Amino Acid Sequence Homology of Mammalian Type C RNA Virus Major Internal Proteins*

(Received for publication, November 18, 1974)

STEPHEN OROSZLAN, TERRY COPELAND, MURRAY R. SUMMERS, GARY SMYTHERS, AND RAYMOND V. GILDER

From Flow Laboratories, Inc., Rockville, Maryland 20852

SUMMARY

The NH$_2$-terminal amino acid sequence of the major group-specific antigen, the major internal virion protein (p30; approximate molecular weight 30,000) of several mammalian type C RNA viruses was determined by the Edman degradation procedure using an automated protein sequenator. All of the proteins analyzed show a high degree of over-all sequence homology and also contain specific regions or single residues. All p30's begin with the sequence prolyl-leucylarginyll (Pro-Leu-Arg) and have an invariant, conserved region from residues 11 to 24. In this region only a single amino acid difference appears between the cat and mouse p30's. At position 17 alanine is found in the cat, and serine in all the mouse proteins. This homologous region starts at position 10 for RD-114 and baboon virus p30s, and at position 18 in the protein of the virus isolated from gibbon ape. The region extending from residue 4 to 10 shows considerable variability between p30s isolated from different mammalian species. Out of 24 residues compared, only a single amino acid difference was found between six different mouse p30s. At position 4, three have leucine, two have alanine, and one has serine. The comparative sequence data demonstrate that the viral p30s are products of related genes in the viruses from various mammalian species.

EXPERIMENTAL PROCEDURE

Materials

Reagents and Solvents—The reagents and solvents used in the sequenator were special grade reagents obtained from Beckman Instruments, Inc. (Palo Alto, Calif.), as were the buffers for amino acid analysis. Phenylthiohydantoin derivatives of pure amino acids used as standard were purchased from Pierce Chemical Co. All other chemicals were reagent grade of highest purity.

Source of Viral Proteins—The proteins used in these studies were obtained in milligram quantities from type C viruses grown in tissue culture and subsequently purified by sucrose density gradient centrifugation. The large scale virus production and purification techniques employing either the K 6 continuous flow rotor in the model K Mark II ultracentrifuge (Electro-Nucleonics) or the CF-32 continuous flow rotor operated by an L-350 ultracentrifuge (Spinco) have been described in detail (7). The concentrated viruses with a density of approximately 1.16 g/ml were reband a second time using Spinco zonal rotors. Viruses isolated from five different mammalian species were studied: mouse, cat, rat, gibbon ape, and baboon.

Murine Leukemia Virus—Several strains of mouse leukemia virus were used. Rauscher virus was grown in monolayer cultures of chronically infected mouse (BALB/c) bone marrow, JLS-V9 cells (8). The virus-shedding cell line was obtained from Electro-Nucleonics, Bethesda, Md. AKR mouse leukemia virus-producing cells were kindly supplied by Dr. Janet Hartley, National Institutes of Health, Bethesda, Md. The La Puente isolate (1501E) of the wild mouse leukemia virus was obtained from Dr. Earle Officer, University of Southern California at Los Angeles, and was grown in monolayer cultures (9). New Zealand black mouse type C virus was grown in suspension cultures established from a fibrosarcoma, SGC 604, at Scripps Clinic, La Jolla, Calif. (10). Moloney mouse sarcoma virus was produced in a rat tissue culture cell line, 18SA (11), obtained through the courtesy of Dr. Maurice Green (Institute of Molecular Virology, St. Louis, Mo.). Passage A Gross leukemia virus was kindly supplied by Dr. Ludwig Gross, Veterans Administration Hospital, Bronx, N.Y., and propagated in mouse embryo fibroblasts.

Rat Type C Virus—Rat type C virus, a pseudotype of murine sarcoma virus, was obtained from the MSB-1 cell line (12) derived from a tumor induced by M MSV in a female rat of the Brown Norway strain, as previously described and characterized (13).

