STUDIORUM PROGRESSUS

On the Relationship Between Glucagon and Secretin

In recent years considerable effort has been devoted to an analysis of the observed alterations of amino acid residues in specific, biologically active hormones, enzymes and proteins as a guide to the zoological cognation and the phylogenetic distance between 2 species[1–4]. Such techniques are dependent upon comparisons between homologous proteins and have been greatly aided by studies with cytochrome *c*[5–8], hemoglobin[8–11] and various proteolytic enzymes[12]. A similar approach is made at this time for the important polypeptides glucagon and secretin.

Glucagon was originally isolated as a byproduct of insulin research[13]. The marked hyperglycemic action of the material led to the name, which means 'mobilizer of sugar'[14]. Later, sufficient amounts of crystalline glucagon were obtained from the α-cells of the pancreas[15] and were used to determine both the composition and the amino acid sequence of the hormone[16]. Several analytical methods for the hormone are available and these serve as useful adjuncts in the treatment of glycogen storage diseases by glucagon[17,18]. The chemical and medical applications of glucagon have been reviewed recently[19,20].

Secretin possesses a shorter history, but due to the marked advancement in the use of physical instrumentation, the identification[21], isolation[22] and structure determination[23] of the material proceeded at a rapid pace. This new hormone is found in the cell walls of the duodenum and acts by stimulating the pancreas to release sodium bicarbonate, hence the name. Within the past year the complete structure of the compound was announced[24] and was followed in turn by a complete synthesis[25]. Contemporary developments in this field have been summarized, too[26].

The nonacosapeptide sequence of glucagon (porcine) bears a remarkable resemblance to the heptacosapeptide sequence of secretin (porcine) (see Scheme 1). Indeed, there are 15 invariant and positionally identical amino acids in both compounds: A_1, A_2, A_4, A_5, A_6, A_7, A_8, A_{11}, A_{15}, A_{16}, A_{18}, A_{20}, A_{24} and A_{26}. Application of the 'index of similarity'[27] to secretin and the first 27 residues of glucagon gives a sum of $^{14}/_{27}$ or 0.52. The result definitely suggests a common origin for the 2 compounds, probably stemming from the process of genetic duplication. A structural parallelism was detected earlier in a computer comparison of the residues in glucagon (bovine?) and secretin (porcine)[28] and it was noted that the pancreas and duodenum have a common embryonic background, a fact of some evolutionary significance[29].

