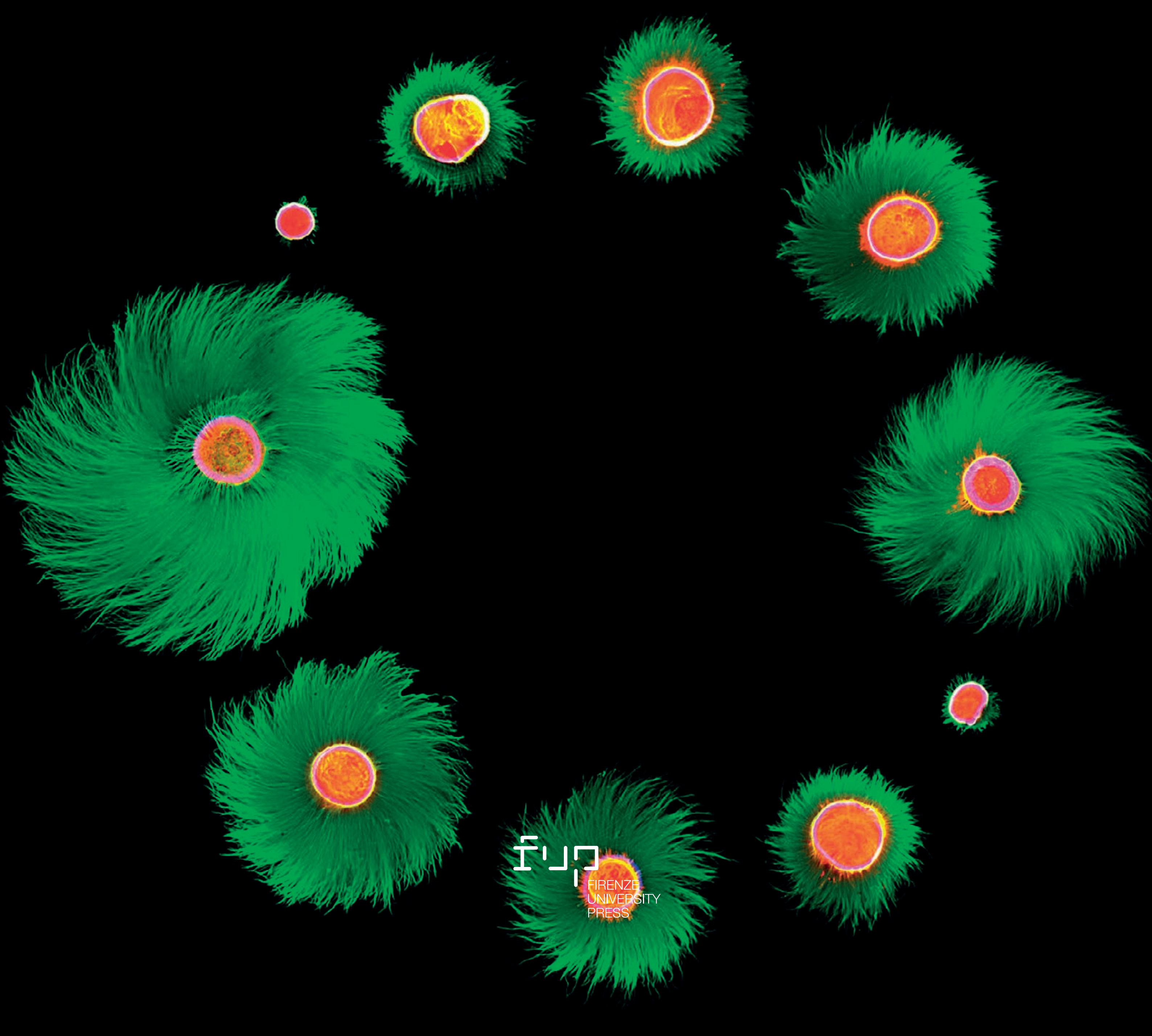
September 2022
Vol. 6 - n. 2

2532-3997

Substantia

An International Journal of the
History of Chemistry



FIRENZE
UNIVERSITY
PRESS

Substantia

An International Journal of the History of Chemistry

Vol. 6, n. 2 - 2022

Firenze University Press

***Substantia*. An International Journal of the History of Chemistry**

Published by
Firenze University Press – University of Florence, Italy
Via Cittadella, 7 - 50144 Florence - Italy
http://www.fupress.com/substantia

Direttore Responsabile: **Romeo Perrotta**, University of Florence, Italy

Cover image: Fluorescence image (magnification 4X) of the radial growth of sensory neurons (axons stained in green, somas in blue and actin filaments in red), by Dr. María Alejandra Lopez-Verrilli, Pontificia Universidad Católica de Chile, Facultad de Ciencias Biológicas, Santiago, Chile. Courtesy of Nikon Small World (Image of Distinction, 2012 Photomicrography Competition, https://www.nikonsmallworld.com).

Substantia is honoured to declare the patronage of:

museo galileo
Istituto e Museo di Storia della Scienza

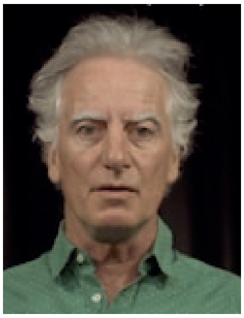
UNIVERSITÀ DEGLI STUDI FIRENZE
DIPARTIMENTO DI CHIMICA "UGO SCHIFF"

Società Chimica Italiana

Substantia
An International Journal of the History of Chemistry

No walls. Just bridges

Substantia is a peer-reviewed, academic international journal dedicated to traditional perspectives as well as innovative and synergistic implications of history and philosophy of Chemistry.
It is meant to be a crucible for discussions on science, on making science and its outcomes.
Substantia hosts discussions on the connections between chemistry and other horizons of human activities, and on the historical aspects of chemistry.
Substantia is published *open access* twice a year and offers top quality original full papers, essays, experimental works, reviews, biographies and dissemination manuscripts.
All contributions are in English.



SCIENTIFIC BOARD

as of 15 June 2022

Ferdinando Abbri
University of Siena, Italy

Francesca Antonelli
University of Bologna, Italy

Nicola Armaroli
CNR, Istituto per la Sintesi Organica
e la Fotoreattivita, Bologna, Italy

Vincenzo Balzani
University of Bologna, Italy

Marco Beretta
University of Bologna, Italy

Paolo Blasi
University of Florence, Italy

Gisela Boeck
Riga Technical University, Latvia;
Institut für Chemie, Universität
Rostock, Germany

Elena Bougleux
University of Bergamo, Italy

Luigi Campanella
University of Rome La Sapienza, Italy

Andrea Cantini
University of Florence, Italy

Louis Caruana SJ
Gregorian University of Rome, Italy

Elena Castellani
University of Florence, Italy

Martin Chaplin
London South Bank University, UK

Marco Ciardi
University of Florence, Italy

Luigi Dei
University of Florence, Italy

Ernesto Di Mauro
Università degli Studi della Tuscia, Italy

Sarah Everts
Carleton University, Canada

Yasu Furukawa
The Graduate University for
Advanced Studies (SOKENDAI),
Hayama, Japan

Juan Manuel Garcia Ruiz
University of Granada, Spain

Andrea Goti
University of Florence, Italy

Antonio Guarna
University of Florence, Italy

Marc Henry
University of Strasbourg, France

Roald Hoffmann
Cornell University, USA

Ernst Homburg
University of Maastricht, The
Netherlands

Juergen Heinrich Maar
Univ. Federal de Santa Catarina, Brasil

Bart Kahr
New York University, NY, USA

Roberto Livi
University of Florence, Italy

Stjepan Marcelja
Australian National University,
Australia

Pierluigi Minari
University of Florence, Italy

Juzo Nakayama
Saitama University, Japan

Barry W. Ninham
Australian National University,
Australia

Mary Virginia Orna
College New Rochelle, USA

Carla Rita Palmerino
Center for the History of Philosophy
and Science, The Netherlands

Adrian V. Parsegian
University of Massachusetts Amherst,
USA

Seth C. Rasmussen
North Dakota State University, USA

Jürgen Renn
Max Planck Institute for the History
of Science, Germany

Adrian Rennie
University of Uppsala, Sweden

Guillermo Restrepo
Max Planck Institute for Mathematics
in the Sciences Leipzig, Germany

Yona Siderer
The Hebrew University of Jerusalem,
Israel

Alberto Striolo
School of Chemical, Biological
and Materials Engineering, The
University of Oklahoma, USA

Brigitte Van Tiggelen
Science History Institute, USA

Barbara Valtancoli
University of Florence, Italy

George Weiss
Georgetown University, USA

EDITORIAL BOARD

Simone Ciofi Baffoni, University of Florence, Italy
Ilaria Palchetti, University of Florence, Italy
Damiano Tanini, University of Florence, Italy

ASSISTANT EDITOR

Duccio Tatini, University of Florence, Italy

MANAGING EDITOR

Alessandro Pierno, Firenze University Press, Italy

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

The Cultural Revolution: Ecological and Social

Vincenzo Balzani

Emeritus Professor in Chemistry - University of Bologna, Italy
Member of the Accademia dei Lincei
Email: vincenzo.balzani@unibo.it

Our world is sick because of the bad relationship between human society and the planet and even more because of the discords within human society itself. We are slipping more and more towards ecological and social unsustainability. Both scientists and philosophers say it, and Pope Francis highlights these views in the *Laudato Sì*: " Doomsday predictions can no longer be met with irony or disdain. [...] The pace of consumption, waste and environmental change has so stretched the planet's capacity that our contemporary lifestyle, unsustainable as it is, can only precipitate catastrophes".[1] Here then, as the Pope writes, " bold cultural revolution" is needed.[2]

The myth of a continuous and permanent growth has dominated human society for several years. An absurd myth that leads us to consider our planet only as a container of resources, without limits. The planet is actually a system with limited resources, consisting of chemical elements and their compounds, some relatively abundant, others scarce. In addition, resources are unevenly distributed across the planet, so there is a strong competition between peoples and between nations to get hold of them.

Investigations by the international agency Oxfam show that the gap in inequality between rich and poor, both at the level of people and nations, continues to widen without restraint.[3] Thus, as Pope Francis wrote in the *Laudato Sì*, " We are faced not with two separate crises, one environmental and the other social, but rather with one complex crisis which is both social and environmental".[4]

In the most recent encyclical *Fratelli Tutti* Pope Francis explains that the cultural revolution that is necessary to achieve ecological and social sustainability cannot be accomplished through some partial changes in the humans-planet relationship or in the international relationships. Instead, the basis on which our cultures rest must be radically changed. We must accept and appreciate diversity, acknowledge our limits and recognize that we are all children of God, brothers and sisters who are born, live and die in the same common home, planet Earth. In other words, the mandatory cultural revolution requires that humans and nations pass from the situation of inhabitants in the same planet, often engaged in commercial competition or even at war with each other, to that of brothers and sisters that love and esteem each other.

The main cause of ecological unsustainability is the use of fossil fuels. It is therefore necessary to resort to alternative energy sources, the renewable energies of the Sun (photovoltaic), wind (wind power) and water (hydroelectric) which, without generating pollution and without causing climate change, provide electricity, a form of energy much more valuable than heat produced by fossil fuels. Renewable energies are not only the answer to the climate crisis,

---

[1] Encyclical Letter "Laudato Sì" of the Holy Father Francis on Fraternity and Social Friendship. 2015. Ch. IV, 161. (https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html)

[2] Encyclical Letter "Laudato Sì" of the Holy Father Francis on Fraternity and Social Friendship. 2015. Ch. III, 114. (https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html)

[3] https://www.oxfam.org/en/reducing-inequality-what-your-country-doing-tackle-gap-between-rich-and-poor

[4] Encyclical Letter "Laudato Sì" of the Holy Father Francis on Fraternity and Social Friendship. 2015. Ch. IV, 139. (https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20150524_enciclica-laudato-si.html)

Substantia. An International Journal of the History of Chemistry 6(2): 5-6, 2022
ISSN 2532-3997 (online) | DOI: 10.36253/Substantia-1741

but also the key to fighting energy poverty. As reported in *Avvenire* by the director of the Power Shift Africa,[5] some developed nations, especially Italy and Germany, instead of supporting Africa in the development of renewable energy, push many African countries to pour their limited financial reserves into the development of an industry based on fossil fuel extraction.

In nature, the energies supplied by the Sun, wind and water are very abundant, but to convert them into electricity we need equipment (photovoltaic panels, wind turbines, dams, etc.) that must be built starting from the material resources that Earth provide. However, the amount of these materials is limited, so we must use them with maximum efficiency and recycle them. For this and other reasons it is necessary to abandon the disposable linear economy, powered by fossil fuels, and adopt a circular economy that uses renewable energy and that is based on reuse, repair and recycling of everything we produce.

Only in this way ecological sustainability can be achieved because the planet will be protected and not degraded, and its resources will be shared in sobriety. A wise policy can also be implemented to reduce inequalities through the development of common services (school, health, transport, etc.) and an economy based on taxes and subsidies aimed at helping the weakest, as every person is worth and must not be forgotten. The awareness that in a globalized world no one is self-sufficient will allow us to undertake fruitful collaborations between nations and to give strength to peace.

Then there is another problem. The limited material resources needed to convert the energies of the Sun, wind and water into electricity are not equally distributed on Earth. For example, some fundamental chemical elements, such as lithium for batteries and neodymium for wind turbines, are not found in Europe, but mainly in Chile, the former, and China, the latter. A country can cope with the scarcity of an important resource in two ways: with war, as often happened in the past for oil, or with collaboration and trade agreements. Will nations continue to be so reckless that they wage wars to conquer the resources they do not have, or will they finally understand that every war is a defeat for all?

Here is another aspect of the mandatory cultural revolution, indicated by Pope Francis in the encyclical *Fratelli tutti*: promoting a "universal aspiration to fraternity"[6] and "social friendship"[7] to build a better world starting from the common belonging to the human family and recognizing that "as children of the one God, we are all brothers and sisters".[8]

---

[5] https://www.powershiftafrica.org/

[6] Encyclical Letter "Fratelli tutti" of the Holy Father Francis on Fraternity and Social Friendship. 2020. Introduction, 8. (https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html)

[7] Encyclical Letter "Fratelli tutti" of the Holy Father Francis on Fraternity and Social Friendship. 2020. Ch. III, 106. (https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html)

[8] Encyclical Letter "Fratelli tutti" of the Holy Father Francis on Fraternity and Social Friendship. 2020. Ch. VIII, 279. (https://www.vatican.va/content/francesco/en/encyclicals/documents/papa-francesco_20201003_enciclica-fratelli-tutti.html)

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: Lekner J. (2022) The spinning electron. *Substantia* 6(2): 7-13. doi: 10.36253/Substantia-1630

Received: Apr 16, 2022

Revised: May 23, 2022

Just Accepted Online: May 24, 2022

Published: September 1, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

The Spinning Electron

John Lekner

School of Chemical and Physical Sciences, Victoria University of Wellington, PO Box 600, Wellington, New Zealand
E-mail: john.lekner@vuw.ac.nz

Abstract. The notion introduced by Ohanian that *spin is a wave property* is implemented, both in Dirac and in Schrödinger quantum mechanics. We find that half-integer spin is the consequence of azimuthal dependence in two of the four spinor components, relativistically and non-relativistically. In both cases the spinor components are free particle wavepackets; the total wavefunction is an eigenstate of the total angular momentum in the direction of net particle motion. In the non-relativistic case we make use of the Lévy-Leblond result that four coupled non-relativistic wave equations, equivalent to the Pauli-Schrödinger equation, represent particles of half-integer spin, with g-factor 2. An example of an exact Gaussian solution of the non-relativistic equations is illustrated.

Keywords: electron, spin, spinor.

1. INTRODUCTION

In his article “What is spin” [1], Ohanian argues that ‘spin may be regarded as an angular momentum generated by a circulating flow of energy in the wave field of the electron’, and that ‘the spin of the electron has a close classical analog: It is an angular momentum of exactly the same kind as carried by the wave field of a circularly polarized electromagnetic wave.’ Ohanian credits Belifante [2] for establishing that ‘this picture of spin is valid not only for electrons but also for photons, vector mesons, and gravitons.’

Dirac [3,4] regarded his four-by four matrices as ‘new dynamical variables…describing some internal motion of the electron, which for most purposes may be taken to be the spin of the electron postulated in previous theories’ [4]. This is how the concept of spin is presented in most texts, as intrinsically relativistic, a mysterious internal angular momentum for which there is no classical analogue. For example, in his “Introduction to quantum mechanics” [5] Griffiths states ‘…the electron also carries *another* form of angular momentum, which has nothing to do with motion in space (and which is not, therefore, described by any function of the position variables r,θ,ϕ) but which is somewhat analogous to classical spin…’.

We shall construct, for a general *relativistic or non-relativistic* wavepacket, an eigenstate of the component of total angular momentum in the net direction of propagation, with eigenvalue $\hbar/2$. Such eigenstates are four-component spinors, of which two components have $e^{i\phi}$ azimuthal dependence. In these formulations the phenomenon of spin is incorporated into ordinary space-time: the twist is in the azimuthal dependence of two of the wavefunctions. To the question: *what does a spinning electron look like*? we answer, in brief, that spin in the spinor formulation, relativistic or nonrelativistic, resides in the azimuthal dependence of two of the spinor components. This contrasts with the usual spin-space formulation, and the decoupling of spin from space-time.

In Sections 2 we construct general relativistic wavepackets with spin half; these are four-component spinors. An important aspect of spin is that *it is not purely a relativistic effect*: Levy-Léblond [6] has proved that the Galileo group has irreducible representations with non-zero spin. A Reviewer has pointed out that Galindo and del Rio [7] show that Galilean fermions are possible, with a four-component spinor linearization of the non-relativistic wave equation and a correct (to lowest order) g-factor. The Galindo and del Rio paper anticipates some of the work of Lévy-Leblond [6] and Gould [14].

Levy-Léblond's four-component nonrelativistic spinors are implemented in Section 3, to construct general angular momentum eigenstates with spin half. An explicit example of a non-relativistic spinning wavepacket is illustrated in Section 4.

2. DIRAC SPINORS

The wavefunction $\Psi(\boldsymbol{r},t)$ of an electron wavepacket in free space is to satisfy the Dirac equation

$$H\Psi(\boldsymbol{r},t)=i\hbar\partial_t\Psi(\boldsymbol{r},t), \quad H=c\boldsymbol{\alpha}\cdot\boldsymbol{p}+\beta mc^2, \quad \boldsymbol{p}=-i\hbar\nabla \tag{2.1}$$

The 4×4 matrices $\boldsymbol{\alpha},\beta$ are written in terms of the Pauli spin matrices $\sigma_x,\sigma_y,\sigma_z$ and the unit 2×2 matrix I as

$$\boldsymbol{\alpha}=\begin{pmatrix}0 & \boldsymbol{\sigma}\\ \boldsymbol{\sigma} & 0\end{pmatrix},\ \sigma_x=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\ \sigma_y=\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\\ \sigma_z=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\quad \beta=\begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix} \tag{2.2}$$

The wave equation (2.1) thus consists of four coupled first-order partial differential equations.