The mammalian type C RNA viruses contain a major internal protein (the major group-specific antigen) of $M_1 = \sim 30,000$, designated as p30 according to the most recent nomenclature (1). These proteins exhibit both "species-specific" and "interspecific" antigenic determinants (2-5). The latter reaction category indicates that the p30s are a series of homologous proteins, a hypothesis substantiated by NH$_2$-terminal sequence analyses of the first 15 residues of three members of this family (6). Extension of these analyses to new virus isolates and also larger stretches of sequence in the individual p30s as reported here further supports the origin of these proteins from related genes and also indicates specific molecular regions useful for design of immunological reagents that should have either species-specific or interspecific properties. The latter should be of special value for surveys for type C viral gene products in human material.

* This work was supported by Contract NOI CP 43388 with the Virus Cancer Program of the National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014.
**Felina Type C Viruses**—Felinel leukemia virus (Theilen strain) was obtained from a chronically infected cat lymphosarcoma cell suspension culture (14). RD-114, an endogenous cat virus, was the isolate from the RD rhabdomyosarcoma cell line of McAllister et al. (15).

**Gibbon Ape Type C Virus**—The virus isolated from a gibbon ape (16) was grown in a human cell line (17).

**Baboon Type C Virus**—The virus isolated from a baboon cell culture of a caesium-derived male baboon (BABB-K) (18) was grown on dog thymus cells. The infected culture was kindly provided by Dr. George Todaro, National Cancer Institute. Purified virus derived from this cell line was obtained from the Fä ler Laboratories through the courtesy of Dr. R. Maysay (Fä ler), under the auspices of the Office of Program Resources and Logistics (Dr. Jack Gruber), Virus Cancer Program, National Cancer Institute.

**Methods**

**Protein Purification**

The p30s of the above virusses were purified by the high resolving power isoelectric focusing technique (19). Initially electrofocuspurified protein obtained from are of gibbon ape virus was used in these studies. These purification procedures were described in detail for the mouse (20, 21), rat (13), and feline type C viruses (3). More recently, we obtained purified p30s with the isoelectric focusing technique after complete denaturation of viral proteins with Na dodecyl sulfate-urea and 2 mercaptoethanol. Before electrofocusing, the anionic detergent was removed by Dowex-3A-1-X2 ion exchange resin (Bio-Rad, Richmond, Calif.) according to the method of Weber and Kuter (22). The disruption of the virus with this procedure is less tedious, and electrofocusing results in a high yield of p30s. Details of the purification procedure have been described (23, 24). We observed a substantial charge heterogeneity of the mouse p30s (21). In the primary structure studies reported here, the major isoelectric component was used from each virus p30.

All p30s purified by electrofocusing were chromatographed on Bio-Gel P-10 or Bio-Gel P-100 (Bio-Rad) columns equilibrated with 0.33 m ammonium acetate (6) or 0.25 m ammonium bicarbonate, and protein was recovered by lyophilization.

Purity of the protein was established by re electrofocusing, gel chromatography on guanidine HCl-agarose columns (25), electrophoresis in Na dodecyl-SO4-polyacrylamide gels (26), and most importantly, by the results of NH4-terminal sequence analysis as reported in part previously (6, 27) and described in more detail in the present communication.

For two viruses (AKR and GalV) the p30s subjected to sequence analysis were purified by guanidine HCl agarose chromatography (25, 28). After rechromatography, these proteins were also homogeneous by the above criteria.

**Determination of NH4-terminal Amino Acids**

The procedure described by Weiner et al. (29) was used without major modifications. Purified proteins (25 to 50 μg) dissolved in 100 μl of 0.1 m NaHCO3 were dansylated, hydrolyzed, and prepared for chromatography. Two-dimensional thin layer chromatography was performed on double-layered Cheng-Chin polyamide sheets (Gallard-Schlesinger, Carle Place, N. Y.) using a solvent system of 1.5% formic acid and benzene-acetic acid (9:1).