[1] E. ZUCKERKANDL and L. PAULING, *Evolving Genes and Proteins* (Academic Press, New York 1965), p. 97.
[2] R. ARCHER, Angew. Chem. Int. Ed. *5*, 798 (1966).
[3] M. FLORKIN, *A Molecular Approach to Phylogeny* (Elsevier, Amsterdam 1966).
[4] T. H. JUKES, *Molecules and Evolution* (Columbia University Press, New York 1966).
[5] E. MARGOLIASH and E. L. SMITH, *Evolving Genes and Proteins* (Academic Press, New York 1965), p. 221.
[6] E. MARGOLIASH and A. SCHEJTER, Adv. Protein Chem. *21*, 113 (1966).
[7] C. R. CANTOR and T. J. JUKES, Biochem. biophys. Res. Commun. *23*, 319 (1966).
[8] C. R. CANTOR and T. J. JUKES, Proc. natn. Acad. Sci., U.S. *56*, 177 (1966).
[9] J. BUETTNER-JANUSCH and R. L. HILL, *Evolving Genes and Proteins* (Academic Press, New York 1965), p. 167.
[10] E. ZUCKERKANDL, *Protides of the Biological Fluids* (Elsevier, Amsterdam 1965), p. 102.
[11] W. M. FITCH, J. molec. Biol. *16*, 17 (1966).
[12] H. NEURATH, K. A. WALSH and W. P. WINTER, Science *158*, 1638 (1967).
[13] C. B. F. GIBBS, E. W. ROOT JR. and J. R. MURLIN, Q. Jl exp. Physiol. *13*, 128 (1923).
[14] C. P. KIMBALL and J. R. MURLIN, J. biol. Chem. *58*, 337 (1923).
[15] A. STAUB, L. E. SINN and O. K. BEHRENS, J. biol. Chem. *214*, 619 (1955).
[16] W. W. BROMMER, L. G. SINN and O. K. BEHRENS, J. Am. chem. Soc. *79*, 2807 (1957).
[17] W. W. BROMMER and O. K. BEHRENS, *Methods in Hormone Research* (Academic Press, New York 1962), p. 459.
[18] E. W. PROBST and R. W. COLWELL, Biochemistry *5*, 1209 (1966).
[19] O. K. BEHRENS and W. W. BROMER, Vitams Horm. *16*, 263 (1958).
[20] P. P. FOÀ and E. GALASINO, *Glucagon: Chemistry and Function in Health and Disease* (Charles C. Thomas, Springfield 1962).
[21] J. E. JORPES and V. MUTT, Ann. intern. Med. *55*, 395 (1961).
[22] J. E. JORPES and V. MUTT, Acta chem. scand., *15*, 1790 (1961).
[23] V. MUTT, S. MAGNUSSON, J. E. JORPES and E. DAHL, Biochemistry *4*, 2358 (1965).
[24] V. MUTT and J. E. JORPES, Fourth Int. Symp. Chem. Nat. Products, Stockholm (1966).
[25] M. BODANSZKY and N. J. WILLIAMS, J. Am. chem. Soc. *89*, 685 (1967).
[26] V. MUTT and J. E. JORPES, Recent Prog. Horm. Res. *23*, 483 (1967).
[27] A. VEGOTSKY and S. W. FOW, *Comparative Biochemistry* (Academic Press, New York 1962), vol. 4, p. 185.
[28] R. V. ECK and M. O. DAYHOFF, *Atlas of Protein Sequence and Structure 1966* (National Biomedical Research Foundation, Silver Spring 1966), pp. 89, 108, 109, 195.
[29] B. M. PATTEN, *Human Embryology* (The Blakiston Company, New York 1953), p. 479.

Seven of the variant amino acid substitutions in glucagon and secretin – A_9, A_{12}, A_{14}, A_{22}, A_{23}, A_{25} and A_{27} – are equivalent on a 'conservative' basis. The term 'conservative' has been used for classes of residues considered to be mutually interchangeable, in which similar chemical structures can fit the same functional or steric role[30]. Such a definition is used here, but with the additional restriction of only a single base mutation being allowed in the 'conservative' substitution pattern[31,32]. A common twofold degeneracy (only 2 coding triplets apiece)[33] is maintained by sA_9(glu)=gA_9(asp), a sixfold-twofold change occurs in sA_{12}(arg)=gA_{12}(lys) and sA_{22}(leu)=gA_{22}(phe), a sixfold-fourfold shift appears with sA_{23}(leu)=gA_{23}(val), and a fourfold-singlefold change is seen in both sA_{25}(gly)=gA_{25}(trp) and sA_{27}(val)=gA_{27}(met), but sA_{14}(arg)=gA_{14}(leu) is the sole example of a sixfold-sixfold degeneracy. Six more residues – A_3, A_{10}, A_{13}, A_{17}, A_{19} and A_{21} – fall into the 'radical' pattern, a classification generalized to include those amino acids whose side chains are entirely different from a chemical viewpoint. A further restraint is inserted now by requiring a mandatory 2 point codon change for each 'radical' interchange. Close attention to the known triplet code combinations reveals several interesting details for this last group in the secretin-glucagon system. For example, sA_3(asp)=gA_3(gln) involves a twofold degeneracy, sA_{10}(leu)=gA_{10}(tyr), sA_{13}(leu)=gA_{13}(tyr) and sA_{21}(arg)=gA_{21}(asp) constitute a sixfold-twofold degeneracy, while sA_{17}(ala)=gA_{17}(arg) and sA_{19}(leu)=gA_{19}(ala) provide a sixfold-fourfold degenerate change. From the above data, it seems that the amino acid variations arose stepwise as follows: asp → glu → gln or asp ← glu ⇄ gln; leu → phe → tyr or leu ← phe ⇄ tyr; ala → gly → arg or ala ← gly ⇄ arg; leu → val → ala or leu ← val ⇄ ala; and arg → gly → asp or arg ← gly ⇄ asp. The proposed alignment in terms of possible base changes per codon amounts to $^{19}/_{27}$ or 0.70. For comparison, cytochrome *c* (human vs. horse) is 0.13 and β-hemoglobin (human vs. horse) is 0.27[4]. A random deviation value of 1.59 is obtained by simply relocating the entire secretin chain 1 residue to the right under the glucagon sequence and then recomputing the base changes. The terminal valinamide group in secretin, sA_{27}(val-NH_2), is equivalent to valine for this analysis. The last 2 residues in glucagon, gA_{28}(asn) and gA_{29}(thr), cannot be discussed from a base change viewpoint, since comparable units are lacking in secretin. Glucagons and secretins from widely varying extant species, using the conclusions drawn here, would be predicted to possess such changes as A_3(glu), A_{10}(phe), A_{13}(phe), A_{17}(gly), A_{19}(val) and A_{21}(gly).