We consider wavepacket motion, predominantly along the z direction. In cylindrical polar coordinates $\rho=(x^2+y^2)^{\frac{1}{2}}$ is the distance from the z-axis, ϕ is the azimuthal angle, and

$$\frac{i}{\hbar}\left(p_x \pm ip_y\right)=\partial_x \pm i\partial_y = e^{\pm i\phi}\left(\partial_\rho \pm i\rho^{-1}\partial_\phi\right) \tag{2.3}$$

The four time-dependent free-space equations for the spinor Ψ read, with $mc/\hbar=K$,

$$(\partial_{ct}+iK)\psi_1+e^{-i\phi}(\partial_\rho-i\rho^{-1}\partial_\phi)\psi_4+\partial_z\psi_3=0 \tag{2.4a}$$

$$(\partial_{ct}+iK)\psi_2+e^{i\phi}(\partial_\rho+i\rho^{-1}\partial_\phi)\psi_3-\partial_z\psi_4=0 \tag{2.4b}$$

$$(\partial_{ct}-iK)\psi_3+e^{-i\phi}(\partial_\rho-i\rho^{-1}\partial_\phi)\psi_2+\partial_z\psi_1=0 \tag{2.4c}$$

$$(\partial_{ct}-iK)\psi_4+e^{i\phi}(\partial_\rho+i\rho^{-1}\partial_\phi)\psi_1-\partial_z\psi_2=0 \tag{2.4d}$$

When the spinor components ψ_j are independent of ϕ, solutions exist only for the ψ_j also independent of ρ. These are the well-known plane wave solutions $\psi_j=a_je^{i(qz-\omega t)}$, where the wavenumber q and the energy $\hbar\omega$ are constrained by $(\omega/c)^2=K^2+q^2$. To attain localized wavepacket solutions, we need to consider azimuthal dependence.

The angular momentum operator $\boldsymbol{L}=\boldsymbol{r}\times\boldsymbol{p}$ does not commute with the Hamiltonian, but the combination $\boldsymbol{J}=\boldsymbol{L}+\frac{\hbar}{2}\boldsymbol{\Sigma}$ does, where $\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & \boldsymbol{\sigma}\end{pmatrix}$. The z component of the total angular momentum operator is

$$J_z=L_z+\frac{\hbar}{2}\begin{pmatrix}\sigma_z & 0\\ 0 & \sigma_z\end{pmatrix}=-i\hbar\, diag(1,1,1,1)\partial_\phi+\\ +\frac{\hbar}{2}diag(1,-1,1,-1) \tag{2.5}$$

Let the spinor components ψ_j have azimuthal dependence $e^{i\nu_j\phi}$; the J_z eigenstate equations for ψ_1,ψ_2 read

$$\begin{pmatrix}-i\partial_\phi+1/2 & 0\\ 0 & -i\partial_\phi-1/2\end{pmatrix}\begin{pmatrix}e^{i\nu_1\phi}\\ e^{i\nu_2\phi}\end{pmatrix}=\\ =\begin{pmatrix}(\nu_1+1/2)e^{i\nu_1\phi}\\ (\nu_2-1/2)e^{i\nu_2\phi}\end{pmatrix} \tag{2.6}$$

This will be an eigenstate of J_z if $\nu_1+1/2=\nu_2-1/2$, $\nu_2-\nu_1=1$, with eigenvalue $(\nu_1+1/2)\hbar$. Similarly for ψ_3,ψ_4 we shall have an eigenstate of J_z if $\nu_3+1/2=\nu_4-1/2$, $\nu_4-\nu_3=1$, with eigenvalue $(\nu_3+1/2)\hbar$. Hence the choice $\nu_{1,3}=0$, $\nu_{2,4}=1$ makes Ψ an eigenstate of J_z with eigenvalue $\hbar/2$. (The choice $\nu_{1,3}=-1$, $\nu_{2,4}=0$ makes Ψ an eigenstate of J_z with eigenvalue $-\hbar/2$.) It is necessary to have integer ν_j, since the spinor components are in real space-time (not in some abstract spin space) so we must have $\psi_j(\phi+2\pi)=\psi_j(\phi)$. The eigenvalues of J_z are thus $\pm\hbar/2,\pm3\hbar/2$ etc.

With spinor components $\psi_{1,3}=f_{1,3}(\rho,z,t)$,$\psi_{2,4}=e^{i\phi}f_{2,4}(\rho,z,t)$, the azimuthal dependence cancels out, and the equations (2.4) read

$$(\partial_{ct}+iK)f_1+(\partial_\rho+\rho^{-1})f_4+\partial_zf_3=0 \tag{2.7a}$$

$$(\partial_{ct}+iK)f_2+\partial_\rho f_3-\partial_z f_4=0 \tag{2.7b}$$

$$(\partial_{ct}-iK)f_3+(\partial_\rho+\rho^{-1})f_2+\partial_z f_1=0 \tag{2.7c}$$

$$(\partial_{ct}-iK)f_4+\partial_\rho f_1-\partial_z f_2=0 \tag{2.7d}$$

The combination $(\partial_{ct}-iK)$(2.7a)-$(\partial_\rho+\rho^{-1})$(2.7d)-$\partial_z$(2.7c) gives

$$(\partial^2_{ct}+K^2-\partial^2_\rho-\rho^{-1}\partial_\rho-\partial^2{}_z)f_1(\rho,z,t)=0 \tag{2.8}$$

Likewise $(\partial_{ct}-iK)$(2.7b)-∂_ρ(2.7c)+∂_z(2.7d) gives us

$$(\partial^2_{ct}+K^2-\partial^2_\rho-\rho^{-1}\partial_\rho+\rho^{-2}-\partial^2{}_z)f_2(\rho,z,t)=0 \tag{2.9}$$

The equations (2.8) and (2.9) are solved respectively by

$$e^{i(qz-\omega t)}J_0(k\rho), \quad e^{i(qz-\omega t)}J_1(k\rho), \quad k^2+q^2+K^2=(\omega/c)^2 \tag{2.10}$$

The function f_3 satisfies the same equation as f_1, and f_4 satisfies the same equation as f_2. The transverse and longitudinal wavenumbers k and q are real, and $\omega \geq cK$, or $\hbar\omega \geq mc^2$. The wavenumbers $k \geq 0$ and $q \geq 0$ are related to $K=mc/\hbar$ and ω by $k^2+q^2+K^2=(\omega/c)^2$; the maximum value of both k and q is $Q=\sqrt{(\omega/c)^2-K^2}$. Hence the general form of the spinor eigenstate of J_z with eigenvalue $\hbar/2$ is

0

$$\psi_{1,3}(\rho,z,t)=\int_{cK}^{\infty}d\omega\int_0^Q dk\, A_{1,3}(\omega,k)e^{i(qz-\omega t)}J_0(k\rho) \tag{2.11}$$

$$\psi_{2,4}(\rho,\phi,z,t)=e^{i\phi}\int_{cK}^{\infty}d\omega\int_0^Q dk\, A_{2,4}(\omega,k)e^{i(qz-\omega t)}J_1(k\rho) \tag{2.12}$$

These are analogues of the acoustic and electromagnetic wavepackets, for which simple closed forms exist ([8], Section 2.6). The author has not found amplitudes $A_j(\omega,k)$ which lead to closed forms for the relativistic spinor components. Bessel beam wavefunctions (not localized enough transversely to have finite energy per unit length) have been studied by Bliokh et al. [9].

3. NON-RELATIVISTIC SPINORS

Lévy-Leblond [6] has shown that four coupled non-relativistic wave equations, equivalent to the Schrödinger equation, are spinors representing spin 1/2 particles, with g-factor 2 (see also Greiner [10]). We shall again construct a general eigenstate of J_z with eigenvalue $\hbar/2$: it is a four-component spinor. It is based on localized wavepacket solutions of the time-dependent Schrödinger equation, with no restriction on the wavepacket parameters. In Section 4 we shall explore some properties of exact Gaussian solutions of the equations satisfied by the spinor components.

Let $\Psi(\boldsymbol{r},t)$ be the four-component spinor, $\Psi=\begin{pmatrix}\psi\\ \chi\end{pmatrix}$, with ψ,χ each having two components. The Lévy-Leblond non-relativistic coupled spinor equations are, with $E=i\hbar\partial_t$, $\boldsymbol{p}=-i\hbar\nabla$,

$$E\psi+\boldsymbol{\sigma}\cdot\boldsymbol{p}\chi=0, \quad \boldsymbol{\sigma}\cdot\boldsymbol{p}\psi+2m\chi=0 \tag{3.1}$$

$\boldsymbol{\sigma}$ are, as before, the Pauli spin matrices defined in (2.2).

Note that the ψ,χ in (3.1) have dimension differing by a speed; we could make them the same by inserting factors $e^2/\hbar$ or c in front of χ, but choose not to do so, in order keep the Lévy-Leblond formulation. Note also that the lower spinor component χ can be eliminated, giving the Pauli-Schrödinger equation $E\psi=\frac{1}{2m}(\boldsymbol{\sigma}\cdot\boldsymbol{p})^2\psi$, with Hamiltonian $H=\frac{1}{2m}(\boldsymbol{\sigma}\cdot\boldsymbol{p})^2$.

For comparison, the Dirac equations (2.1), with $\psi_u=\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$, $\psi_v=\begin{pmatrix}\psi_3\\ \psi_4\end{pmatrix}$, may be written in the form

$$(E-mc^2)\psi_u=c\boldsymbol{\sigma}\cdot\boldsymbol{p}\psi_v, \quad c\boldsymbol{\sigma}\cdot\boldsymbol{p}\psi_u=(E+mc^2)\psi_v \tag{3.2}$$

The non-relativistic limit is obtained from (3.2) by setting $\psi_j(\boldsymbol{r},t)=e^{-imc^2t/\hbar)}F_j(\boldsymbol{r},t)$. Then $E\psi_j=i\hbar\partial_t\ \psi_j=e^{-\frac{imc^2t}{\hbar}}(mc^2+i\hbar\partial_t)F_j$, and the equations (3.2) have the dominant terms

$$EF_u=c\boldsymbol{\sigma}\cdot\boldsymbol{p}F_v, \quad c\boldsymbol{\sigma}\cdot\boldsymbol{p}F_u=2mc^2F_v \tag{3.3}$$

These are the same as (3.1) if we identify F_u with ψ, and cF_u with $-\chi$.

Returning to solutions of the Lévy-Leblond equations (3.10), we write $\psi=\begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$, $\chi=\begin{pmatrix}\psi_3\\ \psi_4\end{pmatrix}$, and consider wavepacket motion, predominantly along the direction, but of course converging onto or diverging from the focal region, which we shall centre at the space-time origin. Again in cylindrical polar coordinates ρ,ϕ, and with use of (2.3), the four time-dependent free-space equations (3.1) for the spinor Ψ read

$$-\partial_t\psi_1+e^{-i\phi}(\partial_\rho-i\rho^{-1}\partial_\phi)\psi_4+\partial_z\psi_3=0 \tag{3.4a}$$

$$-\partial_t\psi_2+e^{i\phi}(\partial_\rho-i\rho^{-1}\partial_\phi)\psi_3+\partial_z\psi_4=0 \tag{3.4b}$$

$$\frac{2im}{\hbar}\psi_3+e^{-i\phi}(\partial_\rho-i\rho^{-1}\partial_\phi)\psi_2+\partial_z\psi_1=0 \tag{3.4c}$$

$$\frac{2im}{\hbar}\psi_4+e^{i\phi}(\partial_\rho-i\rho^{-1}\partial_\phi)\psi_1+\partial_z\psi_2=0 \tag{3.4d}$$

When the spinor components ψ_j are independent of ϕ, solutions exist only for the ψ_j also independent of ρ. These are the plane wave solutions $\psi_j=a_je^{i(qz-\omega t)}$, where the wavenumber k and the energy $\hbar\omega$ are constrained by

$\hbar\omega=\hbar^2q^2/2m$. To attain localized wavepacket solutions, we need to consider azimuthal dependence.

The angular momentum operator $\boldsymbol{L}=\boldsymbol{r}\times\boldsymbol{p}$ does not commute with the free-particle Hamiltonian $H=\frac{1}{2m}(\boldsymbol{\sigma}\cdot\boldsymbol{p})^2$, but the combination $\boldsymbol{J}=\boldsymbol{L}+\frac{\hbar}{2}\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}=\begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & \boldsymbol{\sigma}\end{pmatrix}$ does, as may be verified from the commutators $\boldsymbol{\sigma}\times\boldsymbol{\sigma}=2i\boldsymbol{\sigma}$, $[L,\boldsymbol{\sigma}\cdot\boldsymbol{p}]=i\hbar\boldsymbol{\sigma}\times\boldsymbol{p}$, $[\boldsymbol{\sigma},\boldsymbol{\sigma}\cdot\boldsymbol{p}]=-2i\boldsymbol{\sigma}\times\boldsymbol{p}$. $\boldsymbol{J}$ satisfies the angular momentum commutation relations $\boldsymbol{J}\times\boldsymbol{J}=i\hbar\boldsymbol{J}$. The z component of the total angular momentum operator is again

$$J_z=L_z+\frac{\hbar}{2}\begin{pmatrix}\sigma_z & 0\\ 0 & \sigma_z\end{pmatrix}=-i\hbar\, diag(1,1,1,1)\partial_\phi+\frac{\hbar}{2}\, diag(1,-1,1,-1) \quad (3.5)$$

We shall now construct the non-relativistic spinor eigenstates of J_z.

Let the spinor components ψ_j have azimuthal dependence $e^{i\nu_j\phi}$; the J_z eigenstate equations for ψ_1,ψ_2 are the same as in (2.6):

$$\begin{pmatrix}-i\partial_\phi+1/2 & 0\\ 0 & -i\partial_\phi-1/2\end{pmatrix}\begin{pmatrix}e^{i\nu_1\phi}\\ e^{i\nu_2\phi}\end{pmatrix}=\begin{pmatrix}(\nu_1+1/2)e^{i\nu_1\phi}\\ (\nu_2-1/2)e^{i\nu_2\phi}\end{pmatrix} \quad (3.6)$$

The equations (3.5) and (3.6) have the same form as in the relativistic case, equations (2.5) and (2.6). Hence as before the choice $\nu_{1,3}=0$, $\nu_{2,4}=1$ makes Ψ an eigenstate of J_z with eigenvalue $\hbar/2$ and the choice $\nu_{1,3}=-1$, $\nu_{2,4}=0$ makes Ψ an eigenstate of J_z with eigenvalue $-\hbar/2$. With spinor components $\psi_{1,3}=f_{1,3}(\rho,z,t)$, $\psi_{2,4}=e^{i\phi}f_{2,4}(\rho,z,t)$, the equations (3.4) read

$$-\partial_t f_1+(\partial_\rho+\rho^{-1})f_4+\partial_z f_3=0 \quad (3.7a)$$

$$-\partial_t f_2+\partial_\rho f_3-\partial_z f_4=0 \quad (3.7b)$$

$$\frac{2im}{\hbar}f_3+(\partial_\rho+\rho^{-1})f_2+\partial_z f_1=0 \quad (3.7c)$$

$$\frac{2im}{\hbar}f_4+\partial_\rho f_1-\partial_z f_2=0 \quad (3.7d)$$

The last two equations give $f_{3,4}$ in terms of derivatives of $f_{1,2}$, which in turn satisfy the free-space Schrödinger equation for azimuthal orbital quantum number 0 and 1:

$$(i\hbar\partial_t+\frac{\hbar^2}{2m}[\partial_\rho^2+\rho^{-1}\partial_\rho+\partial_z^2])f_1(\rho,z,t)=0 \quad (3.8)$$

$$(i\hbar\partial_t+\frac{\hbar^2}{2m}[\partial_\rho^2+\rho^{-1}\partial_\rho-\rho^{-2}+\partial_z^2])f_2(\rho,z,t)=0 \quad (3.9)$$

Equations(3.8) and (3.9) are satisfied by $J_n(\kappa\rho)e^{in\phi}e^{iqz}e^{-i\hbar k^2t/2m}$, with n=0,1 respectively, and $\kappa^2+q^2=k^2$; J_n are the regular Bessel functions of order n. Hence spinor components of forward-propagating wavepackets have the form

$$e^{in\phi}\int_0^\infty dk\, e^{-i\hbar k^2t/2m}\int_0^k dq\, F_n(k,q)e^{iqz}J_n(\kappa\rho) \quad (\kappa^2+q^2=k^2) \quad (3.10)$$

The amplitudes $F_n(k,q)$ are complex functions, subject only to the existence of the norm and of the expectation values of energy and momentum of the wave packet. A similar expression gives the wavefunctions of scalar and of electromagnetic pulses [8].

To sum up this Section: a general non-relativistic eigenstate of J_z with eigenvalue $\hbar/2$ has been found: it is a four-component spinor, of which two components have 'twist', with $e^{i\phi}$ azimuthal dependence. In this formulation the spin resides in the azimuthal dependence of two of the wavefunctions, in real space-time.

Any spinor based on localized wavepacket solutions of the time-dependent Schrödinger equation, constructed as above, will be an eigenstate of J_z with eigenvalue $\hbar/2$. The next Section gives an explicit example. Stationary states (energy eigenstates) of the hydrogen atom are briefly discussed in Appendix A.

4. SPINNING GAUSSIAN WAVEPACKETS

A free-particle wavepacket solution of Schrödinger's time-dependent equation dates back to the early days of quantum mechanics (Kennard [11], Darwin [12]). This is the Gaussian wavepacket. It is a compact exact solution, but with a physical flaw, to be discussed below. For propagation along the axis, and with cylindrical symmetry, it has the form

$$g(\rho,z,t)=b^{3/2}[b+i\nu t]^{-3/2}exp\{iQ(z-\frac{ut}{2})-\frac{\rho^2+(z-ut)^2}{4b[b+i\nu t]}\} \quad (4.1)$$

The Gaussian wavepacket (4.1) is normalized so that $g^*g=1$ at the space-time origin. In (4.1) the spatial origin $\rho=0$, $z=0$ is the position of maximal $|g|$ a time t=0, Q is the dominant z component wavenumber, m is the mass of the particle, $u=\hbar Q/m$ is the group speed, and $\nu=\hbar/2mb$ is the spreading speed. The length b gives the spread of the wavepacket at t=0. Earlier and later the longitudinal and lateral spread of the packet is greater, proportional to $[b^2+(\nu t)^2]^{1/2}$. Thus $\rho=0$, $z=0$ can be thought of as the centre of the focal region of the wavepacket, occupied at t=0. As t increases towards zero the wavepacket converges to its most compact form, reaches it at t=0, and then expands as it continues to propagate in the positive z direction. The packet used by Ohanian [1] is equivalent to (4.1) evaluated at Q=0 (zero momentum expectation value) and t=0.

For the Gaussian wavepacket g the momentum operator has the expectation values (see for example [13])

$$\langle p_z\rangle=-i\hbar\langle\partial_z\rangle=\hbar Q,\quad \langle p_x\rangle=0=\langle p_y\rangle,\quad \langle p^2\rangle= =\langle-\hbar^2\nabla^2\rangle=\hbar^2\left(Q^2+\frac{3}{4b^2}\right) \tag{4.2}$$

The wavepacket g is neither an energy nor a momentum eigenstate, but it is an eigenstate of the orbital angular momentum operator $L_z=xp_y-yp_x=-i\hbar(x\partial_y-y\partial_x)=-i\hbar\partial_\phi$. The orbital angular momentum eigenvalue is zero, because g is independent of the azimuthal angle ϕ. Eigenstates of the z component of orbital angular momentum, with eigenvalues which are integer multiples of ħ, may be generated from any such g by differentiation, as shown in [13].

The probability density of the scalar wavepacket is g^*g: the probability that the particle described by $g(\boldsymbol{r},t)$ is within the volume element d^3r is $d^3r\, g^*g$. The *norm* $N=\int d^3r\, g^*g$ (integration over all of space) is independent of time. The probability density flux, or the probability current density vector $\boldsymbol{S}$, satisfies the conservation law

$$\nabla.\boldsymbol{S}+\partial_t(g^*g)=0,\quad \boldsymbol{S}(\boldsymbol{r},t)=\frac{\hbar}{m}Im(g^*\nabla g) \tag{4.3}$$

What are the corresponding relations for spinors? The conservation law is now (Lévy-Leblond [6], Section IIIe, and Appendix B)

$$\nabla.\boldsymbol{S}+\partial_t(\psi^+\psi)=0 \tag{4.4}$$

$$\boldsymbol{S}(\boldsymbol{r},t)=-\psi^+\boldsymbol{\sigma}\chi-\chi^+\boldsymbol{\sigma}\psi=\frac{\hbar}{m}Im(\psi^+\nabla\psi)+\frac{\hbar}{2m}\nabla\times(\psi^+\boldsymbol{\sigma}\psi) \tag{4.5}$$

The first term in the second expression for $\boldsymbol{S}$ corresponds to the Schrödinger current in (4.3), the second is a spin current. Ohanian [1] derived the relativistic analogue of last term in (4.5). He showed that it leads, in the nonrelativistic limit, to an azimuthal current. In his words, "such a circulating flow of energy will give rise to an angular momentum. This angular momentum is the spin of the electron."