**Amino Acid Analysis**

Samples to be analyzed were placed in an ignition tube and dissolved in constant boiling HCl containing 0.06% mercaptoethanol and 0.5% phenol. The tubes were then evaporated and flushed with ultrapure N2 three times before being sealed at <500 μ pressure. Hydrolysis was effected by heating at 110° for varying time periods. Samples were quenched in ice to terminate hydrolysis and then taken to dryness in vacuo. For application to a Beckman 121H amino acid analyzer, the dried samples were redissolved in pH 2.2 sodium citrate buffer. The amino acid solutions were applied and electrophoresed using the Beckman single column methodology, and peak areas were determined electronically with an Autolab System AA computing integrator. For each of the proteins analyzed, the reproducibility of the values for the stable amino acids obtained with this technique was 96% or better based on standard deviation of the mean. The number of residues per mol was calculated by dividing the total recovery from the analyzer in nanograms by the gram molecular weight of the protein × 10^10 to obtain the number per mol. A molecular weight of 30,000 was used for calculation. The nanomolar value for each residue is then divided by the total protein input in nanomoles to give the number of residues per mol.

A complete analysis included samples after 24, 48, and 72 hours of hydrolysis, and a performic acid-oxidized sample hydrolyzed for 24 hours.

**Protein Sequencing**

Automated analysis was carried out (30) in a Beckman protein sequenator, model 890, equipped with a refrigerated fraction collector, as previously described (27), using the Quadrol single coupling double cleavage program provided by the manufacturer without major modifications. Protein samples were dissolved in heptanesulfonic acid at room temperature. After removal of aliquots for protein determination and amino acid analysis, the remaining sample was placed in the cup and dried in vacuo before initiating the automated program. The temperature during the run was maintained at 55-57°.

In several experiments the coupling step with phenylisothiocyanate was repeated prior to cleavage of the last residue, proline, but it did not increase the yield substantially. This is similar to the experiences of others (31). The thiazolinone derivatives were converted into phenylthiohydantoin by incubation in 1 n HCl at 80° for 10 min immediately after they were dried down in the refrigerated fraction collector compartment. The PTH derivatives were immediately analyzed by both gas chromatography and thin layer chromatography to parent amino acid by hydrolysis and subsequent analysis, however, was usually made after storage for several days at 4° or at −70°. Initially a reducing agent was not used, but in most of the runs fresh 1, 4-butanediol (K & K Laboratories, Plainview, N. Y.) (50 μl/liter) was added to the 1-chloro-2-carbitol to stabilize the products (32). When a viral protein was sequenced for the first time the sequenator was run without shut down between cycles.

**Identification of PTH Amino Acids**

**Gas-Liquid Chromatography**—The PTH derivatives of amino acids were identified by gas chromatography according to the method of Pisano et al. (33) with a Beckman model GC-50 gas chromatograph equipped with a flame ionization detector and dual column operation. A polar stationary phase SP-400 (Part No. 56706, Beckman Instruments, Inc.) was used in silylated glass columns (4 feet × 2 mm) and the temperature program consisted of a 2-min isothermal period at 165° followed by 110° rise over a 16-min period. In most cases, each residue was identified immediately after conversion with and without silylation. The PTH derivatives were quantitated by comparing peak heights with those of appropriate standards.

**Thin Layer Chromatography**—A previously described ultrasensitive microtechnique was used (34). Two-dimensional TLC was carried out on polyamide sheets (5 × 5 cm) (Gallard-Schlesinger). The first solvent system (toluene/n-pentane/glacial acetic acid, 60/30/5 v/v) contained the fluorescent indicator 2-4-butylnaphthalene-5(4'-biphenyl)-1, 3, 4-oxadiazole (obtained from Beckman Instruments, Inc.). Aqueous acetic acid (35%) was used in the second dimension. The dried chromatograms were inspected with short wave illumination (Mineralight Lamp, Ultra-Violet Products, Inc., San Gabriel, Calif.), and the PTH derivatives were detected as dark spots on a bright fluorescing background. Each residue was identified for all PTH derivatives with the exception of the pair isoleucine/leucine and arginine/histidine.

**Specific Color Reactions and Spot Tests**—To detect histidine, Pauly's reagent was used in spot tests (35). Arginine was detected by the phenyllisothiocyanate method (36).