Alternatively, a different fit between glucagon and secretin is found on rearranging the residue sequences in the mutual A_{13}-A_{14} area to yield 2 artificial gaps (see Scheme 2). The 2 leucyl residues at gA_{14} and at sA_{13} are immediately superimposable, so the 'index of similarity' improves by $^{15}/_{26}$ or 0.58. The number of invariant amino acids rises to 15 from 14, while the 'conservative' substitutions drop from 7 to 6, and the 'radical' class moves from 6 to 5. In terms of possible base changes per codon the new alignment equals $^{16}/_{26}$ or 0.62 in contrast to a random deviation of 1.63. The gA_{13}(tyr) and sA_{14}(arg) residues

[30] E. L. SMITH and E. MARGOLIASH, Fedn Proc. Fedn Am. Socs exp. Biol. *23*, 1243 (1964).
[31] T. H. JUKES, Am. Sci. *53*, 477 (1965).
[32] *The Genetic Code, Cold Spring Harbor Symp. Quant. Biol.* (Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring 1966).
[33] A. L. GOLDBERG and R. E. WITTES, Science *153*, 420 (1966).

Glucagon

His-Ser-GLN-Gly-Thr-Phe-Thr-Ser-*Asp*-TYR-Ser-*Lys*-TYR-*Leu*-Asp-Ser-ARG-Arg-ALA-Gln-ASP-*Phe*-*Val*-Gln-*Trp*-Leu-*Met*-Asn-Thr
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

Secretin

His-Ser-ASP-Gly-Thr-Phe-Thr-Ser-*Glu*-LEU-Ser-*Arg*-LEU-*Arg*-Asp-Ser-ALA-Arg-LEU-Gln-ARG-*Leu*-*Leu*-Gln-*Gly*-Leu-*Val*-NH_2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Base changes g vs. s

0 0 2 0 0 0 0 0 1 2 0 1 2 1 0 0 2 0 2 0 2 1 1 0 1 0 1

Scheme 1. The primary structures of glucagon (porcine) and secretin (porcine) arranged for direct correspondence. For those positions in which no substitutions have occurred, ordinary characters are used; where 'conservative' substitutions have been observed, the residues are *italicized*; where 'radical' substitutions have been found, the residues are CAPITALIZED. See text for definitions used in this classification.