We shall calculate the radial, azimuthal, and longitudinal components of the probability current density, S_ρ, S_ϕ, S_z in the simplest case, in which the spinor components are $\psi_1=f_1(\rho,z,t)$, $\psi_2=0$, $\psi_3\sim\partial_z\psi_1$, $\psi_4\sim e^{i\phi}\partial_\rho\psi_1$. From Appendix B, the components of the probability current density are given by

$$\frac{m}{\hbar}S_\rho=Im\{f^*{}_1\partial_\rho f_1\},\quad \frac{m}{\hbar}S_\phi=-\frac{1}{2}\partial_\rho|f_1|^2,\quad \frac{m}{\hbar}S_z=Im\{f^*{}_1\partial_z f_1\} \tag{4.6}$$

With $f_1(\rho,z,t)=g(\rho,z,t)$ the probability density and current components are given by

$$g^*g=b^3[b^2+(vt)^2]\ exp\left\{-\frac{\rho^2+(z-ut)^2}{2[b^2+(vt)^2]}\right\} \tag{4.7}$$

$$S_\rho=\frac{\rho t\left(\frac{\hbar}{2mb}\right)^2}{b^2+(vt)^2}g^*g,\quad S_\phi=\frac{\frac{\hbar\rho}{4m}}{b^2+(vt)^2}g^*g,\quad S_z=\frac{\frac{\hbar b^2}{m}\left(Q+\frac{vtz}{2b^3}\right)}{b^2+(vt)^2}g^*g \tag{4.8}$$

The components S_ρ, S_z are the same for the scalar wavepacket, the azimuthal component S_ϕ is zero in the scalar case based on g. The conservation law (4.4) is satisfied.

A problem with the Gaussian solution is apparent in S_z: for positive z and negative t (or vice versa) the longitudinal component is negative if the magnitude of vtz exceeds that of $2Qb^3$. The probability current then flows backward. Far from the focal region (here centred on the space-time origin) there should be no backward flow for free-space propagation. Note that the Gaussian wavepacket cannot be put in the purely forward-propagating form (3.10).

Nevertheless, the Gaussian packets demonstrate the azimuthal current component which arises in the spinor formulation. Figures 1 and 2 show the current components in the focal plane, and at a transverse plane cutting through the wavepacket center at a later time. The azimuthal part gives the electron wavepacket its spin.

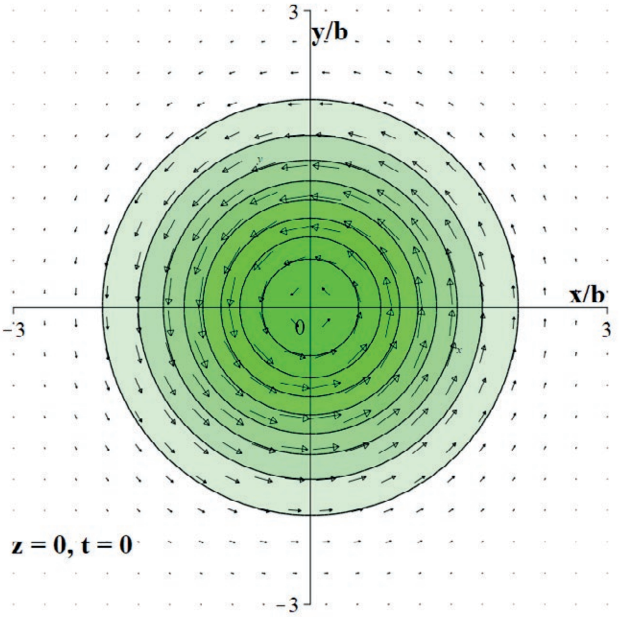


Figure 1. Focal plane section through a Gaussian spinor wavepacket, at t=0. The contours give the probability density, the arrows the transverse current density (the longitudinal current is not shown). The direction of motion is out of the page. The transverse current density is purely azimuthal at this instant.

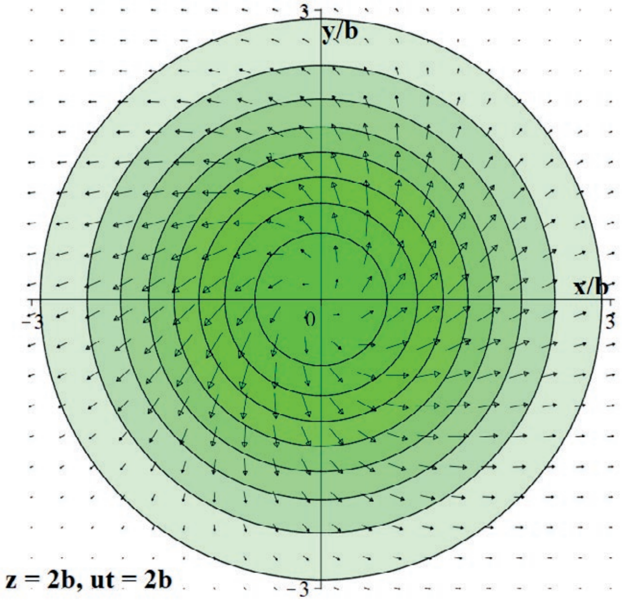


Figure 2. Gaussian spinor wavepacket, at $z=ut=2b$. The transverse current density now has radial and azimuthal components. The group speed is u, so the section is through the centre of the wavepacket. The longitudinal current density is not shown.

5. SUMMARY

The spinning electron may be described by a four-component spinor, depending on space and time coordinates, in both relativistic and non-relativistic quantum theory. The non-relativistic quantum theory and its azimuthal dependence is similar to the relativistic Dirac spinor formulation of Section 2. In both cases the spin is contained in the azimuthal dependence of wavefunctions in ordinary space-time. Gould [14] used the Hamiltonian $H=\frac{1}{2m}(\boldsymbol{\sigma}.\boldsymbol{p})^2$ to show that the magnetic moment follows (correct to lowest order), just as in the Lévy-Leblond spinor formulation. There is thus an alternative formulation to the usual 'spin degree of freedom', and the total wavefunction being a product of space and spin parts, as is done in nonrelativistic quantum theory. Nevertheless, the non-relativistic decoupling of space and spin is usually simpler, as is illustrated by the spinor version of the Hydrogen atom, Appendix A.

ACKNOWLEDGMENT

Constructive comments of the referees have improved the paper, and are much appreciated.

REFERENCES

[1] Ohanian H C "What is spin?", Am. J. Phys. **54**, 500-505 (1986).

[2] Belinfante F J "On the spin angular momentum of mesons", Physica **6**, 887-898 (1939).

[3] Dirac P A M "The quantum theory of the electron", Proc. Roy. Soc. A **117**, 610-624 (1928).

[4] Dirac P A M "The quantum theory of the electron. Part II", Proc. Roy. Soc. A **118**, 351-361 (1928).

[5] Griffiths D J, *Introduction to quantum mechanics*, 2ed (Prentice Hall, New Jersey, 2005), Section 4.4.

[6] Lévy-Leblond J-M "Nonrelativistic particles and wave equations", Commun. Math. Phys. **6**, 286-311 (1967).

[7] Galindo A and del Rio C S "Intrinsic magnetic moment as a nonrelativistic phenomenon", Am. J. Phys. **29**, 582-584 (1961).

[8] Lekner J, *Theory of electromagnetic pulses* (Institute of Physics Publishing 2018), Section 2.6.

[9] Bliokh K Y, Dennis M R and Nori F "Relativistic electron vortex beams: angular momentum and spin-orbit interaction", Phys. Rev. Lett. **107**, 174802 (2011).

[10] Greiner W, *Quantum mechanics*, 3ed (Springer, Berlin, 1994), Chapter 13.

[11] Kennard E H "Zur Quantenmechanik einfacher Bewegungstypen", Z. Physik **44**, 326-352 (1927).

[12] Darwin C G "Free motion in wave mechanics", Proc. Roy. Soc. Lond. **A117**, 258-293 (1927).

[13] Lekner J "Rotating wavepackets", Eur. J. Phys. **29**, 1121-1125 (2008).

[14] Gould R J "The intrinsic magnetic moment of elementary particles", Am. J. Phys. **64**, 597-601 (1996).

[15] Lévy-Leblond J-M "The total probability current and the quantum period", Am. J. Phys. **55**, 146-149 (1987).

[16] Mita K "Virtual probability current associated with the spin", Am. J. Phys. **68**, 259-264 (2000).

[17] Landau L D and Lifshitz E M, *Quantum mechanics*, 2ed. (Pergamon, Oxford, 1985).

APPENDIX A. THE HYDROGEN ATOM IN SPINOR FORM

The equations (3.1) become, with E now an energy eigenvalue, no longer a time derivative,

$$(E+\frac{e^2}{r})\psi+\boldsymbol{\sigma}\cdot\boldsymbol{p}\chi=0, \quad \boldsymbol{\sigma}\cdot\boldsymbol{p}\psi+2m\chi=0 \tag{A.1}$$

$$\left[\frac{(\boldsymbol{\sigma}\cdot\boldsymbol{p})^2}{2m}-\frac{e^2}{r}\right]\psi=E\psi \quad \text{or} \quad [\frac{-\hbar^2}{2m}\nabla^2\text{-}e\frac{e^2}{r}]\psi=E\psi \tag{A.2}$$

Considering the non-degenerate ground state, with J_z eigenvalue $\frac{\hbar}{2}$, ψ_1 and ψ_2 must satisfy the same equation. This is not possible if we choose ψ_2 to have azimuthal dependence $e^{i\phi}$, as in Section 3, unless we also take ψ_2 to be zero. The ground state spinor now consists of ψ_1, the hydrogenic ground state 1S, and $\psi_2=0$, $\psi_3\sim\partial_z\psi_1$, $\psi_4\sim e^{i\phi}\partial_\rho\psi_1$. Because the Lévy-Leblond probability density is defined in terms of the first two spinor components ψ_1,ψ_2, and the probability density current can be expressed in terms of ψ_1,ψ_2, the hydrogenic ground state is, at least in the probability density and the probability density current, equivalent to the scalar ground state. The azimuthal dependence is hidden in the fourth spinor component.

For the first excited states we have a choice of 2S and 2P. The former is set up as above, the latter with $\psi_1=0$, and ψ_2 with $e^{\pm i\phi}$ dependence. Lévy-Leblond [15] and Mita [16] discuss the electron probability current of the 'stationary' states.

APPENDIX B. PROBABILITY DENSITY AND FLUX

In the Dirac case (Section 2), $\Psi^{+}\Psi$ is the probability density, and $\boldsymbol{S}=c\Psi^{+}\boldsymbol{\alpha}\Psi$, with $\boldsymbol{\alpha}$ is defined in (2.2). In the nonrelativistic formulation of Lévy-Leblond we have a time derivative of ψ but not of χ: $i\hbar\partial_t\psi+\boldsymbol{\sigma}\cdot\boldsymbol{p}\chi=0$, $\boldsymbol{\sigma}\cdot\boldsymbol{p}\psi+2m\chi=0$, or $\partial_t\psi\text{-}\boldsymbol{\sigma}\cdot\nabla\chi=0,\text{-}i\hbar\boldsymbol{\sigma}\cdot\nabla\psi+2m\chi=0$. To keep the norm time-independent Lévy-Leblond defines the probability density in terms of ψ only, as $\psi^{+}\psi$. The conservation law is now (Lévy-Leblond [6], Section IIIe)

$$\nabla.\boldsymbol{S}+\partial_t(\psi^{+}\psi)=0 \tag{B.1}$$

$$\partial_t(\psi^{+}\psi)=\psi^{+}(\boldsymbol{\sigma}\cdot\nabla\chi)+(\nabla\chi^{+}\cdot\boldsymbol{\sigma})\psi=\nabla\cdot(\psi^{+}\boldsymbol{\sigma}\chi+\chi^{+}\boldsymbol{\sigma}\psi) \tag{B.2}$$

Hence $\boldsymbol{S}(\boldsymbol{r},t)=\text{-}(\psi^{+}\boldsymbol{\sigma}\chi+\chi^{+}\boldsymbol{\sigma}\psi)$. We may express this current purely in terms of the top two spinor components ψ, since $\chi=\frac{i\hbar}{2m}\boldsymbol{\sigma}\cdot\nabla\psi$. This gives

$$\boldsymbol{S}(\boldsymbol{r},t)=\frac{\hbar}{2im}\{\psi^{+}\boldsymbol{\sigma}(\boldsymbol{\sigma}\cdot\nabla\psi)\text{-}(\boldsymbol{\sigma}\cdot\nabla\psi)^{+}\boldsymbol{\sigma}\psi\} \tag{B.3}$$

On using the commutation relations of the Pauli matrices, $\boldsymbol{\sigma}\times\boldsymbol{\sigma}=2i\boldsymbol{\sigma}$, the probability density current becomes

$$\boldsymbol{S}(\boldsymbol{r},t)=\frac{\hbar}{2im}[\psi^{+}\nabla\psi\text{-}(\nabla\psi^{+})\psi]+\frac{\hbar}{2m}\nabla\times(\psi^{+}\boldsymbol{\sigma}\psi) \tag{B.4}$$

The first term in this expression for $\boldsymbol{S}$ corresponds to the Schrödinger current in (3.3), the second is a spin current, which gives the correct g factor at leading order [6]. The spin term is the curl of a vector, and so does not contribute to the conservation law (B.1). See also Landau and Lifshitz [17] Section 114, and Mita [16] for the spin current term.

We shall calculate the radial, azimuthal, and longitudinal components of the probability current density, S_ρ,S_ϕ,S_z. The corresponding spin matrix components are

$$\sigma_\rho=\boldsymbol{\sigma}.\hat{\boldsymbol{\rho}}=\begin{pmatrix}0 & e^{-i\phi}\\ e^{i\phi} & 0\end{pmatrix}, \quad \sigma_\phi=\boldsymbol{\sigma}.\hat{\boldsymbol{\phi}}=\begin{pmatrix}0 & -ie^{-i\phi}\\ ie^{i\phi} & 0\end{pmatrix}, \quad \sigma_z=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{B.5}$$

Let f_1,f_2 be solutions of (3.8) and (3.9), respectively, and $\psi_1=f_1$, $\psi_2=e^{i\phi}f_2$. We can set $f_2=a\partial_\rho f_1$ [12]; a is a length parameter. We shall first calculate $\psi^{+}\boldsymbol{\sigma}\psi$; this has the cylindrical components $(2a\text{Re}\{f^{*1}\partial_\rho f_1\}$, $2a\text{Im}\{(\partial_\rho f^{*}{}_1)f_1\}$, $|f_1|^2\text{-}a^2|\partial_\rho f_1|^2)$. Note that there is no ϕ dependence. The curl of this vector is

$$\begin{aligned}\nabla\times(\psi^{+}\boldsymbol{\sigma}\psi)=(&\text{-}2a\partial_z\text{Im}\{(\partial_\rho f_1^{*})f_1\},\ 2a\partial_z\text{Re}\{f_1^{*}\partial_\rho f_1\}\text{-}\\ &\text{-}\partial_\rho[|f_1|^2\text{-}a^2|\partial_\rho f_1|^2],\ 2a\partial_\rho\text{Im}\{(\partial_\rho f_1^{*})f_1\}+2a\rho^{-1}\text{Im}\{(\partial_\rho f_1^{*})f_1\})\end{aligned} \tag{B.6}$$

When the length a is zero, just the azimuthal component remains, $\nabla\times(\psi^{+}\boldsymbol{\sigma}\psi)_{\text{a=0}}=(0,\text{-}\partial_\rho|f_1|^2,0)$. In that special case the Schrödinger current is proportional to $\text{Im}\{f^{*}{}_1\nabla f_1\}=\text{Im}\{f^{*}{}_1(\partial_\rho f_1,0,\partial_z f_1)\}$, and the components of the probability current density are given by

$$\frac{m}{\hbar}S_\rho=\text{Im}\{f^{*}{}_1\partial_\rho f_1\}, \quad \frac{m}{\hbar}S_\phi=\text{-}\frac{1}{2}\partial_\rho|f_1|^2, \quad \frac{m}{\hbar}S_z=\text{Im}\{f^{*}{}_1\partial_z f_1\} \tag{B.7}$$

As in the hydrogen ground state, the a=0 spinor now consists of ψ_1, and $\psi_2=0$, $\psi_3\sim\partial_z\psi_1$, $\psi_4\sim e^{i\phi}\partial_\rho\psi_1$. The fourth component contributes to the azimuthal current, and to the angular momentum.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: Tatini D., Raudino M., Sarri F. (2022) Light-Modulated Rheological Properties in Green Innovative Formulations. *Substantia* 6(2): 15-26. doi: 10.36253/Substantia-1736

Received: Mar 07, 2022

Revised: Jul 11, 2022

Just Accepted Online: Jul 12, 2022

Published: September 1, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Light-Modulated Rheological Properties in Green Innovative Formulations

Duccio Tatini*, Martina Raudino, Filippo Sarri

Department of Chemistry "Ugo Schiff" and CSGI, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino (Firenze), Italy
*E-mail: duccio.tatini@unifi.it

Abstract. The addition of azorubine to a viscoelastic aqueous dispersion of sodium oleate (NaOL, 0.43 M, 13% w/w) and KCl (up to 4% w/w) leads to a green gel-like system whose rheological behavior can be efficiently and reversibly triggered from remote by using UV light. Rheology, Differential Scanning Calorimetry (DSC) measurements and phase behavior studies indicate that the original texture of the NaOL dispersion is significantly hardened upon UV irradiation for 8 hours in the presence of azorubine, showing a seven hundred-fold increase in viscosity. The UV treatment brings about the *trans* to *cis* isomerization of azorubine, which modifies the structure of the NaOL wormlike micellar system, leading to a more entangled, close-textured network. The cooperative effect of KCl on the fluid viscosity is found to be concentration-dependent. The system slowly reverts to its original rheological behaviour after standing for about 1 day. These results are relevant for the development of stimuli-responsive innovative systems based on biocompatible, non expensive and commercially available materials that can be used in a wide range of applications, such as in drug delivery or enhanced oil recovery, where a quick change in the physico-chemical features of the system is required but difficult to be performed.

Keywords: green chemistry, sodium oleate, azorubine, viscosity, stimulus-responsive.

INTRODUCTION

Responsive or "smart" materials are functional materials whose properties can undergo controlled and reversible changes in response to an external stimulus [1–4]. The applied stimulus or external field include thermal, electrical, magnetic, pH, UV-visible light, ionic or metallic interactions or combinations thereof [5,6]. The formulation of gel-based stimuli-responsive systems with specific performances is crucial for a great number of applications, and particularly when it is impossible or very difficult to switch on/off their properties, and a remotely controlled trigger is necessary [5,7].

Among these systems, viscoelastic surfactant (VES) solutions have attracted great attention due to their unique features and versatility that can be harnessed in a variety of high-tech and everyday applications [8–15].

The amphiphilic nature of VES molecules leads to the self-assembly in solution into small aggregates, which show a complex phase behavior: the simplest structure are spherical micelles, but also hexagonal, lamellar, vesicular, cubic, reverse phases can be observed, depending on the surfactant packing parameter and global packing constraints [2,16–19].