**Amino Acid Analysis of Sequenator Fractions after Acid Hydrolysis**

To regenerate the parent amino acids from their PTH derivatives, hydrolysis was carried out with hydroiodic acid (Fisher Chemical) according to the method described by Smithies et al. (37). In most instances, PTH-norleucine added to the residue served as internal standard. Amino acid analyses were done with a Beckman model 121 analyzer as described above. Both ethyl acetate and aqueous phases were analyzed. The aqueous phases were run on a short column only to detect histidine and arginine.
RESULTS

Physical and Immunochemical Characterization of p30s

The approximate molecular size of the major group-specific antigens (p30s) has been determined by Na dodecyl-SO₄-polyacrylamide gel electrophoresis. In Table I the molecular weights of p30s from different species are given. These data indicate notable size differences as measured in this system. Fig. 1 illustrates the resolution obtainable in co-electrophoresis experiments.

Species differences in the isoelectric points (pI) of p30s have been reported previously (38). More recently, strain differences within a single species (e.g. mouse) have been observed and evidence was obtained for a substantial intrastrain charge heterogeneity (24). The isoelectric point values are listed in Table II for the major isoelectric p30 components from all of the various viruses for which amino acid sequence analyses were carried out.

All of the mammalian type C virus p30s show substantial serological cross-reactions and carry well defined species-specific antigenic determinants on the same molecule as demonstrated by various immunochemical techniques (2, 3, 13, 20, 21, 23, 24, 38-40). Using citraconylated proteins we were able to distinguish at least three species-specific and three interspecies-specific antigenic determinants in immunodiffusion experiments (41). More recently, type-specific (strain-specific) antigenic determinants have also been detected within a single species (5, 42). In addition, peptide mapping provided clear-cut evidence for the existence of type-specific chemical differences (24).

Amino Acid Composition

Amino acid compositional data of p30s of several mouse virus strains and of the two feline viruses are given in Table III. A

### Table I

<table>
<thead>
<tr>
<th>Virus</th>
<th>Molecular weight*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MuLV (all mouse strains)</td>
<td>31,000</td>
</tr>
<tr>
<td>FeLV (Theilen)</td>
<td>27,000</td>
</tr>
<tr>
<td>RD-114</td>
<td>29,000</td>
</tr>
<tr>
<td>M-MSV (RaLV)</td>
<td>27,000</td>
</tr>
<tr>
<td>GaLV</td>
<td>29,000</td>
</tr>
</tbody>
</table>

* Determined by Na dodecyl-SO₄-polyacrylamide gel electrophoresis analysis.

**Fig. 1.** Co-electrophoresis in Na dodecyl-SO₄-polyacrylamide (10%) gels of several type C virus p30s purified by isoelectric focusing. A: 1, FeLV; 2, RD-114; 3, the mixture of 1 and 2. B: 1, FeLV; 2, WMLV; 3, the mixture of 1 and 2. C: 1, AKR; 2, GaLV; 3, the mixture of 1 and 2.

### Table II

| Isoelectric points of mammalian type C virus p30s |
|------------------------|-------------------|
| Virus                  | pI    |
| R-MuLV                 | 6.7   |
| AKR                    | 6.7   |
| WMLV                   | 6.7   |
| SLV                    | 5.9   |
| M-MSV                  | 6.2   |
| G-MuLV                 | 8.6   |
| M-MSV (RaLV)           | 8.3   |
| FeLV (Theilen)         | 9.1   |
| RD-114                 | 6.2   |
| GaLV                   | 6.9   |
| BAB8-K                 |      |

### Table III

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Number of residues per mol</th>
<th>R-MuLV</th>
<th>AKR</th>
<th>WMLV</th>
<th>M-MSV</th>
<th>FeLV</th>
<th>RD-114</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lysine</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>13</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Histidine</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Arginine</td>
<td>20</td>
<td>27</td>
<td>27</td>
<td>25</td>
<td>22</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>31</td>
<td>28</td>
<td>30</td>
<td>30</td>
<td>23</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Threonine</td>
<td>14</td>
<td>16</td>
<td>13</td>
<td>16</td>
<td>14</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Serine</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>44</td>
<td>48</td>
<td>47</td>
<td>48</td>
<td>50</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Proline</td>
<td>19</td>
<td>18</td>
<td>19</td>
<td>17</td>
<td>23</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Glycine</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Alanine</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>16</td>
<td>20</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Half cystine</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Valine</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Methionine</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Isoleucine</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Leucine</td>
<td>32</td>
<td>33</td>
<td>33</td>
<td>30</td>
<td>37</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Tyrosine</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Total            | 261                        | 262    | 264 | 264  | 264   | 264  |

