ON INFORMATION TRANSFER
FROM NUCLEIC ACIDS
TO PROTEINS

BY

G. GAMOW

København 1955
i kommission hos Ejnar Munksgaard
Det Kongelige Danske Videnskabernes Selskab udgiver følgende publikationsrækker:

L’Académie Royale des Sciences et des Lettres de Danemark publie les séries suivantes:

Oversigt over selskabets virksomhed (8°)
(Annuaire)


Historisk-filologiske Meddelelser (8°)
Historisk-filologiske Skrifter (4°)
(Histoire et Philologie)


Arkæologisk-kunsthistoriske Meddelelser (8°)
Arkæologisk-kunsthistoriske Skrifter (4°)
(Archéologie et Histoire de l’Art)


Filosofiske Meddelelser (8°)
(Philosophie)


Matematisk-fysiske Meddelelser (8°)
(Matématiques et Physique)


Biologiske Meddelelser (8°)
Biologiske Skrifter (4°)
(Biologie)


Selskabets sekretariat og postadresse: Dantes plads 5, København V.

L’adresse postale du secrétariat de l’Académie est:

Det Kongelige Danske Videnskabernes Selskab,
Dantes plads 5, København V, Danmark.

Selskabets kommissionær: Ejnar Munksgaard’s forlag, Nørregade 6, København K.

Les publications sont en vente chez le commissionnaire:

Ejnar Munksgaard, éditeur, Nørregade 6, København K, Danmark.
DEDICATED TO PROFESSOR NIELS BOHR ON THE OCCASION OF HIS 70TH BIRTHDAY

ON INFORMATION TRANSFER FROM NUCLEIC ACIDS TO PROTEINS

BY

G. GAMOW

København 1955
i kommission hos Ejnar Munksgaard
One of the most important problems of today's theoretical biology is the problem of the transfer of the hereditary informations from the chromosomes, where they are presumably stored in the form of long polynucleotide sequences, to the enzymes which, as all other proteins, are represented by long polypeptide chains. Mathematically the problem reduces to finding a procedure by which a long number written in fourdigital system (four bases forming the molecules of nucleic acid) can be translated in a unique way into a long word formed by twenty different letters (twenty amino acids which form protein molecules). The fact that number 20 represents the number of different triplets (with disregard to order) which can be formed out of 4 different elements suggests that each amino acid in the resulting polypeptide sequence is determined by a group of three bases in the corresponding section of polynucleotide chain. Some time ago the author attempted to establish such a translation mechanism on the basis of Crick and Watson's model of Deoxyribonucleic Acid (DNA) molecule. It turned out, however, that the translation mechanism suggested by DNA structure ("diamond code") leads to a contradiction with the actually observed sequences of amino acids in certain protein molecules. This negative result is presumably due to over-simplification of the original picture, since it seems, indeed, that the transfer of informations from chromosomes to enzymes is a two-step process. First the informations stored in DNA molecules are transmitted to the molecules of RNA (Ribonucleic Acid) which move out into the cytoplasm of the cell and form the so-called microsomes. The second part

of the process consists in synthesizing the proteins according to the informations carried by RNA-base sequences. The presence of such double coding makes the analysis of information transfer much more difficult.

The most promising biological material for the study of information transfer is presented by viruses, since it has been shown that in this case (at least for bacterial viruses, or phages) the virus protein for the progeny is synthesized directly by the nucleic acid of the original virus particles which had penetrated into the cytoplasm of the host cell\(^1\). If we limit our studies to plant viruses in which nucleic acid is of RNA-type (rather than DNA-type, as in all animal viruses or phages), we may also hope that we are dealing with only one-step information transfer.

Very careful analysis of both nucleic acid and protein constitution is available in the case of Tobacco Mosaic Virus\(^2\), and is shown in Tables I and II.