Under certain conditions (*i.e.* surfactant concentration, salinity, temperature, presence of different counterions, change in the composition of the solvent, etc.) spherical micelles may undergo uniaxial growth and form elongated and flexible structures, usually referred to as "wormlike" micelles (WLM) [19–21].

Above a critical concentration these systems show viscoelastic properties, like polymer solutions or bicontinuous three-component ionic microemulsions, as a result of the formation of a densely entangled network [19,22–26].

In our previous series of works we extensively investigated the main physicochemical properties and the phase behavior of wormlike micellar systems based on sodium oleate (NaOL), a safe, eco-friendly and cost-effective surfactant [19,26]. In the first part we reported on the structure, thermal properties and rheological behavior of NaOL aqueous dispersions in the presence of a single salt (KCl) via cryo-TEM, rheology and DSC experiments [19]. In part 2 we illustrated the specific ion effect induced by the addition of different salts on viscoelastic dispersions of sodium and potassium oleate, and we systematically discussed it in terms of the Hofmeister series [26]. In a previous work we imparted a voltage-dependent responsiveness to an NaOL aqueous dispersion through the addition of carbon black particles [27].

Prompted by these studies on green formulations with suitable responsiveness to different physical stimuli, in the present contribution we developed a NaOL-based viscoelastic dispersions that, in the presence of minimal amounts of azorubine, a biocompatible dye, and a salt (KCl) undergo a remarkable change in their rheological properties upon irradiation at a specific wavelength.

NaOL finds application in a wide number of industrial products and commercial formulations, like healthcare products, cleansers, thickening agents, emulsifiers, lubricants and fluids for enhanced oil recovery [28–32]. Thanks to its negative charge, NaOL is more biodegradable and less harmful for the environment compared to cationic surfactants [33].

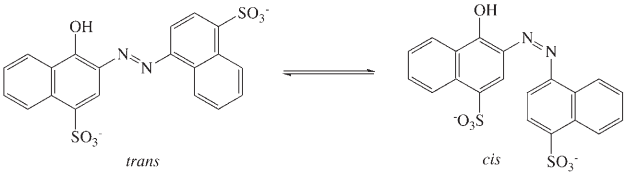


Figure 1. Trans-cis isomerization of azorubine.

NaOL shows a very interesting phase behavior [34] and forms different nano- and micro-structures in solution upon the addition of electrolytes or as a result of pH variation [19]. For an extensive discussion about the structural features and the rheological behavior of NaOL dispersions please see references [19,26,35–41] and references therein.

In this work we report on a moderately concentrated (0.43 M, 13% w/w) dispersion of NaOL in water that gives rise to a wormlike micellar network with peculiar structural properties and rheology. The choice of this surfactant concentration is related to the NaOL/water systems' phase diagram: at this concentration and 25° C the dispersion converts from a simple fluid micellar L_1 to a viscous L_1^* phase that shows shear-dependent birefringence [19].

Azorubine is a synthetic azo dye approved for food decorations and coatings and as drink additive [42]. Its chemical structure is shown in Figure 1. The presence of the azo moiety enables a *trans-cis* isomerization upon irradiation with light at an appropriate wavelength, usually in the UV region [43–46].

The process may revert spontaneously upon heating since the *trans* isomer is thermodynamically more stable, or can be induced through irradiation with a visible light [47].

Previous studies reported on the introduction of different chromophores in VES systems in order to obtain light-responsive fluids with tunable rheological properties. These photo-active molecules include synthetically modified azobenzenes, [48–54] *p*-coumaric acid [55,56] and cinnamic acid derivatives [57–59]. We selected azorubine because of its unique advantages in terms of availability, simple manipulation, low cost and, above all, safety.

This work is a proof-of-concept that shows the efficacy of combining completely biocompatible and nontoxic materials to obtain a versatile formulation with a viscosity and rheological behaviour that can be remarkably modified through the irradiation with UV light.

Moreover, all these features make these systems very attractive for a wide range of applications, including agriculture, food industry, cosmetics, "smart" materials, enhanced oil recovery, shale gas extraction, drug delivery and controlled release.

MATERIALS AND METHODS

Materials

Sodium oleate (ACS reagent grade, Riedel-De Haën, Seelze, Germany) and potassium chloride (> 99 %, Sigma-Aldrich, Milan, Italy) were used as received without any further purification. Azorubine (Carmoisine, Food red 3 or E 122, disodium 4-hydroxy-3-((4- sulphonatonaphthyl)azo) naphthalenesulfonate), food grade quality) was supplied by F.lli Rebecchi S.r.l. (Piacenza, Italy) and used without any further purification. All solutions and dispersions were prepared with Milli-Q water (resistivity > 18 MΩ cm at 25°C).

Sample Preparation

Sodium oleate viscoelastic formulations were prepared by the addition of a weighted amount of surfactant to KCl aqueous solutions at different concentrations (0, 0.1, 0.5, 1, 2, 3, 3.5, 4 % w/w) under constant stirring at room temperature. The final concentration of NaOL was 0.43 M (13 % w/w) in all the samples. The dye-loaded samples were prepared following a similar procedure: to a 0.18 % w/w ($3.6\ 10^{-3}$ M) azorubine aqueous solution the proper amount of potassium chloride and then of the surfactant were added. The final concentrations of KCl were 0.1, 0.5, 1, 2, 3, 3.5, 4% w/w. For all the samples we used Milli-Q water, which was boiled for 4 h, filtrated and stocked under argon. As reported in our previous works the samples were freshly prepared and tested within 1 hour for the rheological and DSC experiments [19,26]. All these precautions are necessary in order to avoid the uptake of atmospheric CO_2 by the samples [60].

UV Irradiation

About 10 mL of sample were placed in a quartz container and irradiated using a Camag UV lamp (wavelength 254 nm, 8 W, Muttenz, Switzerland) for 8 hours. All the experiments were conducted at 25 °C, and the distance between the sample and light source was fixed at 5 cm.

Rheology Measurements

Rheology experiments were carried out on a Paar Physica UDS 200 rheometer using a plate-plate geometry with a diameter of 40 mm and a measurement gap of 300 µm. The temperature was fixed at 25.0 ± 0.1 °C using a Peltier control system for all the measurements. The samples were equilibrated for 15 min at the set temperature before being tested. Frequency sweep measurements were carried out within the linear viscoelastic range at a strain value of 1%, which was previously determined by means of an amplitude sweep test. The storage and loss moduli (G' and G", respectively) were measured over the frequency range of 10^{-3} and 10^{2} Hz. The flow curves were acquired in a torque range between 10^{-1} and $5\cdot10^{3}$ mN·m. The experimental viscosity (η) was fitted with the Cross model (see the Results and Discussion) to obtain the zero-shear viscosity (η_0), the infinite-shear viscosity (η_∞), the shear relaxation exponent (m) and the consistency (C). The experiments were repeated at least three times, and silicon oil was applied to the rim of the measurement geometry to prevent water evaporation from the sample.

Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) was performed by means of a DSC-Q2000 by TA Instruments (Philadelphia, PA). The samples were sealed in aluminum hermetic pans and the measurements were conducted under N_2 atmosphere, with a flow rate of 50 mL/min. The samples were first cooled from 20 °C to − 60 °C at 10 °C/min, then heated up to 50 °C at 5°C/min. The thermograms which show overlapping endothermic peaks were analyzed using a linear combination of exponentially modified Gaussian (EMG) functions, as reported elsewhere [19,26,61,62].

RESULTS AND DISCUSSION

Flow curve experiments

Figure 2 shows the flow curves obtained before (solid lines) and after (dashed lines) UV irradiation from the aqueous dispersions of sodium oleate (NaOL, 13% w/w) in the presence of KCl at different concentrations (0, 0.5, 2 and 3% w/w). The curves for the other KCl concentra-

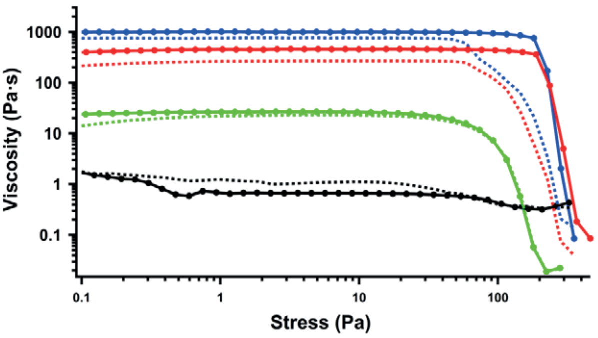


Figure 2. Flow Curves before (solid lines) and after (dashed lines) UV irradiation for 13% w/w NaOL (0.43 M, black) and in the presence of KCl at 0.5% (green), 2% (blue) and 3% w/w (red). The flow curves for non-irradiated samples are reprinted from [19] with permission from Elsevier. Copyright 2021.

tions are reported in Figure S1 in the Supporting Information. The experimental data on non-irradiated samples are reproduced from [19] with permission.

All the samples show the typical flow behavior of wormlike micellar networks, *i.e.,* a Newtonian plateau followed by a steep decrease in the viscosity at high shear stress, due to the shear-induced alignment of cylindrical aggregates. Moreover, a concentration-dependent increase in the viscosity is observed upon the addition of KCl [19,63]. The fitting of the experimental viscosity (η) data was performed using the Cross model [64]:

$$\eta = \eta_\infty + \frac{\eta_\infty - \eta_0}{1 + (C\dot{\gamma})^m} \tag{1}$$

Here η_0 is the zero-shear viscosity, m the shear relaxation exponent, η_∞ the infinite-shear viscosity, the shear rate and C the consistency. The extracted fitting parameters are reported in Table S1 in the Supporting Information. The results nicely agree with the literature data and confirm the shear thinning behavior of the dispersions, originating from the micelles' entanglement and their progressive alignment at high shear rates [19,26,34,65,66].

The UV irradiation does not affect the flow properties of the formulations: the profiles, as well as the values of η_0 exhibit minor fluctuations within the experimental uncertainty before and after the UV treatment.

The flow curves for the NaOL samples at different KCl concentration (0, 0.1%, 0.5% and 4%) in the presence of 0.18% w/w (3.6 10^{-3} M) azorubine are reported in Figure 3. The flow curves at the other KCl concentrations are shown in Figure S2 (see the Supporting Information).

In the presence of azorubine all the formulations exhibit a shear-thinning behaviour that is similar to that found for the NaOL-KCl systems. The results obtained from the fitting of the experimental data with the Cross model are reported in Table S1 in the Supporting Information. After UV irradiation the viscosity of the samples with 0.1%, 0.5% and 4% KCl increases from 6.94 to 13.6, from 278.8 to 728.3 and from 1.852 to 730.1 Pa·s, respectively. For the other KCl concentrations no significant variations in the formulation viscosity are observed after the UV treatment.

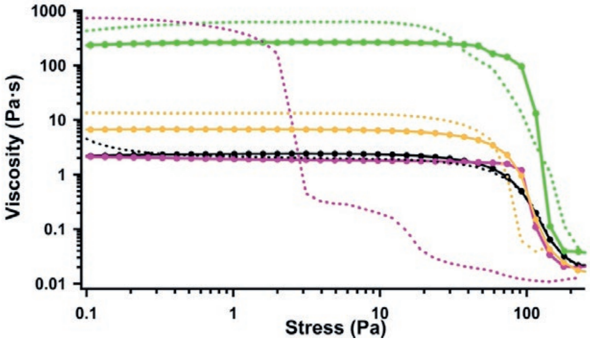


Figure 3. Flow curves before (solid lines) and after (dashed lines) UV irradiation for NaOL 0.43 M mixture in the presence of 0.18% w/w azorubine (3.6·10^{-3} M) at different KCl concentration: 0 (black), 0.1% (orange), 0.5% (green) and 4% w/w (pink).

These findings lead to two important conclusions: (i) azorubine has a remarkable effect as photo-active molecule in modifying the viscosity of the fluids after UV irradiation. A similar light-modulated viscosity change is reported for binary mixtures of NaOL and a light-responsive cationic azobenzene dyes [50,54]. In the work of Lu *et al.* the UV irradiation and the resulting *trans*-to-*cis* isomerization of the azo dye (1-[2-(4-phenylazophenoxy)-ethyl]-3-methylimidazolium bromide) induces a decrease in the viscosity of the system, with a transition from a gel-like structure to a Newtonian fluid [50]. In our case the opposite effect on the viscosity and the strengthening of the wormlike three-dimensional network is observed: a similar behavior is reported by Liu and coworkers on dilute NaOL dispersions in the presence of three different imidazolium surfactants upon UV irradiation [54]. (ii) The presence of KCl plays a key role in modulating the effect of azorubine. The experimental data reported in Table S1 show that the *trans*-to-*cis* isomerization of azorubine induces an increase in the formulation viscosity when KCl concentration is lower than 1% w/w. Between 1% and 3.5% no remarkable changes are observed before and after the UV irradiation, suggesting that the major contribution to the strength of the system is provided by the salt.

The behavior of the sample containing azorubine and 4% KCl is peculiar and deserves a deeper analysis. The viscosity increases by two orders of magnitude, but the rheological profile shows some differences after the UV treatment (see Figure 2, pink curves). A first Newtonian plateau appears in the low stress regime, then the viscosity rapidly decreases at a critical shear stress which is considerably lower than the viscosity breakdown point before the irradiation. After this initial decrease the flow curve shows a second, less pronounced plateau, followed by the typical shear-thinning region at high stress values. A similar behavior was reported by Griffiths *et al.* in 2004 for carbon black particles dispersed in an acrylic polymeric matrix [67]. In this case the presence of an intermediate (secondary) Newtonian plateau was ascribed to the interaction between the polymer layer and the polymer matrix, in particular to the viscous drag between the polymer chains adsorbed on the particles with the polymer in solution. In a more recent work Sochi depicts this intermediate region as a characteristic feature of viscoelastic fluids in porous media flow, that may be attributed to the time-dependent nature of the viscoelastic fluid when the relaxation time of the fluid and the characteristic time of the flow become comparable [68]. Polacco and coworkers observed two distinct shear-thinning phenomena in polymer-modified asphalts: the first shear-thinning was ascribed to a rigid rearrangement of the aggregate structure, that involves a temporary detachment of polymer chains. As a result, the polymer is able to move between different domains, inducing a transitory strengthening of the network [69].

In our system the appearance of this secondary Newtonian plateau may reflect the formation of ordered structures with a different degree of organization. To the best of our knowledge this is the first time that such behavior is found and reported for wormlike micellar dispersions. As a matter of fact the cited literature sources deal with polymeric blends [67,69] or to a more general viscoelastic behavior in porous media flow [68]. Further work is necessary to deepen and clarify this phenomenon. The high salt concentration may be responsible for the formation of complex, more entangled structures after UV exposure. These structures are responsible for the remarkable increase in the viscosity of the formulation and exhibit a low resistance to the applied stress, as confirmed by the low critical stress value. This hypothesis is also confirmed by the fact that during the sample preparation the azorubine solution becomes turbid at this salt concentration, suggesting the formation of aggregated structures. The same salt-induced behavior is observed in pure azorubine solutions, without sodium oleate, and presumably reflects the formation of piled up structures stabilized by π-stacking interactions [70,71].

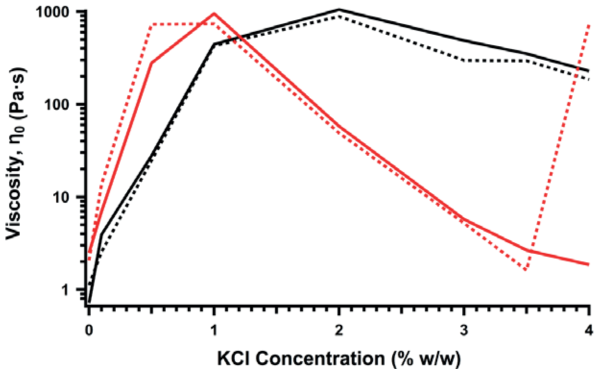


Figure 4. Zero-shear viscosity (η_0) for the formulations containing 0.18% w/w azorubine (red) and the reference samples without the dye (black) as a function of the salt concentration (a) before (solid line) and after (dashed line) UV irradiation. The trend for non-irradiated NaOL-KCl samples (black solid line) are reprinted from [19] with permission from Elsevier. Copyright 2021.

For all the irradiated samples the system recovers its original rheological behavior after standing for about 1 day, indicating a complete reversibility of the process.

Figure 4 reports the values of η_0 as a function of KCl concentration before (solid lines) and after (dashed lines) the UV treatment.

For the two sets of samples the viscosity steeply increases in the dilute regime, then it reaches a maximum and progressively decreases at higher concentrations of salt. This behavior has been widely reported for a large number of wormlike micellar systems [22,39,50,63,72–80]. The initial viscosity increase is ascribed to the formation, growth and entanglement of the cylindrical aggregates. The decrease in the fluid viscosity after the maximum upon salt addition is due to the lateral branching along the rod-like micelles, that provide an extra route for mechanical stress relief [72,81,82]. A detailed discussion about the thermodynamic justification and the driving forces behind branches formation can be found in our previous works [19,26].

The experimental results show that azorubine induces two distinct effects on the salt curves (Figure 3). The first effect is the shift of the peak maximum to lower KCl concentrations. Azorubine is an amphiphilic molecule that can penetrate across the micellar interface at least partially. This results in a flattening of the micelle/water interface with a significant lowering in the surface curvature. Moreover, the electrostatic repulsion between different tubular micelles is screened and the viscosity increases even at lower concentrations of salt.

The second effect is the modification of the salt curve shape after the maximum. In the presence of azorubine the viscosity decreases more rapidly, as a consequence of the increased branching density. Rogers *et al.* reported that for several mixed surfactant/salt viscoelastic systems, the branching effect is more evident with sodium salicylate, a hydrotropic salt that can penetrate more efficiently underneath the water/micellar interface salt respect to a simple inorganic salt like KCl [83]. This effect due to branched structures is quite common and found also in nonionic or zwitterionic systems [41].

Oscillatory-shear measurements

Oscillatory shear experiments were performed to explore the viscoelastic behavior of NaOL-KCl-azorubine systems upon UV irradiation. The storage (*G*') and loss (*G*") moduli before and after UV irradiation for the dispersions of sodium oleate in the presence of KCl at different concentrations are shown in Figure 5. The fre-

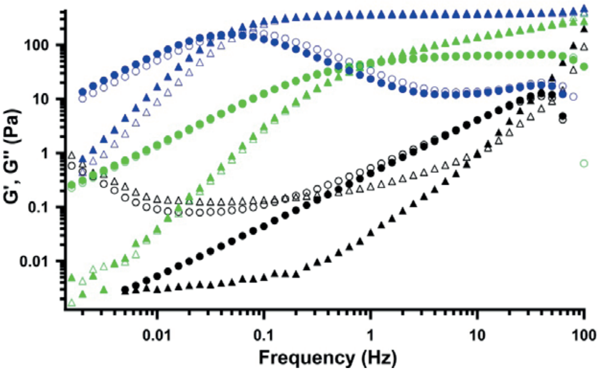


Figure 5. Storage (triangles) and loss (circles) moduli before (filled markers) and after (empty markers) UV irradiation for NaOL 0.43 M in the presence of KCl at 0 (black), 0.5% (green) and 2% (blue). The frequency sweep curves for non-irradiated samples are reprinted from [19] with permission from Elsevier. Copyright 2021.