* Molecular weight of 30,000 was used for calculation.
* Based on a single complete analysis.
* Based on a duplicate complete analysis.
* Corrected for hydrolytic losses by extrapolation to 0 time.
* Determined as cysteic acid. Preliminary experiments by cleavage with -SH specific reagent NCNSPhCOOH indicate that there may be 2 cysteines/mol (S. Oroszlan and T. Copeland, unpublished observations).
* Determined as methionine sulfone.
gros examination of the number of individual residues per molecule indicates a similar amino acid composition for all of the proteins. All are rich in residues with polar side chains, e.g. arginine, lysine, aspartic, and glutamic acid (or their amides), and poor in cysteine, methionine, histidine, and isoleucine. An attempt to explain the differences observed in the isoelectric points (ranging from pH 5.9 to 9.1) cannot be made, since the number of amides (e.g. asparagine and glutamine) has not been determined.

Amino Acid Sequences

**Mouse Virus p30**—Fig. 2 shows a semilogarithmic plot of PTH derivatives in nanomoles quantitated by GLC at each cycle of the sequenator run with R MuLV p30. The yields of selected PTH derivatives—usually the most stable ones—were plotted versus the residue number. Data obtained for aspartic acid, asparagine, and glutamine are also included. In spite of the fact that these amino acids are not easily quantitated (30, 31), their omission would not significantly alter the slope (0.89) of the line. All figures were used without correction for overlap except for leucine at position 20 (see below). Repetitive yield calculated with figures obtained for leucine in Steps 2 and 10 was 92%, and similar calculation using the nanomole yields in Step 2 for leucine and Step 15 for phenylalanine gave 91%. The quantitative GLC data compared with the TLC results are given in Table IV. These combined data permit an unambiguous amino acid assignment for the first 24 residues. With the exception of the usual overlap and background appearing at higher numbered cycles, all sequenator fractions were found to contain a single major amino acid. Leucine seems to accumulate after Step 4 but shows a sharp increase above background value only at positions 10 and 20 (Fig. 3). In this run, proline was the other amino acid that appeared as background at a relatively high level, but again a sharp rise in yield at Cycle 14 and its subsequent fall (Fig. 4) identified this amino acid as the authentic residue 14. In fact, proline was the only amino acid seen by TLC in this step. The TLC method (34) proved also extremely useful in identifying PTH-serine. Amino acid analyses of hydroiodic acid-hydrolyzed PTH derivatives confirmed the results obtained by GLC and TLC. Assignment of lysine at position 24 is based on TLC results corroborated with amino acid analysis after back conversion. Similar data are shown in Table V for WXILV p30. Again, amino acid analyses of all of the sequenator samples did not reveal any inconsistency. Reduced and carboxymethylated R MuLV p30...
was also analyzed. No discrepancies in this run were noticed from the analysis of the nonreduced R-MuLV p30.

Purified p30s from other strains of murine type C virus were also subjected to sequence analysis. The major amino acid residues found by GLC and TLC are listed in Table VI for the proteins of SLV, M-MSV, and AKR viruses. The data shown for AKR p30 were obtained with OnHCl-agarose purified protein. Electrosed p30 was used from all other mouse strains. The AKR run was not followed by amino acid analysis, but the results obtained with this detection method for SLV and M-MSV were entirely consistent with the data given in Table VI.