*Table I.*

Relative amounts of bases (in moles) in RNA from TMV.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TMV</td>
<td>0.31</td>
<td>0.25</td>
<td>0.18</td>
<td>0.26</td>
</tr>
</tbody>
</table>

*Table II.*

Relative amounts of amino acids (in moles) in proteins from TMV.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>TMV</th>
<th>Amino Acid</th>
<th>TMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Alanine</td>
<td>0.070</td>
<td>11. Leucine</td>
<td>0.087</td>
</tr>
<tr>
<td>2. Arginine</td>
<td>0.069</td>
<td>12. Lysine</td>
<td>0.012</td>
</tr>
<tr>
<td>3. Asp. acid</td>
<td>0.030</td>
<td>13. Methionine</td>
<td>0.000</td>
</tr>
<tr>
<td>4. Asparagine</td>
<td>0.090</td>
<td>14. Phenylalanine</td>
<td>0.062</td>
</tr>
<tr>
<td>5. Cysteine</td>
<td>0.010</td>
<td>15. Proline</td>
<td>0.061</td>
</tr>
<tr>
<td>6. Glu. acid</td>
<td>0.047</td>
<td>16. Serine</td>
<td>0.083</td>
</tr>
<tr>
<td>7. Glutamine</td>
<td>0.047</td>
<td>17. Threonine</td>
<td>0.102</td>
</tr>
<tr>
<td>8. Glycine.</td>
<td>0.031</td>
<td>18. Tryptophane</td>
<td>0.012</td>
</tr>
<tr>
<td>9. Histidine</td>
<td>0.000</td>
<td>19. Tyrosine</td>
<td>0.026</td>
</tr>
<tr>
<td>10. Isoleucine</td>
<td>0.061</td>
<td>20. Valine</td>
<td>0.097</td>
</tr>
</tbody>
</table>


If we assume the hypothesis that each amino acid in the protein structure is determined by a triplet of bases in RNA chain, we may expect that relative abundance of different amino acids will be given by the products of relative abundances of the three corresponding bases. (Because of multiplicity of various types of triplets the probability of $aaa$, $aab$, and $abc$ types of triplets should be given the weights 1, 3, and 6).

The calculated abundance of various triplets, and the observed abundances of various amino acids, are plotted in Figure 1
in decreasing order, and we see that two sets of data are in a reasonably good agreement. If this agreement is not coincidental, it should permit the establishment of a correlation between individual units forming polynucleotide and polypeptide chains. Thus, for example, Valine would correspond to (Ad.Cy.Ur.), while Tyrosine would correspond to (Ad.Ad.Ad.).

Unfortunately, there are very few informations concerning RNA’s and protein constitution of plant viruses, and the only other example available at present is that of Turnip Yellow Virus (TYV)1. Using the assignments established by Figure 1 for the case of TYV, one finds that the observed abundances of amino acids do not correspond to calculated abundances of different triplets of bases. This may be due, however, to the deviation of base abundances observed in TYV from the Chargaff’s rule: \((A + C)/(G + U) = 1\) which seems to be well established for all other samples of RNA2.

Indeed, whereas in the case of TMV we have

\[
\begin{align*}
\text{Ad + Cy} & = 0.31 + 0.18 = 0.49 \\
\text{Gu + Ur} & = 0.25 + 0.26 = 0.51 \sim 1,
\end{align*}
\]

the data for TYV2 lead to

\[
\begin{align*}
\text{Ad + Cy} & = 0.23 + 0.38 = 0.61 \\
\text{Gu + Ur} & = 0.17 + 0.22 = 0.39
\end{align*}
\]

Thus, in order to test the proposed correlation, more data on RNA and protein constitution of plant viruses are needed.

The author considers it his pleasant duty to express his thanks to Dr.’s W. E. Cushen, J. Dunitz, and A. Rich for helpful discussion.

Note added in April 1955.