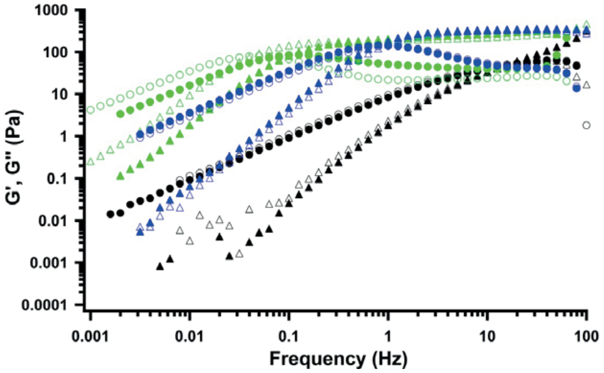


Figure 6. Storage (triangles) and loss (circles) moduli before (filled markers) and after (hollow markers) UV irradiation for NaOL 0.43 M + azorubine 0.18 % mixture in the presence of KCl at 0 (black), 0.5 % (green) and 2 % (blue).

quency sweep curves for other KCl concentrations are reported in the Supporting Information.

For all the samples two distinct regimes are observed, *i.e.* a predominant viscous behavior at low frequencies (*G*" > *G*') and mainly an elastic behavior at higher frequencies (*G*" < *G*'). The *crossover point* of the storage and loss moduli (and the corresponding crossover frequency ω_c) marks the transition between the two different regions in the viscoelastic spectrum [84]. The decrease in ω_c upon salt addition reflects the formation of a more entangled WLM structure with slower relaxation mechanisms [19]. No remarkable variations in *G*' and *G*" are observed before and after the UV irradiation with the exception of NaOL alone, that exhibits higher values of the storage and loss moduli in the low frequency region after the UV treatment.

Figure 6 reports the storage and loss moduli obtained in the frequency sweep experiments before and after UV irradiation for the samples containing azorubine in the presence of KCl at different concentrations (0, 0.5 and 2%). The frequency sweep curves for other KCl concentrations are reported in the Supporting Information. The formulations exhibit a viscoelastic behavior that looks similar to those reported in Figure 5, indicating that the two frequency-dependent regimes are present. The crossover frequency decreases when KCl concentration is increased from 0 to 0.5%, then it progressively shifts to higher values upon further increase in the salt concentration.

The UV treatment has a very minor effect on the viscoelastic properties of the formulations, with the exception of the sample containing KCl 4% (Figure S10 in the Supporting Information). In this case before the UV irradiation the storage and loss moduli show a crossover in the high-frequency region, followed by a remarkable drop at medium and low frequencies, which indicates a predominantly viscous behavior. After the UV treatment *G*' and *G*" overlap for most part of the whole frequency range and after a slight decrease at high frequencies they level off to a relatively high constant value in the medium and low-frequency regime. This result is consistent with the remarkable viscosity increase observed after the irradiation (Figure 2) due to the presence of light-induced ordered structures.

A Cole-Cole plot analysis was performed to describe the viscoelastic behavior of the NaOL dispersions in terms of a Maxwell model with a single relaxation time (results not shown). Unfortunately, this approach does not provide an accurate prediction for the experimental values due to the presence of additional relaxation modes. For this reason, we calculated the continuous time-weighted relaxation spectrum using the values of the storage and loss moduli (Trios software, 5.2 version, TA instruments) [85,86]. The relaxation times τ_R estimated from the spectra are reported in Figure 7.

In the relaxation spectra an intense primary peak is observed for all the samples, and it was used to extrapolate the relaxation times. Additional secondary peaks are clearly detectable, demonstrating the presence of concurrent relaxation modes (reptation, breaking and recombination, Rouse motion, etc.) [19].

The comparison between the relaxation times and the viscosity values (see Figure 4) shows a similar effect on both the flow and the viscoelastic properties of the dispersions induced by azorubine: the UV irradiation, thanks to the presence of the dye, increases the structuredness of the tridimensional micellar network, resulting in similar trends of η_0 and τ_R.

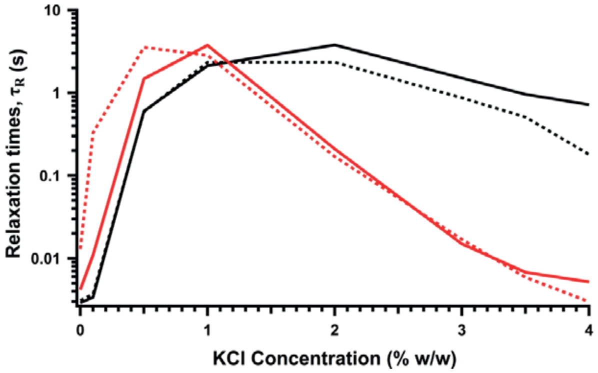


Figure 7. Relaxation time (τ_R) for the formulations containing azorubine (red) and the reference samples without the dye (black) as a function of the salt concentration before (solid line) and after (dashed line) the UV irradiation. The trend for non-irradiated NaOL-KCl samples (black solid line) are reprinted from [19] with permission from Elsevier. Copyright 2021.

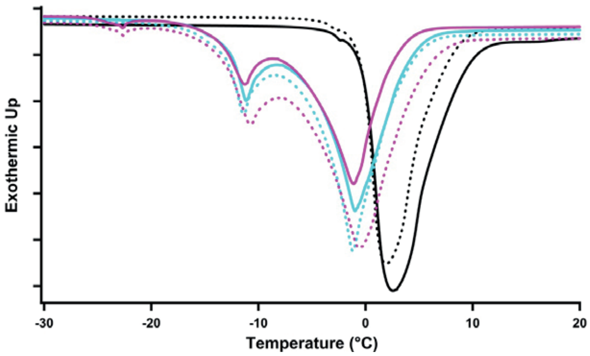


Figure 8. DSC heating curves before (solid lines) and after (dashed lines) UV irradiation for NaOL 0.43 M mixture in the presence of KCl at 0 (black), 3.5% (light blue) and 4% (pink).

The only exception is represented by NaOL + KCl 4% in the presence of azorubine: in this case the relaxation time after the UV treatment is very close to the value of the non-irradiated sample. As we mentioned in the previous section, the remarkable increase in the viscosity observed at this salt concentration is probably due to the presence of aggregated structures.

Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) experiments were performed to investigate the thermal properties of the NaOL-KCL-azorubine systems. The thermograms of the NaOL dispersions in the presence of KCl at different concentrations (0, 3.5 and 4%) before and after UV irradiation are reported in Figure 8. For the DSC curves of the other investigated KCl concentrations see the Supporting Information.

All the DSC thermograms show a free water melting peak at around 0°C. For KCl concentrations above 0.5 % a secondary endothermic peak occurs between -15 and -11°C, and it is related to the melting of the "interfacial" water molecules that are confined in the solvent-rich domains between the entangled cylindrical micelles [87–91]. A third endothermic peak appears when KCl concentration reaches 4%, as observed in our previous work [19]. This thermal transition is attributed to the melting of freezable bound water, *i.e.* the water molecules that closely interact with the micellar surface and show a melting temperature remarkably different from bulk water [61,92–96].

The melting temperature, the relative enthalpy change and amount (%) of free water (ΔH_{mf}, T_{mf} and W_f), interfacial water (ΔH_{mi}, T_{mi} and W_i) and freezable bound water (ΔH_{mb}, T_{mb} and W_b) for all the examined samples, before and after UV irradiation are listed in Table S2 in the Supporting Information. A detailed discussion about the theoretical background and the procedures for the calculation of the thermal parameters can be found in [19,26,61].

For the pristine non-irradiated samples W_f rapidly decreases, passing from 96% to 46% as the salt concentration increases. Similar values are obtained for the NaOL-KCl systems after the UV irradiation, demonstrating that the endothermic process (*i.e.* the melting of the free water) is not affected by the UV treatment, as expected.

When KCl concentration increases, the melting peak temperatures T_{mi} shift to lower values; conversely ΔH_{mi} shows a progressive increase. As previously reported, the addition of KCl (and the consequent Na^+/K^+ exchange at the micellar surface) leads to the formation of elongated cylindrical micelles, and above a critical concentration (see Figure 4) to the branching of the WLM network [19]. For this reason the number of connections and junctions within the micellar network increase, as well as the number of water molecules confined within the intermicellar domains, as evidenced by the values of W_i.

By comparing the values of the peak temperatures and of the enthalpy changes related to the melting of the interfacial and bound water we conclude that the UV treatment does not alter the hydration state of the NaOL/KCl systems.

For the samples containing azorubine the values of ΔH_{mf}, T_{mf}, W_f, ΔH_{mi}, T_{mi}, W_i, ΔH_{mb}, T_{mb} and W_b are reported in Table S2 in the Supporting Information.

The presence of the dye does not induce any remarkable variation in the amount of free and interfacial water, as well as in the values of peak temperatures and the relative enthalpy changes. The main difference is represented by the freezable bound water melting peak, that appears starting from KCl 3%. As indicated by the viscosity measurements, the negatively charged sulfonate groups in the azorubine molecule have a cooperative effect with KCl in screening the electrostatic repulsion between the micellar surfaces. This synergistic action reduces the amount of salt that is required for the formation of the micellar entangled network (and eventually the branched structures) and affects the thermal transition of the water molecules that closely interacts with the micellar surface, *i.e.* the freezable bound water.

CONCLUSIONS

In this work we illustrate a simple and inexpensive procedure for the formulation of green photo-responsive viscoelastic fluids starting from non-toxic, biocompatible, commercially available materials. The addition of a food azo dye, azorubine, to a dispersion of sodium oleate in the presence of KCl enables the modification of the formulation rheological properties through an external UV light stimulus. The UV treatment brings about the *trans-cis* isomerization of azorubine, which is partially intercalated between the surfactant polar heads, resulting in a modification of the wormlike micellar structure of NaOL. This gives rise to a remarkable increase in the formulation viscosity and a variation of the rheological properties, as evidenced by viscosity and oscillatory-shear measurements. The effect of azorubine is mediated by the presence of the salt, since it occurs only in a specific range of KCl concentration. DSC measurements confirm the formation of wormlike micelles above a critical salt concentration, which is lower in the presence of azorubine. The UV treatment does not affect the thermal transitions of the NaOL dispersions, since no significant variation is detected before and after the irradiation.

All these features, combined with the complete biocompatibility and non-toxicity break new ground in a wide range of applications, from shale gas extraction to controlled release and drug delivery, where a remote control on the mechanical and physico-chemical properties of the formulation is crucial.

ACKNOWLEDGEMENTS

The authors acknowledge funding from the European Union Horizon 2020 research and innovation program “Shale for Environment” (SXT) under grant agreement no. 640979.

REFERENCES

[1] M. Wei, Y. Gao, X. Li, M.J. Serpe, Stimuli-responsive polymers and their applications, Polym. Chem. 8 (2016) 127–143. https://doi.org/10.1039/C6PY01585A.

[2] C.A. Dreiss, Wormlike micelles: where do we stand? Recent developments, linear rheology and scattering techniques, Soft Matter. 3 (2007) 956–970. https://doi.org/10.1039/B705775J.

[3] S. Guragain, B.P. Bastakoti, V. Malgras, K. Nakashima, Y. Yamauchi, Multi-Stimuli-Responsive Polymeric Materials, Chem. Weinh. Bergstr. Ger. 21 (2015) 13164–13174. https://doi.org/10.1002/chem.201501101.

[4] E. Moulin, L. Faour, C.C. Carmona-Vargas, N. Giuseppone, From Molecular Machines to Stimuli-Responsive Materials, Adv. Mater. 32 (2020) 1906036. https://doi.org/10.1002/adma.201906036.

[5] S. Ahn, R.M. Kasi, S.-C. Kim, N. Sharma, Y. Zhou, Stimuli-responsive polymer gels, Soft Matter. 4 (2008) 1151–1157. https://doi.org/10.1039/B714376A.

[6] D. Wang, G. Wei, R. Dong, J. Hao, Multiresponsive Viscoelastic Vesicle Gels of Nonionic C12EO4 and Anionic AzoNa, Chem. – Eur. J. 19 (2013) 8253–8260. https://doi.org/10.1002/chem.201300132.

[7] M.A.C. Stuart, W.T.S. Huck, J. Genzer, M. Müller, C. Ober, M. Stamm, G.B. Sukhorukov, I. Szleifer, V.V. Tsukruk, M. Urban, F. Winnik, S. Zauscher, I. Luzinov, S. Minko, Emerging applications of stimuli-responsive polymer materials, Nat. Mater. 9 (2010) 101–113. https://doi.org/10.1038/nmat2614.

[8] J. Yang, Viscoelastic wormlike micelles and their applications, Curr. Opin. Colloid Interface Sci. 7 (2002) 276–281. https://doi.org/10.1016/S1359-0294(02)00071-7.

[9] H. Rehage, H. Hoffmann, Rheological properties of viscoelastic surfactant systems, J. Phys. Chem. 92 (1988) 4712–4719. https://doi.org/10.1021/j100327a031.

[10] H. Shi, Y. Wang, B. Fang, Y. Talmon, W. Ge, S.R. Raghavan, J.L. Zakin, Light-Responsive Threadlike Micelles as Drag Reducing Fluids with Enhanced Heat-Transfer Capabilities, Langmuir. 27 (2011) 5806–5813.

[11] M. Samuel, D. Polson, D. Graham, W. Kordziel, T. Waite, G. Waters, P.S. Vinod, D. Fu, R. Downey, Viscoelastic Surfactant Fracturing Fluids: Applications in Low Permeability Reservoirs, in: Society of Petroleum Engineers, 2000. https://doi.org/10.2118/60322-MS.

[12] C. Moitzi, N. Freiberger, O. Glatter, Viscoelastic Wormlike Micellar Solutions Made from Nonionic Surfactants: Structural Investigations by SANS and DLS, J. Phys. Chem. B. 109 (2005) 16161–16168. https://doi.org/10.1021/jp0441691.

[13] Z. Zhai, X. Yan, J. Xu, Z. Song, S. Shang, X. Rao, Phase Behavior and Aggregation in a Catanionic System Dominated by an Anionic Surfactant Containing a Large Rigid Group, Chem. – Eur. J. 24 (2018) 9033–9040. https://doi.org/10.1002/chem.201800628.

[14] P. Brown, C.P. Butts, J. Eastoe, Stimuli-responsive surfactants, Soft Matter. 9 (2013) 2365–2374. https://doi.org/10.1039/C3SM27716J.

[15] H. Sakai, K. Tsuchiya, K. Sakai, Stimuli-Responsible Viscoelastic Surfactant Solutions, in: Stimuli-Responsive Interfaces, Springer, Singapore, 2017: pp. 19–36. https://doi.org/10.1007/978-981-10-2463-4_3.

[16] B.W. Ninham, K. Larsson, P. Lo Nostro, Two sides of the coin. Part 1. Lipid and surfactant self-assembly revisited, Colloids Surf. B Biointerfaces. 152 (2017) 326–338. https://doi.org/10.1016/j.colsurfb.2017.01.022.

[17] B.W. Ninham, K. Larsson, P. Lo Nostro, Two sides of the coin. Part 2. Colloid and surface science meets real biointerfaces, Colloids Surf. B Biointerfaces. 159 (2017) 394–404. https://doi.org/10.1016/j.colsurfb.2017.07.090.

[18] B.W. Ninham, The Biological/Physical Sciences Divide, and the Age of Unreason, Substantia. 1 (2017) 7–24. https://doi.org/10.13128/Substantia-6.

[19] D. Tatini, M. Raudino, M. Ambrosi, E. Carretti, I. Davidovich, Y. Talmon, B.W. Ninham, P. Lo Nostro, Physicochemical characterization of green sodium oleate-based formulations. Part 1. Structure and rheology, J. Colloid Interface Sci. 590 (2021) 238–248. https://doi.org/10.1016/j.jcis.2021.01.040.

[20] Y. Lin, Y. Qiao, Y. Yan, J. Huang, Thermo-responsive viscoelastic wormlike micelle to elastic hydrogel transition in dual-component systems, Soft Matter. 5 (2009) 3047–3053. https://doi.org/10.1039/B906960G.

[21] P.L. Nostro, N. Stubicar, S.-H. Chen, Isotopic Effect in Phase Separation of Dioctanoylphosphatidylcholine/Water Micellar Solutions, Langmuir. 10 (1994) 1040–1043. https://doi.org/10.1021/la00016a015.

[22] T.M. Clausen, P.K. Vinson, J.R. Minter, H.T. Davis, Y. Talmon, W.G. Miller, Viscoelastic micellar solutions: microscopy and rheology, J. Phys. Chem. 96 (1992) 474–484. https://doi.org/10.1021/j100180a086.

[23] P.A. Hassan, S.R. Raghavan, E.W. Kaler, Microstructural Changes in SDS Micelles Induced by Hydrotropic Salt, Langmuir. 18 (2002) 2543–2548. https://doi.org/10.1021/la011435i.

[24] R. Byron Bird, P.J. Carreau, A nonlinear viscoelastic model for polymer solutions and melts—I, Chem. Eng. Sci. 23 (1968) 427–434. https://doi.org/10.1016/0009-2509(68)87018-6.

[25] R.H. Colby, Structure and linear viscoelasticity of flexible polymer solutions: comparison of polyelectrolyte and neutral polymer solutions, Rheol. Acta. 49 (2010) 425–442. https://doi.org/10.1007/s00397-009-0413-5.

[26] D. Tatini, D. Ciardi, C. Sofroniou, B.W. Ninham, P. Lo Nostro, Physicochemical characterization of green sodium oleate-based formulations. Part 2. Effect of anions, J. Colloid Interface Sci. 617 (2022) 399–408. https://doi.org/10.1016/j.jcis.2022.01.135.

[27] F. Sarri, D. Tatini, M. Raudino, M. Ambrosi, E. Carretti, P. Lo Nostro, Electro-Responsive Green Gels for Lower Environmental Impact Shale Gas Extraction, Energy Fuels. 33 (2019) 2057–2066. https://doi.org/10.1021/acs.energyfuels.8b04321.

[28] O.P. Artykulnyi, V.I. Petrenko, L.A. Bulavin, L. Almasy, N.A. Grigoryeva, M.V. Avdeev, V.L. Aksenov, On the Impact of Polyethylene Glycol on the Structure of Aqueous Micellar Solutions of Sodium Oleate According to Small-Angle Neutron Scattering, J. Surf. Investig. X-Ray Synchrotron Neutron Tech. 12 (2018) 1142–1148. https://doi.org/10.1134/S1027451018050543.

[29] D. Mijaljica, F. Spada, I.P. Harrison, Skin Cleansing without or with Compromise: Soaps and Syndets, Molecules. 27 (2022) 2010. https://doi.org/10.3390/molecules27062010.

[30] H. Bunjes, M.H.J. Koch, K. Westesen, Influence of emulsifiers on the crystallization of solid lipid nanoparticles, J. Pharm. Sci. 92 (2003) 1509–1520. https://doi.org/10.1002/jps.10413.

[31] M. Ratoi, H.A. Spikes, Lubricating Properties of Aqueous Surfactant Solutions, Tribol. Trans. 42 (1999) 479–486. https://doi.org/10.1080/10402009908982244.

[32] X.P. Li, L. Yu, Y.Q. Ji, B. Wu, G.Z. Li, L.Q. Zheng, New type flooding systems in enhanced oil recovery, Chin. Chem. Lett. 20 (2009) 1251–1254. https://doi.org/10.1016/j.cclet.2009.05.007.

[33] G.C. Kalur, S.R. Raghavan, Anionic Wormlike Micellar Fluids that Display Cloud Points: Rheology and Phase Behavior, J. Phys. Chem. B. 109 (2005) 8599–8604. https://doi.org/10.1021/jp044102d.

[34] F.E. Antunes, L. Coppola, D. Gaudio, I. Nicotera, C. Oliviero, Shear rheology and phase behaviour of sodium oleate/water mixtures, Colloids Surf. Phys-

icochem. Eng. Asp. 297 (2007) 95–104. https://doi.org/10.1016/j.colsurfa.2006.10.030.

[35] Q. Cao, L. Yu, L.-Q. Zheng, G.-Z. Li, Y.-H. Ding, J.-H. Xiao, Rheological properties of wormlike micelles in sodium oleate solution induced by sodium ion, Colloids Surf. Physicochem. Eng. Asp. 312 (2008) 32–38. https://doi.org/10.1016/j.colsurfa.2007.06.024.