Cat Virus p30s—The p30 of the Thelrien strain of FeLV was degraded in the automatic sequenator and after residue aliquots were taken for GLC and TLC the remainder of the PTH derivatives was hydrolyzed immediately with HI and subjected to amino acid analysis. The quantitative results are summarized in Table VII together with the TLC data. Fig. 5 shows a repetitive yield plot for an input of 64 nmol of protein (the slope of the line is 0.946). The repetitive yield calculated from residue 2 (leucine) to residue 15 (phenylalanine) is 96%. The over-all quantitative recovery data are in good agreement with the TLC results and are similar to those obtained for the mouse proteins by GLC. With the exception of the usual overlap from one residue to the other, the general background was well within the acceptable range but appeared somewhat higher than that obtained by GLC analysis. Leucine, alanine, and proline were the amino acids that appeared consistently in almost every cycle. The background values for these amino acids are given in Table VII. The yields for leucine from Cycles 1 to 30 are plotted in Fig. 3, and for alanine and proline from Steps 11 to 21 in Fig. 4. GLC data (not shown) completely supported the amino acid analysis and TLC results for the major components at each cycle. An unambiguous identification of residues up to position 26 was therefore possible. The assignment beyond this point is tentative. At positions 13 and 23 tryptophan was degraded to alanine and glycine during HI hydrolysis, but tryptophan was the major component by TLC and GLC. Similarly serine at Cycles 16 and 18 degraded to alanine, but it was the major component by TLC in both sequenator fractions. At position 24, lysine is assigned because it was the major component on TLC, and it was a newly appearing amino acid on a general background (e.g. leucine, alanine, and proline). Serine at position 25 was the major component by TLC and was also identifiable by GLC.

The other feline type C virus protein sequenced was the p30 of the endogenous cat virus, RD-114. The results are shown in Table VIII. The repetitive yield calculated from the recovery data for leucine in Cycle 2 and for phenylalanine in Step 14 was 92%. The yield for PTH-leucine in each cycle is plotted in Fig. 3, and for PTH-alanine from Steps 11 to 21 in Fig. 4. The identification and assignment to positions 4, 8, and 24 of threonine, an amino acid not appearing in the NH2-terminal sequence of the other p30s except baboon, and not easily quantitated by GLC, was greatly facilitated by the detection of α-aminobutyric acid in the amino acid analyzer after HI hydrolysis of the sequenator fractions. Residue 28 could not be determined.

p30s of Viruses from Other Species—A single automated Edman degradation was carried out with p30s of type C viruses from rat, gibbon ape, and baboon. The results of GLC and TLC analyses for rat and gibbon ape viruses are given in Table IX. For MSV(RaLV) p30 a definitive amino acid assignment could not be made at positions 7, 9, and 14. Of the first 27 residues of GaLV p30, assignments could not be made for positions 11, 23, 24, and 26.

The alignment of the NH2-terminal amino acid sequences of p30s from several type C RNA viruses, including five different species, is shown in Table X. A comparison of these sequences reveals that a high degree of genetic conservatism exists in the reservoir of viral genes coding for these proteins. Each protein starts with an identical sequence of Pro-Leu-Arg. A longer conserved region extends from residues 11 to 24. It may be assumed that a similar homology will be maintained for those proteins for which only shorter sequences are available at this time. In this almost completely homologous region only a single amino acid difference appears between cat and mouse p30s: at position 17 alanine is found in the former, and serine in the latter viral protein. This requires only a single base change in the codon. Between FeLV and RD-114 p30s a substantial homology exists beyond residue 24 up to the last known residue in position 30. In order to match RD-114 and baboon p30s with the FeLV p30,
TABLE X
NH₂-terminal amino acid sequences of p30s from several type C RNA viruses