It was indicated to the author by his colleague Dr. M. Ycas that a reasonable agreement between amino acid constitution of both viruses (TMV and TYV), and the probabilities of base-triplets can be achieved if the “burden of fit” is distributed equally between the two cases. Indeed, if one makes the following assignments of the most abundant amino acids to the most probable base-triplets:

2 D. Elson and E. Chargaff. Biochim. et Biophys. Acta (in press). The author is grateful to Dr. Chargaff for the opportunity to see this manuscript prior to publication.
the observed abundances of these amino acids in both TMV and TYV agree with the calculated probabilities of base-triplets with a mean error of 20 per cent. This error can be easily explained by experimental errors in measurements of amino acid and base percentages in the viruses in question. The correctness of the proposed assignment, and the possibility of any assignment of that kind, can be verified, of course, only after the data for at least one more plant virus will become available.
## Matematisk-fysiske Meddelelser

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Titel</th>
<th>Forfatter</th>
<th>Pris (kr.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Moments of Inertia of Rotating Nuclei</td>
<td>Born, Max</td>
<td>3.00</td>
</tr>
<tr>
<td>2</td>
<td>Continuity, Determinism, and Reality</td>
<td>Boggild, Hooper, Ortel, and Scharff</td>
<td>3.50</td>
</tr>
<tr>
<td>3</td>
<td>Some Studies on Heavy Meson Events in Stripped Emulsions</td>
<td>Bøggild</td>
<td>3.00</td>
</tr>
<tr>
<td>4</td>
<td>Hall Coefficient and Resistivity of (\alpha) and (\beta)-Brass from 20—600° C</td>
<td>Born</td>
<td>3.00</td>
</tr>
<tr>
<td>5</td>
<td>On the Potential Collective Flow of a Rotating Nucleus with Non-Ellipsoidal Boundary</td>
<td>Hermansen</td>
<td>2.00</td>
</tr>
<tr>
<td>6</td>
<td>An Algorithm for Thickening Control in the Production of Interference Filters</td>
<td>Gustafson</td>
<td>5.75</td>
</tr>
<tr>
<td>7</td>
<td>On the Mathematical Structure of T. D. Lee's Model of a Renormalizable Field Theory</td>
<td>Källén, Pauli</td>
<td>3.25</td>
</tr>
<tr>
<td>8</td>
<td>Total Charges and Electron Capture Cross-Sections of Fission Fragments in Gases</td>
<td>Lassen</td>
<td>1.50</td>
</tr>
<tr>
<td>9</td>
<td>Quantitative Beta Track Autoradiography with Nuclear Track Emulsions</td>
<td>Levi, Hilde, Høgbøn</td>
<td>2.50</td>
</tr>
<tr>
<td>10</td>
<td>Old Problems in the General Theory of Relativity Viewed from a New Angle</td>
<td>Møller</td>
<td>3.50</td>
</tr>
<tr>
<td>11</td>
<td>Multipole Order of the (\gamma)-Rays from ({}^{81})Te</td>
<td>Nielsen, O. B.</td>
<td>2.00</td>
</tr>
<tr>
<td>12</td>
<td>Den klassiske mekanik i geometrisk beskrivelse. With an English Summary</td>
<td>Pihl, Mogens</td>
<td>3.00</td>
</tr>
<tr>
<td>13</td>
<td>Spectroscopic Investigations of Separated Krypton Isotopes</td>
<td>Rasmussen, Middelboe</td>
<td>3.25</td>
</tr>
</tbody>
</table>

## Biologiske Meddelelser

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Titel</th>
<th>Forfatter</th>
<th>Pris (kr.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>On Information Transfer from Nucleic Acids to Proteins</td>
<td>Gamow</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>Conservation of Skeletal Calcium Atoms through Life</td>
<td>Hévesy</td>
<td>2.00</td>
</tr>
</tbody>
</table>

### Total Price

- **Prisen for det samlede festskrift er kr. 41.-**
- The price of the whole volume is **41 Dan. Cr.**