[36] P. Koshy, V.K. Aswal, M. Venkatesh, P.A. Hassan, Unusual Scaling in the Rheology of Branched Wormlike Micelles Formed by Cetyltrimethylammonium Bromide and Sodium Oleate, J. Phys. Chem. B. 115 (2011) 10817–10825. https://doi.org/10.1021/jp2053986.

[37] C. Flood, C.A. Dreiss, V. Croce, T. Cosgrove, G. Karlsson, Wormlike Micelles Mediated by Polyelectrolyte, Langmuir. 21 (2005) 7646–7652. https://doi.org/10.1021/la050326r.

[38] S. Khandavalli, J. Hendricks, C. Clasen, J.P. Rothstein, A comparison of linear and branched wormlike micelles using large amplitude oscillatory shear and orthogonal superposition rheology, J. Rheol. 60 (2016) 1331–1346. https://doi.org/10.1122/1.4965435.

[39] N.V. Thampi, R.P. John, K. Ojha, U.G. Nair, Effect of Salts, Alkali, and Temperature on the Properties of Sodium Oleate Hydrogel, Ind. Eng. Chem. Res. 55 (2016) 5805–5816. https://doi.org/10.1021/acs.iecr.6b00158.

[40] J. Huang, S. Zhang, Y. Feng, J. Li, H. Yan, F. He, G. Wang, Y. Liu, L. Wang, Rheological properties and application of wormlike micelles formed by sodium oleate/benzyltrimethyl ammonium bromide, Colloids Surf. Physicochem. Eng. Asp. 500 (2016) 222–229. https://doi.org/10.1016/j.colsurfa.2016.03.074.

[41] H. Lu, Q. Shi, Z. Huang, pH-Responsive Anionic Wormlike Micelle Based on Sodium Oleate Induced by NaCl, J. Phys. Chem. B. 118 (2014) 12511–12517. https://doi.org/10.1021/jp506809m.

[42] J. König, 2 - Food colour additives of synthetic origin, in: M.J. Scotter (Ed.), Colour Addit. Foods Beverages, Woodhead Publishing, Oxford, 2015: pp. 35–60. https://doi.org/10.1016/B978-1-78242-011-8.00002-7.

[43] H.M.D. Bandara, S.C. Burdette, Photoisomerization in different classes of azobenzene, Chem. Soc. Rev. 41 (2012) 1809–1825. https://doi.org/10.1039/C1CS15179G.

[44] R. Dąbrowski, P. Kula, J. Herman, High Birefringence Liquid Crystals, Crystals. 3 (2013) 443–482. https://doi.org/10.3390/cryst3030443.

[45] D. Statman, I. Jánossy, Study of photoisomerization of azo dyes in liquid crystals, J. Chem. Phys. 118 (2003) 3222–3232. https://doi.org/10.1063/1.1538598.

[46] H. Chen, W. Chen, Y. Lin, Y. Xie, S.H. Liu, J. Yin, Visible and near-infrared light activated azo dyes, Chin. Chem. Lett. 32 (2021) 2359–2368. https://doi.org/10.1016/j.cclet.2021.03.020.

[47] H. Rau, Chapter 4 - Azo Compounds, in: H. Dürr, H. Bouas-Laurent (Eds.), Photochromism, Elsevier Science, Amsterdam, 2003: pp. 165–192. https://doi.org/10.1016/B978-044451322-9/50008-7.

[48] H. Oh, A.M. Ketner, R. Heymann, E. Kesselman, D. Danino, D.E. Falvey, S.R. Raghavan, A simple route to fluids with photo-switchable viscosities based on a reversible transition between vesicles and wormlike micelles, Soft Matter. 9 (2013) 5025–5033. https://doi.org/10.1039/C3SM00070B.

[49] Y. Lin, X. Cheng, Y. Qiao, C. Yu, Z. Li, Y. Yan, J. Huang, Creation of photo-modulated multi-state and multi-scale molecular assemblies via binary-state molecular switch, Soft Matter. 6 (2010) 902–908. https://doi.org/10.1039/B916721H.

[50] Y. Lu, T. Zhou, Q. Fan, J. Dong, X. Li, Light-responsive viscoelastic fluids based on anionic wormlike micelles, J. Colloid Interface Sci. 412 (2013) 107–111. https://doi.org/10.1016/j.jcis.2013.09.014.

[51] Y. Bi, H. Wei, Q. Hu, W. Xu, Y. Gong, L. Yu, Wormlike Micelles with Photoresponsive Viscoelastic Behavior Formed by Surface Active Ionic Liquid/Azobenzene Derivative Mixed Solution, Langmuir. 31 (2015) 3789–3798. https://doi.org/10.1021/acs.langmuir.5b00107.

[52] J. Jiang, G. Wang, Y. Ma, Z. Cui, B.P. Binks, Smart worm-like micelles responsive to CO2/N2 and light dual stimuli, Soft Matter. 13 (2017) 2727–2732. https://doi.org/10.1039/C7SM00185A.

[53] M. Salzano de Luna, V. Marturano, M. Manganelli, C. Santillo, V. Ambrogi, G. Filippone, P. Cerruti, Light-responsive and self-healing behavior of azobenzene-based supramolecular hydrogels, J. Colloid Interface Sci. 568 (2020) 16–24. https://doi.org/10.1016/j.jcis.2020.02.038.

[54] T. Liu, Q. Lei, J. Dong, X. Li, Y. Lu, T. Zhou, Light-responsive vesicles based on azobenzene containing imidazolium surfactants and sodium oleate, Colloid Polym. Sci. 297 (2019) 1489–1497. https://doi.org/10.1007/s00396-019-04557-8.

[55] D. Wang, R. Dong, P. Long, J. Hao, Photo-induced phase transition from multilamellar vesicles to wormlike micelles, Soft Matter. 7 (2011) 10713–10719. https://doi.org/10.1039/C1SM05949A.

[56] R. Kumar, S.R. Raghavan, Photogelling fluids based on light-activated growth of zwitterionic wormlike micelles, Soft Matter. 5 (2009) 797–803. https://doi.org/10.1039/B809452G.

[57] J. Li, M. Zhao, H. Zhou, H. Gao, L. Zheng, Photo-induced transformation of wormlike micelles to spherical micelles in aqueous solution, Soft Matter. 8 (2012) 7858–7864. https://doi.org/10.1039/C2SM25218J.

[58] H. Sakai, S. Taki, K. Tsuchiya, A. Matsumura, K. Sakai, M. Abe, Photochemical Control of Viscosity Using Sodium Cinnamate as a Photoswitchable Molecule, Chem. Lett. 41 (2012) 247–248. https://doi.org/10.1246/cl.2012.247.

[59] M. Zhao, M. Gao, C. Dai, C. Zou, Z. Yang, X. Wu, Y. Liu, Y. Wu, S. Fang, W. Lv, Investigation of Novel Triple-Responsive Wormlike Micelles, Langmuir. 33 (2017) 4319–4327. https://doi.org/10.1021/acs.langmuir.7b01011.

[60] E. Nabi, M. Drechsler, M. Gradzielski, Phase behaviour and vesicle formation in catanionic mixtures of Na oleate and alkyl trimethyl ammonium bromide and its salt-free version, Colloid Polym. Sci. 293 (2015) 3119–3130. https://doi.org/10.1007/s00396-015-3737-y.

[61] D. Tatini, F. Sarri, P. Maltoni, M. Ambrosi, E. Carretti, B.W. Ninham, P. Lo Nostro, Specific ion effects in polysaccharide dispersions, Carbohydr. Polym. 173 (2017) 344–352. https://doi.org/10.1016/j.carbpol.2017.05.078.

[62] D. Tatini, P. Tempesti, F. Ridi, E. Fratini, M. Bonini, P. Baglioni, Pluronic/gelatin composites for controlled release of actives, Colloids Surf. B Biointerfaces. 135 (2015) 400–407. https://doi.org/10.1016/j.colsurfb.2015.08.002.

[63] B.A. Schubert, E.W. Kaler, N.J. Wagner, The Microstructure and Rheology of Mixed Cationic/Anionic Wormlike Micelles, Langmuir. 19 (2003) 4079–4089. https://doi.org/10.1021/la020821c.

[64] M.M. Cross, Rheology of non-Newtonian fluids: A new flow equation for pseudoplastic systems, J. Colloid Sci. 20 (1965) 417–437. https://doi.org/10.1016/0095-8522(65)90022-X.

[65] Toshiyuki. Shikata, Hirotaka. Hirata, Tadao. Kotaka, Micelle formation of detergent molecules in aqueous media: viscoelastic properties of aqueous cetyltrimethylammonium bromide solutions, Langmuir. 3 (1987) 1081–1086. https://doi.org/10.1021/la00078a035.

[66] T. Shikata, H. Hirata, T. Kotaka, Micelle formation of detergent molecules in aqueous media. 2. Role of free salicylate ions on viscoelastic properties of aqueous cetyltrimethylammonium bromide-sodium salicylate solutions, Langmuir. 4 (1988) 354–359. https://doi.org/10.1021/la00080a019.

[67] C. L Barrie, P.C. Griffiths, R.J. Abbott, I. Grillo, E. Kudryashov, C. Smyth, Rheology of aqueous carbon black dispersions, J. Colloid Interface Sci. 272 (2004) 210–217. https://doi.org/10.1016/j.jcis.2003.12.004.

[68] T. Sochi, Non-Newtonian flow in porous media, Polymer. 51 (2010) 5007–5023. https://doi.org/10.1016/j.polymer.2010.07.047.

[69] G. Polacco, S. Filippi, F. Merusi, G. Stastna, A review of the fundamentals of polymer-modified asphalts: Asphalt/polymer interactions and principles of compatibility, Adv. Colloid Interface Sci. 224 (2015) 72–112. https://doi.org/10.1016/j.cis.2015.07.010.

[70] M.F. Nazar, S. Murtaza, B. Ijaz, M. Asfaq, M.A. Mohsin, Photophysical Investigations of Carmoisine Interacting with Conventional Cationic Surfactants Under Different pH Conditions, J. Dispers. Sci. Technol. 36 (2015) 18–27. https://doi.org/10.1080/01932691.2014.884465.

[71] A. Basu, G. Suresh Kumar, Minor Groove Binding of the Food Colorant Carmoisine to DNA: Spectroscopic and Calorimetric Characterization Studies, J. Agric. Food Chem. 62 (2014) 317–326. https://doi.org/10.1021/jf404960n.

[72] S.J. Candau, E. Hirsch, R. Zana, Light scattering investigations of the behavior of semidilute aqueous micellar solutions of cetyltrimethylammonium bromide: Analogy with semidilute polymer solutions, J. Colloid Interface Sci. 105 (1985) 521–528. https://doi.org/10.1016/0021-9797(85)90327-3.

[73] L.J. Magid, Z. Han, Z. Li, P.D. Butler, Tuning Microstructure of Cationic Micelles on Multiple Length Scales: The Role of Electrostatics and Specific Ion Binding, Langmuir. 16 (2000) 149–156. https://doi.org/10.1021/la990686c.

[74] R.D. Koehler, S.R. Raghavan, E.W. Kaler, Microstructure and Dynamics of Wormlike Micellar Solutions Formed by Mixing Cationic and Anionic Surfactants, J. Phys. Chem. B. 104 (2000) 11035–11044. https://doi.org/10.1021/jp0018899.

[75] H. Hoffmann, Viscoelastic Surfactant Solutions, in: Struct. Flow Surfactant Solut., American Chemical Society, 1994: pp. 2–31. https://doi.org/10.1021/bk-1994-0578.ch001.

[76] M. In, G.G. Warr, R. Zana, Dynamics of Branched Threadlike Micelles, Phys. Rev. Lett. 83 (1999) 2278–2281. https://doi.org/10.1103/PhysRevLett.83.2278.

[77] A. Parker, W. Fieber, Viscoelasticity of anionic wormlike micelles: effects of ionic strength and small hydrophobic molecules, Soft Matter. 9 (2012) 1203–1213. https://doi.org/10.1039/C2SM27078A.

[78] C. Rodriguez-Abreu, K. Aramaki, Y. Tanaka, M.A. Lopez-Quintela, M. Ishitobi, H. Kunieda, Wormlike micelles and microemulsions in aqueous mixtures of sucrose esters and nonionic cosurfactants., J. Col-

loid Interface Sci. 291 (2005) 560–569. https://doi.org/10.1016/j.jcis.2005.05.018.

[79] N. Naito, D.P. Acharya, J. Tanimura, H. Kunieda, Rheological Behavior of Wormlike Micellar Solutions in Mixed Nonionic Systems of Polyoxyethylene Phytosterol-Polyoxyethylene Dodecyl Ether, J. Oleo Sci. 53 (2004) 599–606. https://doi.org/10.5650/jos.53.599.

[80] R.G. Shrestha, C. Rodriguez-Abreu, K. Aramaki, Wormlike Micelles in Mixed Amino Acid Surfactant/Nonionic Surfactant Aqueous Systems and the Effect of Added Electrolytes, J. Oleo Sci. 58 (2009) 243–254. https://doi.org/10.5650/jos.58.243.

[81] H. Rehage, H. Hoffmann, Viscoelastic surfactant solutions: model systems for rheological research, Mol. Phys. 74 (1991) 933–973. https://doi.org/10.1080/00268979100102721.

[82] S.J. Candau, R. Oda, Linear viscoelasticity of salt-free wormlike micellar solutions, Colloids Surf. Physicochem. Eng. Asp. 183–185 (2001) 5–14. https://doi.org/10.1016/S0927-7757(01)00535-0.

[83] S.A. Rogers, M.A. Calabrese, N.J. Wagner, Rheology of branched wormlike micelles, Curr. Opin. Colloid Interface Sci. 19 (2014) 530–535. https://doi.org/10.1016/j.cocis.2014.10.006.

[84] M. Buchanan, M. Atakhorrami, J.F. Palierne, F.C. MacKintosh, C.F. Schmidt, High-frequency microrheology of wormlike micelles, Phys. Rev. E. 72 (2005) 011504. https://doi.org/10.1103/PhysRevE.72.011504.

[85] J. Honerkamp, J. Weese, A nonlinear regularization method for the calculation of relaxation spectra, Rheol. Acta. 32 (1993) 65–73. https://doi.org/10.1007/BF00396678.

[86] E. Carretti, C. Matarrese, E. Fratini, P. Baglioni, L. Dei, Physicochemical characterization of partially hydrolyzed poly(vinyl acetate)–borate aqueous dispersions, Soft Matter. 10 (2014) 4443–4450. https://doi.org/10.1039/C4SM00355A.

[87] N. Garti, A. Aserin, S. Ezrahi, I. Tiunova, G. Berkovic, Water Behavior in Nonionic Surfactant Systems I: Subzero Temperature Behavior of Water in Nonionic Microemulsions Studied by DSC, J. Colloid Interface Sci. 178 (1996) 60–68. https://doi.org/10.1006/jcis.1996.0093.

[88] N. Garti, Thermal Behavior of Dispersed Systems, CRC Press, 2000.

[89] P.C. Schulz, DSC analysis of the state of water in surfactant-based microstructures, J. Therm. Anal. Calorim. 51 (1998) 135–149. https://doi.org/10.1007/BF02719017.

[90] M. Fernandez-Tarrio, C. Alvarez-Lorenzo, A. Concheiro, Calorimetric approach to tetronic/water interactions, J. Therm. Anal. Calorim. 87 (2007) 171–178. https://doi.org/10.1007/s10973-006-7834-5.

[91] A. Yaghmur, A. Aserin, I. Tiunova, N. Garti, Subzero temperature behaviour of non-ionic microemulsions in the presence of propylene glycol by DSC, J. Therm. Anal. Calorim. 69 (2002) 163–177. https://doi.org/10.1023/A:1019997909346.

[92] D. Senatra, G. Gabrielli, G.G.T. Guarini, Experimental Evidence of Interphasal Water in Water-in-Oil Microemulsions, EPL Europhys. Lett. 2 (1986) 455. https://doi.org/10.1209/0295-5075/2/6/008.

[93] Chaplin, Martin, Polysaccharide hydration, (n.d.). http://www1.lsbu.ac.uk/water/polysaccharide_hydration.html (accessed September 27, 2016).

[94] L. Guan, H. Xu, D. Huang, hatak, J. Polym. Res. 18 (2010) 681–689. https://doi.org/10.1007/s10965-010-9464-7.

[95] T. Hatakeyama, M. Tanaka, H. Hatakeyama, Thermal Properties of Freezing Bound Water Restrained by Polysaccharides, J. Biomater. Sci. Polym. Ed. 21 (2010) 1865–1875. https://doi.org/10.1163/092050610X486946.

[96] L. Zhao, L. Pan, A. Ji, Z. Cao, Q. Wang, Recrystallization of freezable bound water in aqueous solutions of medium concentration, Chin. Phys. B. 25 (2016) 075101. https://doi.org/10.1088/1674-1056/25/7/075101.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

OPEN ACCESS

Citation: Uskoković V. (2022) Indigenous Sand Drawings as Predictors of the Cell Response to Nanoparticle Therapy. *Substantia* 6(2): 27-42. doi: 10.36253/Substantia-1586

Received: Feb 28, 2022

Revised: Jun 17, 2022

Just Accepted Online: Jun 18, 2022

Published: September 1, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Indigenous Sand Drawings as Predictors of the Cell Response to Nanoparticle Therapy

Vuk Uskoković[1,2,*]

[1] *Advanced Materials and Nanobiotechnology Laboratory, TardigradeNano, 7 Park Vista, Irvine, CA 92604, USA*

[2] *Department of Mechanical Engineering, San Diego State University, 5500 Campanile Dr., San Diego, CA 92182, USA*

*E-mail: vuk.uskokovic@tardigradenano.com; vuskokovic@sdsu.edu; vuk21@yahoo.com.

Abstract. A technique for predicting the response of cells and tissues to a physicochemical stimulus without the use of expensive molecular markers and at time points before any morphological changes can be visibly spotted would be a meaningful addendum to the current set of bioimaging tools. One such method was developed here based on correlating transformed distance matrices of populations of cultured cells and digital checkerboard patterns derived from traditional central African drawings in the sand. Similarity measurements were made at an early time point in the therapy administered to bone cancer cells in the form of composite magnetic nanoparticles. At this early of a time point, the cell viability was mildly reduced, but no gross alterations to the cell morphology or density were visible yet. Similarity score evaluation demonstrated a significantly higher degree of similarity between the patterns derived from the sand drawings and the cells subjected to the treatment than between the former patterns and the untreated cell controls. The treated cells produced more ordered and symmetric patterns than the control ones after the processing of their pairwise distant matrices, explaining their better geometric correlation with the ancestral sand arabesques, which were monolinear and commonly comprised multiple mirror planes and rotational symmetry axes. This has suggested that the course of the therapy could be predicted by a relatively simple comparison between raw optical images of cells and indigenous ideographs using the metrics postulated here. The interdisciplinary method developed in this study may prove applicable for *in situ* monitoring of the response of cells and tissues to various therapies, allowing for the early indications of adverse effects to be noticed based on the simple optical observations of cells and acted upon before the progression toward nonviable states becomes irreversible. The method elaborated here may also provide an impetus for a broader search for solutions to problems plaguing the modern medicine outside of the scope of its mainstream analytical frameworks and in the ancestral heritage of relatively obscure ethnic traditions.

Keywords: Anthropology; Bioimaging; Ethnoscience; Lusona; Nanomedicine; Pattern recognition; Sona; Tchokwe.