Glucagon

His-Ser-GLN-Gly-Thr-Phe-Thr-Ser-*Asp*-TYR-Ser-*Lys*-**Tyr**-Leu——Asp-Ser-ARG-Arg-ALA-Gln-ASP-*Phe*-*Val*-Gln-*Trp*-Leu-*Met*-Asn-Thr
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

Secretin

His-Ser-ASP-Gly-Thr-Phe-Thr-Ser-*Glu*-LEU-Ser-*Arg*——Leu-**Arg**-Asp-Ser-ALA-Arg-LEU-Gln-ARG-*Leu*-*Leu*-Gln-*Gly*-Leu-*Val*-NH_2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Base changes g vs. s

0 0 2 0 0 0 0 0 1 2 0 1 – 0 – 0 0 2 0 2 0 2 1 1 0 1 0 1

Scheme 2. The primary structures of glucagon (porcine) and secretin (porcine) arranged for a shifted correspondence. For those positions in which no substitutions have occurred, ordinary characters are used; where 'conservative' substitutions have been observed, the residues are *italicized*; where 'radical' substitutions have been found, the residues are CAPITALIZED; and where 'insertions' or 'deletions' are indicated, the residues are in **bold-face**. See text for definitions used in this classification.

have no partners in the modified analysis and can correspond either to direct insertion or deletion mutations in the 2 coding cistrons. The action of ionizing radiation (guanine-guanine cross-linking ?), random crossover mechanisms or viral infections may have introduced these 'holes'[34], once their reality is granted and sufficient reason existed for their formation at one time in Nature. This point is significant because the biochemical specificity of these 2 related hormones must be maintained by minimum deviations in the tertiary folded form of the polypeptide. Similarly, an 'active site' in these compounds will be dependent on proper spacing of the functional side chains on one side or the other of a possible α-helix conformation of the molecule. The addition or subtraction of amino acids would naturally change the distance between the remaining residues and so produce a marked shift in specific hormone activity. Valid support for the deletion scheme can come from finding new glucagons or secretins with additional residues or equivalent substitutions in the common A_{13}-A_{14} region. It may be noted that on statistical grounds the gaps are improbable, since they create new problems with respect to additional homology comparisons[35].

The sum of the mutational amino acid shifts between secretin and glucagon amounts to either 19 residues ($7 \times 1 + 6 \times 2 + 0 \times 3$) or 18 residues ($6 \times 1 + 5 \times 2 + 0 \times 3 + 2$). Assuming gene duplication occurred 200×10^6 years ago (200 Myr.)[36] and if the time needed for a random amino acid change in the polypeptide chain is about 10 Myr, years, then 20 such incidents would be expected in the secretin-glucagon (porcine) system. A simple statistical calculation of the probable number of mutational events based on either the 13 (non-shifted model) or 12 (shifted model) observed variant sites in secretin yields the same sum of 20, a result in excellent agreement with the actual count[37]. It is noted the first procedure ignores possible back mutations and the dangers inherent in taking the simplest route between codons, while the second method disregards possible cyclic, convergent changes in a residue of glucagon that are not matched at a corresponding residue in secretin. These arguments necessarily invoke the fact of equal evolutionary variations at all mutable sites, a probably untrue point[38]. Bearing these difficulties in mind, and noting the close correspondence between the calculated residue changes and the observed residue values, it is postulated the ancestral gene for the glucagon-secretin system duplicated in the early Mesozoic era. Glucagon, or a physiological fraction possessing glucagon activity, is widely distributed in mammals, birds and some amphibians, a fact that substantiates these speculations. It would be interesting to learn whether glucagon can be rigorously identified in fish; if so, the glucagon-secretin split may have taken place in the late Paleozoic.