<table>
<thead>
<tr>
<th>Virus</th>
<th>NH₂-terminus</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>FeLV</td>
<td>PLREGPNRPQYWPFSASDLYNWKTSHBPPF</td>
<td>1-30</td>
</tr>
<tr>
<td>RD 114</td>
<td>PLRT*VNRUVQYWPFSASDLYNWKTHTNPX</td>
<td>1-30</td>
</tr>
<tr>
<td>BABOON</td>
<td>PLRT*VNRUVQYWXM</td>
<td>1-30</td>
</tr>
<tr>
<td>MOUSE</td>
<td>PLRLGGNGQLQYWPFSSSDLY</td>
<td>1-30</td>
</tr>
<tr>
<td>R-MuLV</td>
<td>PLRLGGNGQLQYWPFSSSDLY</td>
<td>1-30</td>
</tr>
<tr>
<td>AKR</td>
<td>PLRLGGNGQLQYWPFSSSDLY</td>
<td>1-30</td>
</tr>
<tr>
<td>G-MuLv</td>
<td>PLRLGGNGQLQYWPFSSSDLY</td>
<td>1-30</td>
</tr>
<tr>
<td>WMLV</td>
<td>PLRLGGNGQLQYWPFSSSDLY</td>
<td>1-30</td>
</tr>
<tr>
<td>SLV</td>
<td>PLRAAGGNGQLQYWPFSSSDLY</td>
<td>1-30</td>
</tr>
<tr>
<td>M-MSV</td>
<td>PLRAAGGNGQLQYWPFSSSDLY</td>
<td>1-30</td>
</tr>
<tr>
<td>RAT</td>
<td>PLRQGAGXGMQYWXF</td>
<td>1-30</td>
</tr>
<tr>
<td>MSV (RaLV)</td>
<td>PLRQGAGXGMQYWXF</td>
<td>1-30</td>
</tr>
<tr>
<td>GIBBON</td>
<td>PLRAIGPPAEQYWFPXXAXLY</td>
<td>1-30</td>
</tr>
<tr>
<td>APE</td>
<td>(XNGLVLVP)</td>
<td>1-30</td>
</tr>
</tbody>
</table>

The one letter code used for amino acids is: A, alanine; B, aspartic acid or amide; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; X, unidentified; Y, tyrosine (see Ref. 50).

Based on TLC analysis only.

Assigned on the basis of homology to the other mouse viruses.

The comparative sequence data leave no doubt that the p30s are products of related genes in the viruses from various species. Temin (47, 48) has proposed that the information for vegetative virus (provirus) originates in the form of normal cell DNA sequences (protovirus), which by complex information transfers—mediated by reverse transcriptase—evolve into the provirus. Genetic homology among components of inherited viruses, such as p30, would then be viewed as the result of their ultimate origin in related segments of host cell DNA. The concept of co-evolution of host and viral genomes (38) based on the viral oncogene theory (49) is compatible with the protovirus theory if one assumes that the type C viruses are derived from host cell nucleic acid sequences, either early in evolutionary history or as a continuous process. Degrees of genetic homology between homologous viral components would be expected to be a reflection of evolutionary events occurring at the pro- or proto-virus levels. Horizontal interspecies transmission can clearly distort any anticipated regular order in such an event. This may explain the unexpected close relationship between endogenous viruses of cats and old world monkeys (18, 45, 46).

The sequence data must be extended further into the respective molecules before final conclusions are drawn; however, the original estimate of ~80% sequence relatedness between p30s from different species appears reasonable. The differences and
similarities seen correlate well with known immunochemical observations. Thus, species specificity could well result from reactivities directed against regions such as 4 to 10, and interspecific cross-reactions are presumably based on regions such as 11 to 24 (17 to 30 in the gibbon virus). Evidence that interspecific reactions of RD-114 and gibbon virus are distinctive (more distant than those of the other mammalian type C viruses) is consistent with the gap proposed for RD-114 p30 and the large interspecies differences by effects on conformation of p30. The relationship of these regions to immunological specificity is currently being studied using synthetic peptides in collaboration with Dr. S. Sallay, Purdue University. Forested antibody production against the synthetic conserved region may provide valuable reagents in the continued search for type C virus gene products in human tumor tissues.

Acknowledgments—We wish to thank Mr. Robert Toni, Mr. Zachary Gilstein, Mrs. Carol Foreman, Mrs. Michele Hanson, and Mr. Dave Bova for their excellent technical assistance.

REFERENCES

50. IUPAC-IUB (1968) J. Biol. Chem. 243, 3557-3559
Amino acid sequence homology of mammalian type C RNA virus major internal proteins.
  S Oroszlan, T Copeland, M R Summers, G Smythers and R V Gilden
  

Access the most updated version of this article at http://www.jbc.org/content/250/16/6232

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/250/16/6232.full.html#ref-list-1