1. INTRODUCTION

Some time in the early 1930s, a Canadian neoclassical composer and ethnomusicologist, Colin McPhee visited an island in the Pacific called Bali[1]. During the couple of months he spent there during his first visit, he got enchanted by the traditional gamelan music of the indigenous islanders to such an extent that he returned to the West with the decision to transform the classical music with the percussive style and unconventional tunings and scales of the music of the islanders. Although McPhee's fusion music would be largely neglected and unrecognized by the mainstream and the classical music scene of his times, it would end up setting the scene for the birth of rhythmically similar avant-garde concepts by the likes of John Cage, Lou Harrison, Benjamin Britten, Philip Glass and Steve Reich, handing us a powerful example of how delving into the deep past and the places inhabited by the poor, the unheeded and the underprivileged could pave way toward the most advanced sci-fi futures imaginable.

This story is one of many that have served as an inspiration for this author's diligent effort to provide an impetus for rejuvenation of today's scientific practice in the west, which continues to hold an elitist, intrinsically arrogant, neocolonial stance with respect to the scientific methods and sources of knowledge emerging from older, less technologically developed cultures, looking down on them from unjustifiable heights. More than three centuries have passed since the Newtonian astronomers proved that the ecclesiastical idea that the Sun is moving while the Earth is still and the Galilean view of the Sun at rest and the moving Earth must be merged to get a correct picture[2]; a century has passed since the demonstration of the particle/wave dualism in quantum mechanics and, consequently, of the correctness of both Newtonian and Huygens' models of electromagnetic radiation; more than eighty years have passed since the valence bond and the molecular orbital theories explaining chemical bonding were proven equivalent[3]; more than fifty years have passed since the topological geometry unified the Euclidean, metric geometry and analytical, projective geometry[4], and yet the openness of the western science to alternative models of physical reality and tools for assessing it appears never to have been lesser than today. Rare exceptions aside, the uniformities of methods, models, lifestyles and worldviews have taken over the world, which increasingly closes itself to the opportunities of their fundamentally novel analogues, notwithstanding that it needs them for its own good more than ever. As a consequence, curricula in natural sciences in virtually every university across the globe have been fundamentally the same[5], reflecting a false impression that scientific thinking evolved outside of cultural boundaries[6]. As these boundaries are erased, individual cultures get trapped inside the dogma that a single model of development suits them all[7], creating a potentially dangerous state where a leader of the cultures in development, if it only turns out to be led astray, could drag all of its myrmidons down a cliff. In contrast, a healthier and more sustainable model for the global development would be rooted in the fosterage of a diversity of approaches and trajectories of growth[8].

The ethos of the story about McPhee and the gamelan music instructs us through an analogy to seek local and largely forgotten scientific practices on the map of the world and translate their premises and methods to the contemporary scientific settings in the West. One such approach goes a step ahead of conventional ethnoscience, which busies itself with rediscovering the scientific methods native to specific cultures, and tries instead to cross-fertilize the scientific language and methods employed in the West with the ethnic ones confined to either history books or cultures that often find themselves on the brink of extinction. In doing so, it is deemed that we would not only help revitalize the western science, but also preserve the cultural and epistemological diversities of humanity, alongside unraveling the often tangled yarn of politics and spotlighting the issues of globalism, economic neoliberalism, neocolonialism and neo-imperialism, which affect every aspect of our existence ever so unnoticeably and ever so profoundly[9]. For example, D'Ambrosio's seminal work on ethnomathematics was inextricably tied to the issues of colonialism and so did Bishop's studies on language and education in Papua New Guinea lead to the insight about "the cultural invasion in colonized countries by western mathematics"[10]; similarly, delving into the specifics of any ethnoscience increases the awareness of the colonial history or ethnic struggles behind it, which sheds an important historical and political connotation to science *per se*. This lesson in history has been symptomatically skipped by the educators and practitioners of the western science, which is unsurprising given how easily such sociological determinants of the evolution of science can be discarded as inconsequential when we stand atop an edifice built out of piles of paradigms growing so tall that the foundations have become buried deep out of sight.

At this point in the introduction, gears are suddenly shifted and the cell biology problematics tackled in this study briefly described, as follows. Rare exceptions aside, all physiologically normal eukaryotic cells require a contact with adjacent cells to properly proliferate and engage in normal metabolic activities. Consequently, an increasing distance between cells populating a surface, such as that of a cell culture plate, normally signals an abnormality and an adverse response to the treatment, if any. However, this relationship between the cell distance and their viability is rather simplistic and applicable only in late stages of the treatment, when the adverse effects have become largely irreversible. More holistic attempts to correlate the broader distribution parameters of cell aggregates with their long-term viability have rarely been made. However, the benefits of a model capable of predicting the cell fate based on cell distance distribution in the early stages of the treatment, when average gross distances between the individual cells have not increased yet, are many and include the ability to recognize any adverse signs in the given cells during their monitoring *in situ*, which would increase the overall positive effects of the therapy. This study is underlain by the premise that the spatial ordering of the cells in culture, describable solely by the translational symmetry parameters such as pairwise distances, can be a marker for their future states in the course of the treatment. In search of this pairwise distance distribution model, an ethnocentric route is being followed, taking the searcher away from the western mathematics or molecular biology and into the heart of Africa, to drawings in the sand with which the local indigenes have engaged in storytelling.

2. EXPERIMENTAL PART

2.1 Cell culture

K7M2-pCl murine osteosarcoma cell line was purchased from the American Type Culture Collection (Manassas, VA, USA). Cells were grown to confluency before being plated on 12 mm circular glass cover slips or in 48-well culture plates. Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% antibiotic-antimycotic (Life Technologies, Carlsbad, CA). The medium was replaced every 48 h, and the cultures were incubated at 37 °C in a humidified atmosphere containing 5% CO_2. Upon confluency, the cells were detached from the cell culture flask surface using 0.25 wt.% trypsin, washed, centrifuged (1000 rpm x 3 min), resuspended in fresh media and subcultured. The cultures were regularly examined under an optical microscope to check for their growth and possible contamination.

2.2 Nanoparticle synthesis and characterization

Composite nanoparticles consisted of iron oxide cores, silica shells and carbon crusts and were modeled after the stratified structure of the planet Earth[11]. Their synthesis using a hydrothermal method and characterization using a variety of physicochemical techniques are described in more detail elsewhere[12]. Briefly, iron oxide nanoparticles were precipitated from an aqueous solution containing 10 mM $FeCl_3$, 5 mM $FeCl_2$ and 0.1 vol.% Triton X-100 by reacting it with a mixture of ammonia and NaOH. The resulting dispersion of the nanoparticles continued to be stirred and aged at 80 °C for 1 hour. A 1:1 mixture of tetraethylorthosilicate and (3-aminopropyl)triethoxysilane was then added to the suspension to deposit the silica layer. Carbon coating was deposited in a hydrothermal reactor (Parr), using citric acid as the carbon precursor. The reaction was run at 200 °C for 1 h. The suspension was concentrated via centrifugation in Amicon Ultra-4 centrifugal filter tubes (Ultracel 100-K, 100,000 M_w) to yield stable ferrofluids. High resolution transmission electron microscopy (HR-TEM) analysis was carried on a JEOL 2100F microscope equipped with Schottky type field emission source and the cryo-polepiece operating at 200 keV. All images were recorded using Gatan OneView camera with point-to-point resolution of 0.26 nm, lattice resolution of 0.1 nm, and information limit of 0.124 nm.

2.3 Cell viability assay

Near confluence, K7M2-pCl cells were divided to two groups: the control one and the to-be-treated one. The cells in the treatment group were treated with 5 mg/ml of composite magnetic nanoparticles dispersed in the cell culture medium and incubated at 37 °C and 5 % CO_2 for 24 hours. After the given incubation time, the cell viability was measured using the 3-[4, 5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide (MTT) assay. A 12 mM MTT stock solution was first prepared by adding 1 ml of sterile phosphate buffered saline (PBS) to a 5 mg vial of MTT and vortex-mixing to ensure complete dissolution. Cells were washed with PBS and 275 μl of 1:10 vol./vol. MTT/media were added into each well. After 4 h of incubation at 37 °C, 211 μl of the solution was removed and 125 μl of dimethyl sulfoxide was added to each well. The plates were placed in a 37 °C incubator shaker at 120 rpm for 30 min before measuring the absorbance at 570 nm using the BMG Labtech Fluostar Omega microplate spectrophotometer. Viability was expressed in percentages and normalized to the absorbance of the negative control, containing the growth medium alone. To account for the effect of nanoparticles *per se* on the absorbance in the lysate, the absorbance of the wells containing the growth medium and the nanoparticles, but no cells, was subtracted from that of the cells treated with the nanoparticles.

2.4. Fluorescent cell staining

Control and nanoparticle-treated cell cultures on glass cover slips were fixed with 4% paraformaldehyde, then washed with the wash buffer (0.1 wt.% Triton X and 0.1 wt.% bovine serum albumin in PBS), and stained with 0.2 ml per well of the staining solution containing 1:4000 v/v AlexaFluor 568 Phalloidin as an actin-staining reagent and one drop of NucBlue ReadyProbes as a cell nucleus-staining reagent. Fixed cells were incubated at room temperature for 2 hours before being mounted and cured with ProLong Diamond antifade mounting agent. Fluorescent cell images were acquired on a Nikon T1-S/L100 confocal optical microscope and analyzed for cell morphology and pairwise distances between cells using ImageJ (NIH, Bethesda, MD).

3. RESULTS

Nanoparticles that the cells were challenged with are displayed in Fig.1a, while their size distribution histogram is shown in Fig.1b. The particles were cuboid in shape and very narrowly dispersed, ranging from 8 to 16 nm in size, with the distribution peaking at 12 – 13 nm and the average size equaling 12.0 ± 2.3 nm. In spite of the absence of any amphiphilic stabilizers as well as the high surface-to-volume ratio of the nanoparticles and the magnetic attraction between their ferromagnetic cores, the nanoparticles were stable in suspension owing to the electrostatic repulsion provided by the monoatomically thin carbon coating (Fig.1a). Monodisperse in size and highly stable in suspension, the nanoparticles in this form ensured the reproducibility of the therapy and the evenness of the effect exhibited on cells in the culture. Earlier studies demonstrated that these nanoparticles exhibit moderate anticancer effects against various cancer cell lines[13]. On one hand, most nanoparticles exhibit slower modes of action than small molecules, in part because of their lower pharmacodynamic specificity and in part because of their lower diffusion coefficients. On the other hand, this slower mode of action allows for more facile *in situ* monitoring of the effects that the therapy has on the cells. It enables the segmentation of the therapy into narrow timeframes so as to capture the fine changes occurring in the cells as the result of the treatment. One such interruption of the treatment at its early stages was accomplished in this study with the goal of capturing the earliest sign that the cells have entered the trajectory leading from viability to obliteration.

Consequently, fluorescent optical images of the cells were acquired at a very specific time point, at which a partial loss of viability in the treatment sample group has started to occur, as demonstrated in Fig.2a, but no

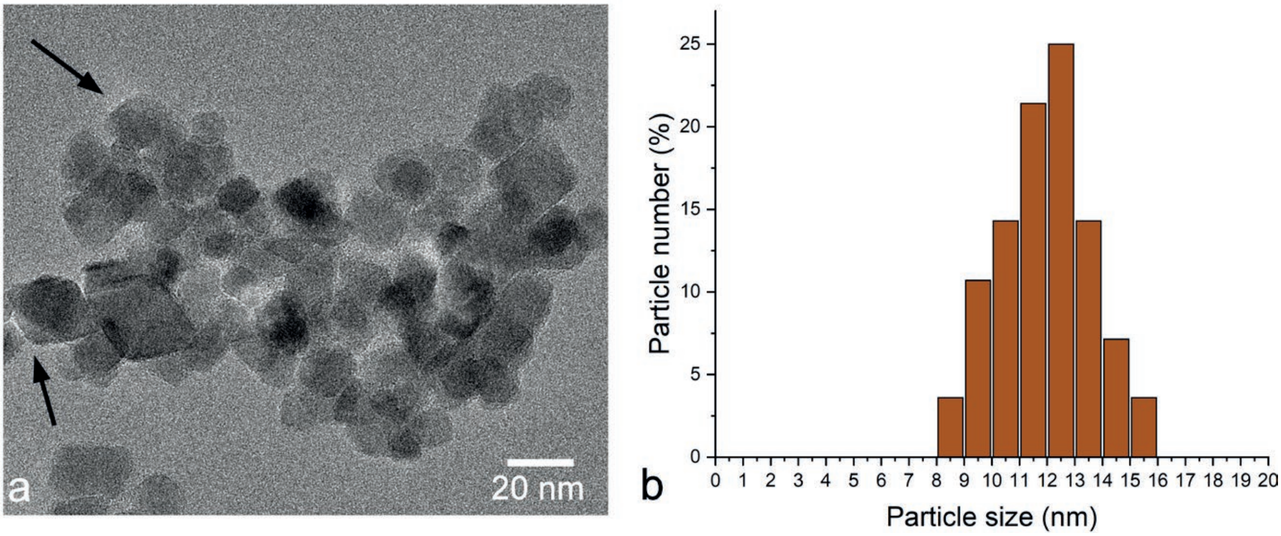


Figure 1. A representative transmission electron micrograph of the composite magnetic nanoparticles (a) and the corresponding particle size histogram (b). Arrows in (a) denote the bright regions on the particle surface originating from the carbon coating that protects the particles against segregation.

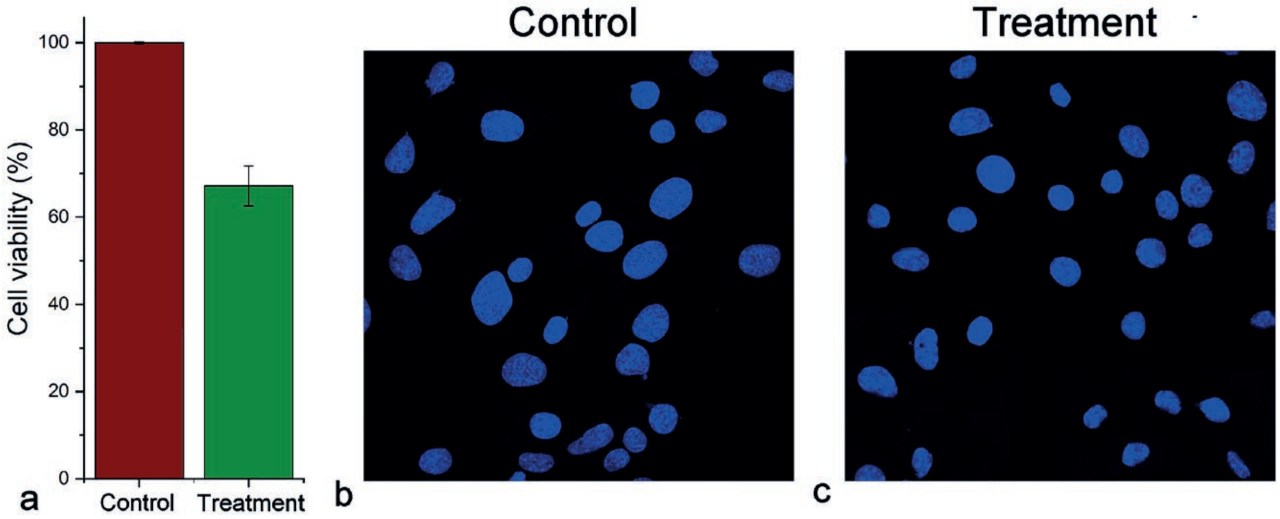
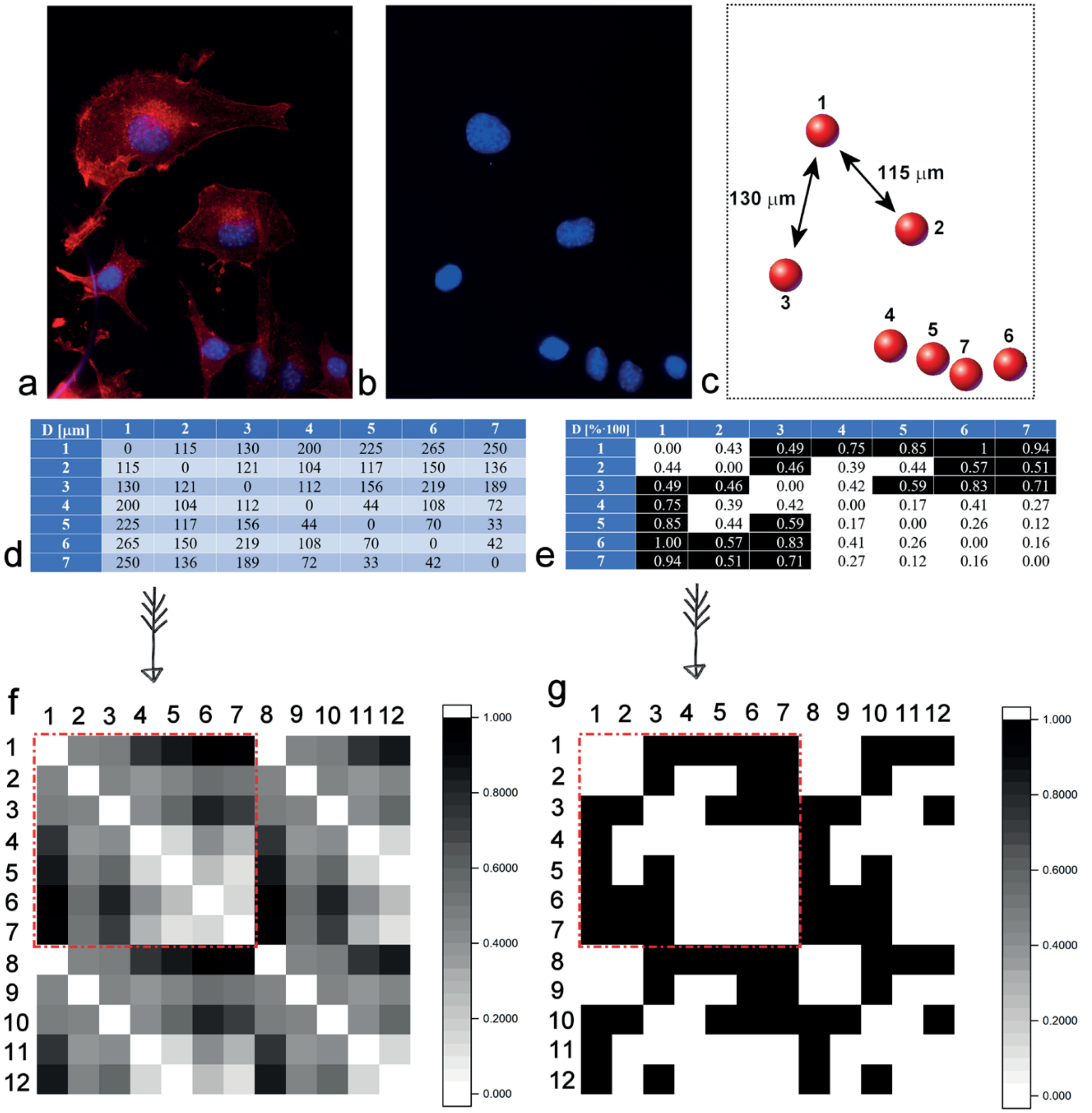


Figure 2. K7M2-pCl bone cancer cell viability (a) and fluorescent optical images (b-c) showing the K7M2-pCl cell nuclei stained in blue 24 h after the onset of the treatment for the control (a, b) and the treated (a, c) population. Dimensions of the images (b) and (c) are 270 x 270 μm.

detrimental effects were evident by simply observing the cells under the microscope. Hence, posed side by side, the control and the treated cells after the 24 h incubation with the magnetic nanoparticles looked indistinguishable despite the beginning of the loss of viability of the treated cells, as assessed by their levels of the mitochondrial activity[14] measured in an MTT test (Fig.2a). Correspondingly, the image of control cells shown in Fig.2b and the image of the treated cells shown in Fig.2c present the identical number of cells per image, in this case 30, and the only difference between them can lie in their spatial relations with the neighboring cells, which cannot be discerned with the naked eye and requires the transformation to an imaginary, computational space. The central biological question underlying this study was whether these imperceptible changes to the spatial arrangement, if any, might hide a key to predicting the cell fate at one such early time point in the treatment as that captured in Fig.2.