Attention is drawn to the recent suggestion concerning a common structural background between glucagon and insulin[39]. If gA_1(his) of glucagon (porcine) is placed next to iB_{10}(his) of insulin (porcine) and the alignment continued for 12 more residues, there is seen a spatial coincidence between 6 amino acids in glucagon – gA_1(his), gA_2(ser), gA_7(thr), gA_8(ser), gA_{11}(ser), and gA_{13}(tyr) – and a similar number in insulin – iB_{10}(his), iB_9(gln), iA_8(ser), iA_{12}(ser) and iA_{14}(tyr). The same situation exists with secretin with the exception of sA_{13}(leu). Although intriguing, the original comparison fails on 3 counts: first, a reversed –COCHRNH– amide alignment in glucagon is equated with a normal –NHCHRCO– amide pattern in insulin; second, the complete change of each coding triplet dictated by this mechanism would imply a massive, favorable genetic error; and third, the many constraints required to conserve the secondary and tertiary structures of this new peptide fragment would appear to be incompatible with a favorable biological role. It is concluded from these considerations that glucagon and secretin are not derived from insulin. A similar argument is used to discard the suggestion of a resemblance between some small glucagon subunits and related sequences in the heme proteins[40].

Finally, secretin and glucagon contain an extremely high content of polar to apolar residues (glucagon, $^{17}/_6$ or 2.8; secretin, $^{17}/_8$ or 2.1), which is characteristic of a fibrillar material whose interactions are towards other polar substances[41]. Indeed, a molecular model (Corey-Pauling-Koltun) of glucagon in a normal α-helix conformation has almost all the polar residues reside on one side, while the bulk of the non-polar residues lie on the opposite face. A similar conclusion has been reached on the basis of both X-ray diffraction and optical rotatory dispersion evidence[42–44]. When more sequence information is in hand, a more profitable discussion can ensue on the significance of the hydrophobic or hydrophylic clustering in glucagon and secretin, evolutionary aspects of structure-function relationships, and other phenotype characteristics. It is anticipated that the information accumulating from phylogenetic tree construction[45,46] for cytochrome *c*, hemoglobin and other proteins[47] should be duplicated in the near future with both glucagon and secretin[48].

Zusammenfassung. Die Polypeptidhormone Glukagon und Sekretin wurden auf der Basis des genetischen Codes verglichen. Es wurde daraus geschlossen, dass die beiden Substanzen ursprünglich von einem gemeinsamen Gen erzeugt wurden.

B. WEINSTEIN

Department of Chemistry, University of Washington, Seattle (Washington 98105, USA),
1 November 1967.

[34] E. BRAUNITZER, V. BRAUN, K. HILSE, E. HOBOM, V. RUDLOFF and G. v. WETTSTEIN, *Evolving Genes and Proteins* (Academic Press, New York 1965), p. 183.
[35] C. R. CANTOR, personal communication (20 October 1967).
[36] K. RANKAMA, Nature *214*, 634 (1967).
[37] W. FELLER, *An Introduction to Probability Theory and Its Applications*, 2nd edn (John Wiley and Sons, New York 1957), vol. 1, p. 93.
[38] A Poisson distribution calculation leads to variable results.
[39] T. M. SCHUSTER, Nature *209*, 302 (1966).
[40] F. SŐRM, Colln Czech. chem. Commun. *27*, 1604 (1962).
[41] F. T. HATCH, Nature *206*, 777 (1965).
[42] M. V. KING, J. molec. Biol. *11*, 549 (1965).
[43] M. H. BLANCHARD and M. V. KING, Biochem. biophys. Res. Commun. *25*, 298 (1966).
[44] W. B. GRATZER, E. BAILEY and E. H. BEAREN, Biochem. biophys. Res. Commun. *28*, 914 (1967).
[45] W. M. FITCH, J. molec. Biol. *16*, 9 (1966).
[46] W. M. FITCH and E. MARGOLIASH, Science *155*, 279 (1967).
[47] G. H. DIXON, *Essays in Biochemistry* (Academic Press, New York 1966), vol. 2, p. 147.
[48] This investigation was supported by the National Science Foundation (No. GB-3208) and the National Institutes of Health (No. AM 10671-01).