In response to the stimulus, which in this case comes from the addition of the nanoparticles to the culture at a considerable concentration of 5 mg/ml, the cells begin to migrate and rearrange. However, in spite of the typical migration rates of dozens of micrometers per hour[15], such rearrangements for most inorganic nanoparticle treatments produce changes that make it impossible to induce the occurrence of adversity at early time points. The hypothesis underlying this study was that with an appropriate computational model, subtle changes in the translational symmetry could be used to predict the cell fate before the latter becomes phenotypically evident. In real life, this would help in noticing the early disease progression during *in situ* monitoring of organs and tissues, at which point the pathological process is still reversible. An even bolder hypothesis posed atop of this hypothesis is that the key to producing computational methods for achieving this analytical prediction lies dormant in ideograms drawn with sticks or fingers in the sand as a part of indigenous storytelling. By descending deep into the heart of Africa and selecting ancestral figure drawings in the sand as sources of such methods, it was deemed that the latter must be original, so as to comply with the originality of the study, but also primitivistic, thereby evoking the ideal held onto by Paul Gauguin, who found his niche away from the western world and on the island of Tahiti, which was to "go back beyond the horses of the Parthenon to the rocking-horse of the childhood"[16].

This simplistic, manualized process of converting the raw cell images into digital checkerboard patterns and comparing them against similar patterns derived from the indigenous drawings proceeded as follows. In the first step, pairwise distances between cells in a single image were used to construct the corresponding distance matrices. To do so, as it is illustrated in Fig.3, the originally captured cells (Fig.3a) were stripped of anything but their nuclei (Fig.3b) and assigned numbers starting from the top left to the bottom right in the image, so that their numbering in the distance matrix corresponded approximately to their location in the culture dish from the viewer's perspective. The pairwise distances were subsequently measured in pixels or micrometers using ImageJ (Fig.3c) and used to create the distance matrices (Fig.3d), which were then converted



D [μm]	1	2	3	4	5	6	7
1	0	115	130	200	225	265	250
2	115	0	121	104	117	150	136
3	130	121	0	112	156	219	189
4	200	104	112	0	44	108	72
5	225	117	156	44	0	70	33
6	265	150	219	108	70	0	42
7	250	136	189	72	33	42	0

D [%·100]	1	2	3	4	5	6	7
1	0.00	0.43	0.49	0.75	0.85	1	0.94
2	0.44	0.00	0.46	0.39	0.44	0.57	0.51
3	0.49	0.46	0.00	0.42	0.59	0.83	0.71
4	0.75	0.39	0.42	0.00	0.17	0.41	0.27
5	0.85	0.44	0.59	0.17	0.00	0.26	0.12
6	1.00	0.57	0.83	0.41	0.26	0.00	0.16
7	0.94	0.51	0.71	0.27	0.12	0.16	0.00

Figure 3. A representative optical image of K7M2-pCl osteosarcoma cells fluorescently stained in red for f-actin and blue for the nucleus (a). The image showing only the positions of the cell nuclei (b) is converted to dots corresponding to centers of nuclei (c) and used to measure the pairwise distances between cells (c). These distances are used to plot an absolute distance matrix (d) and a relative one, where distances are normalized to the largest distance in the image (e). The former matrix can be converted to an absolute heat map (f). To digitalize one such heat map, distances ranging from 0 to the matrix average are assigned 0 value (white), while distances ranging from the matrix average to 100 % are assigned the value of 1 (black) in order to create a matrix (g) that is comparable for similarity with the sona drawing patterns based on the methodology described in Figs.4 and 5. The distance matrix heat maps were of the n x n dimensions, where n is the number of cells used to construct the matrix. Here, n = 7 and the basic 7 x 7 patterns obtained from the distance matrices are demarcated in (f) and (g) by the red dashed squares. Their extension resulted in 12 x 12 checkerboard patterns (f, g) comparable against the identically dimensioned sona drawing patterns.

to 2D heat maps (OriginPro 2016). Each square in these heat maps represented a value ranging from 0 to 1 after being normalized to the range between the zero distance and the greatest distance between two cells in the image (Fig.3f). To create digital checkerboard patterns, the value of 1 was assigned to each square in the heat map whose distance value was higher than the average for the given matrix and the value of 0 to each square whose distance value was lower than the average (Fig.3e, g). All patterns were plotted in 12 x 12 dimensions. Although this could be achieved by extending the pattern to the missing rows and columns when there are less than 12 cells in an image, as it is illustrated in Fig.3, all patterns used in the comparisons were created by using the first 12 cells from the top left to the bottom right in an image. Such patterns were then compared for similarity against similar checkerboard patterns obtained from hand-drawn contours in the sand corresponding to selected (lu)sona (sing. lusona, pl. sona) ideograms of Tchokwe, Lunda, Lwena, Xinge, Minungo, Ngangela and Luchazi indigenes from the territories of today's Angola, Congo and Zambia.

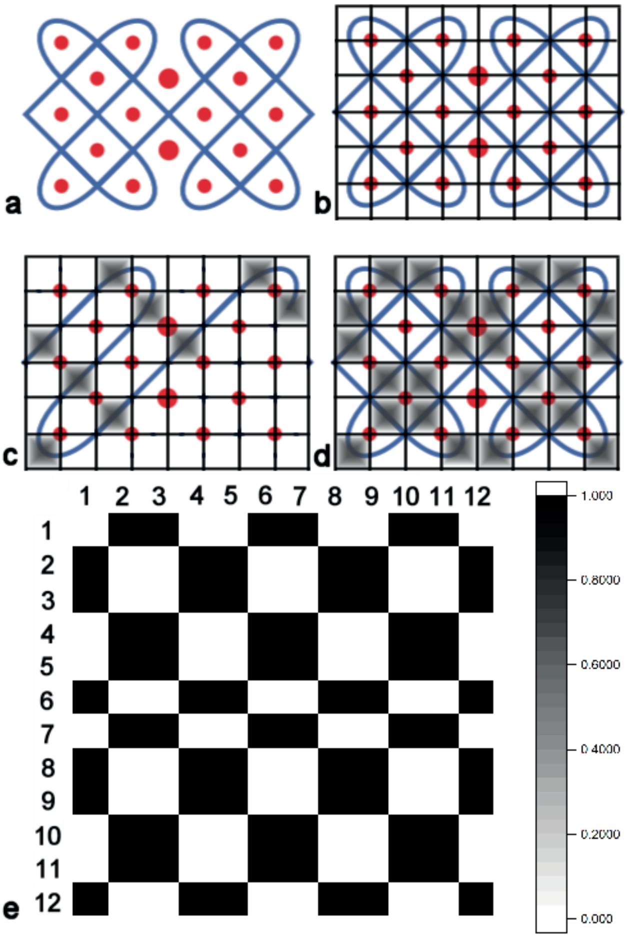


Figure 4. Steps involved in the conversion of a lusona drawing into a simplistic checkerboard pattern. The lusona in (a) represents a nest of doves. A grid defined by dots in the sand (red circles) is then formed over the drawing (b). The squares of the grid crossed during the drawing of the continuous curve are alternately marked as black or white (c), yielding a distinct black-and-white pattern (d), which is then redrawn without the underlying lusona and extended to the dimensions comparable with the cell distance matrix (in this case 12 x 12 starting from a 8 x 6 sona) (e).

Out of all these indigenous West Central Bantu cultures, the sand drawings of the Tchokwe remain the most historically preserved[17]. Their current forms trace a few hundred years back in time, not long before or after the Tchokwe broke away from the Lunda chiefdom of today's Congo. The sona arabesques represent the product and the means of the art of storytelling that combines verbal expression with drawing continuous lines in the sand around equidistant dots impressed in the sand by the index and the ring finger. Because of the vanishing of local traditions under the colonial rules, it is mostly older men and women that are proficient in drawing the sona figures in the sand. Although such drawings serve as a common source of storytelling to children, they have been traditionally used to amuse and educate adults during social gatherings. Individual lusona figure drawings are used to represent flora or fauna, manmade objects, natural or social events, or entire myths. Most, but not all, lusona ideographs are monolinear, that is, formed by drawing a single curved line around a series of dots impressed in the sand without touching them. All of the sona drawings used in this study were monolinear and they represented a range of objects and events, as reported by Gerdes[18], including a nest of doves, a coop for carrying birds, the belly of the lion, the chased chicken's path, the bridge, the antelope's paws, and others. Despite the large amount of research that went into collecting and preserving the Tchokwe sand drawings compared to many other similar storytelling and artistic traditions, there is still some ambiguity with respect to the annotation of meaning to specific ideograms. For example, the lusona representing fire is highly similar to the sona representing the end of a fight and a fishing trap; the sona illustrating the fight, a snake on a tree trunk, and a dead one in a coffin bear high levels of semblance[19]; the lusona taken here to represent tiptoeing in a forest has also had multiple meanings assigned to it, including that of a forest where a local, unidentified bird lives.

The method for transforming the drawings in the sand to digital checkboard patterns comparable with those formed from the cell micrographs was adapted from Gerdes[20]. It involved the alternate assignment of

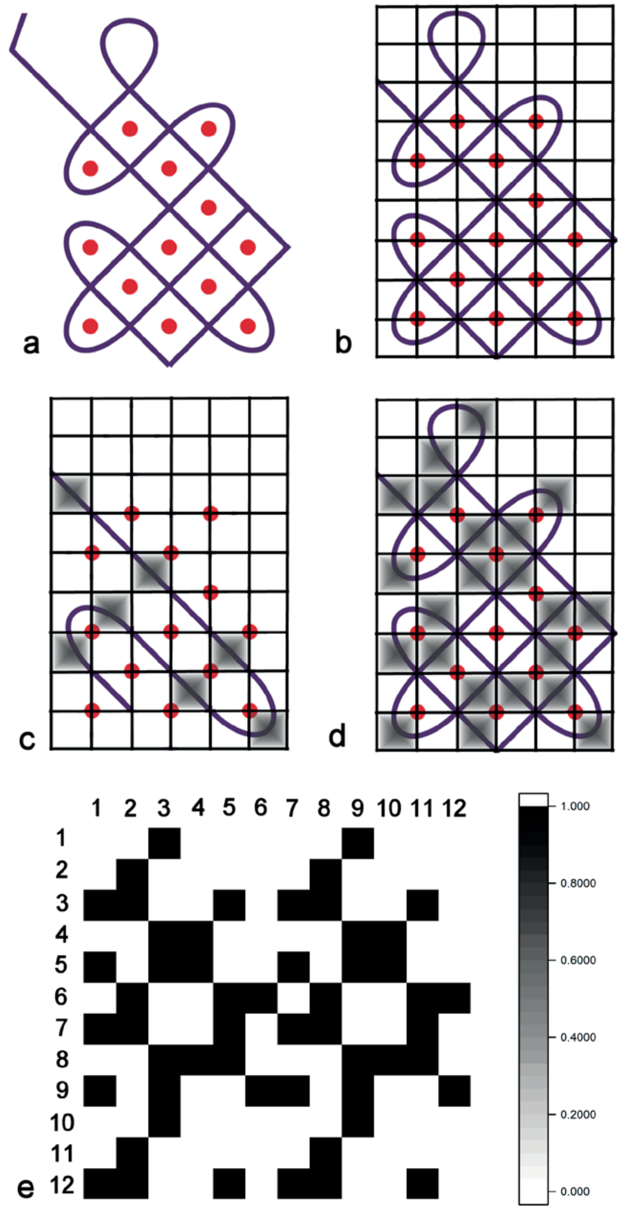


Figure 5. Steps involved in the conversion of a lusona drawing into a more intricate checkerboard pattern. The lusona in (a) represents a human. The same process as that applying to the pattern in Fig.4 is carried out (b-d) to create an extended pattern corresponding to the lusona drawing, with the dimensions comparable with the cell distance matrix (in this case 12 x 12 starting from a 6 x 9 sona) (e).

black and white colors to the squares of a grid formed around the nods of the drawing, as depicted in Figs.4 and 5. Here, Fig.4 represents the process of transforming a lusona drawing representing a nest of doves to a digitalized heat map, giving the simplest checkerboard pattern derivable from sona drawings. Identical patterns were obtained from numerous other sona drawings, including those representing the cricket's party and a basket for carrying birds. In contrast, Fig.5 represents the derivation of a more complex and unique, less translationally symmetric pattern from the lusona drawing representing a human being. In both cases, a grid is drawn using the equidistant dots in the sand as nodes, after which the curved line forming the drawing is traced and the squares of the grid are assigned alternate values of 1 and 0. The typically rectangular sona checkerboard patterns were converted to 12 x 12 squares not by multiplying the original matrices with their transposes, but by extending the patterns to the missing rows and columns or occasionally, as in the case of very large drawings, such as that representing finding oneself in the belly of a lion, cropped down to the 12 x 12 size. In any case, it was ensured that the dimensions of the cell distance maps and the sona patterns were both of the 12 x 12 size prior to the comparison. It was also ascertained that the comparison was performed for cells before the significant morphological deformations and loss of viability have occurred because only in such a way does the similarity, if observed, becomes a meaningful and reliable predictor of the treatment effect.

The comparison was based on the assignment of scores measuring similarity between the control cells and the cells challenged with the nanoparticles on one side and 10 selected sona ideograms (Fig.6) on the other. The ideograms chosen for similarity matching were selected from the aforementioned sona collections by Gerdes, mostly based on the aesthetics of their representation of a subject and without any prior idea of how well they would perform on the similarity test. This was done to ensure that the analysis proceeds in good faith and in a bias-free manner. Any of the countless computational algorithms, ranging from the codes used in facial feature[21] or fingerprint[22] recognition software to the Bhattacharyya similarity coefficient integral to programs for personalized product or service recommendation[23], could have been utilized for similarity measurements in a far less time- and effort-consuming manner, but that, it was deemed, would defy the primitivistic premises of this study. To be loyal to the spirit of simplicity intrinsic to the idea of using indigenous sand drawings to predict the cell response to the nanoparticle therapy, the similarity score evaluation was performed using a self-devised method based on manually counting the number of elementary shapes present in compared patterns. Eleven of such basic shapes were constructed and they were divided based on the number of squares that they comprised: 3, 4 or 5 (Fig.7). Diagonal shapes were not included in the comparison because of the centrality of the diagonal lines to distances matrices in gen-

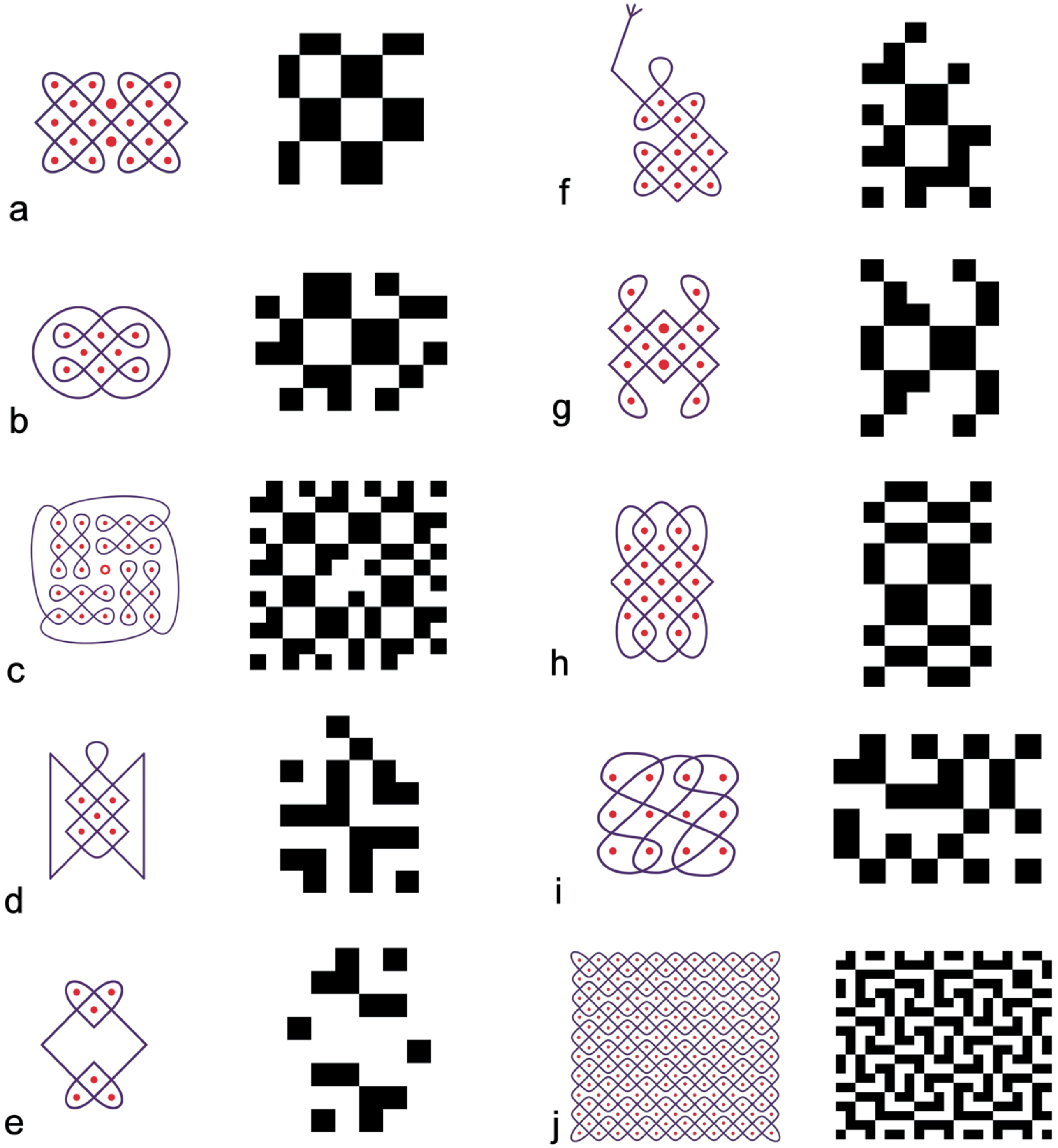


Figure 6. All of the Tchokwe sona ideograms and their corresponding checkerboard patterns used in this study: (a) a nest of doves, (b) an antelope's paws, (c) walking on tiptoes in a forest, (d) a fairy swallow (a.k.a. a wading bird or a stork), (e) a bridge made of sticks and lianas, (f) a human being, (g) the story about a bat and the sun, (h), fire, (i) looking sick, and (j) inside the belly of a lion.

eral (Fig.3). Similarity scores were measured manually for 3 cell images obtained in 3 separate cell experiments (n = 3 x 3) for each sample group, including the control one and the treatment one, with respect to each of the 10 sona patterns. The resulting similarity scores were then averaged within each sample group and compared against different treatments. The following equation was used to calculate the similarity score, S:

$$S = 3k + 3^2 \cdot k' + 4l + 4^2 \cdot l' + 5m + 5^2 \cdot m' \qquad (1)$$

Here, k, l and m represent the numbers of all the elementary shapes composed of 3, 4 and 5 squares, respectively, that were present in the two compared checkerboard patterns, one derived from the cell distance matrix and the other one from a lusona drawing. In the same equation, k', l' and m' represent the numbers of all the distinct elementary shapes composed of 3, 4 and 5 squares, respectively, in the two compared checkerboard patterns, one derived from the cell distance matrix and the other one from a lusona drawing. The difference between k, l and m values and k', l' and m' values is that while the former correspond to all shapes of a given type, including both the distinct ones and those that are integral to larger shapes in the pattern, the latter correspond only to shapes that are distinct and clearly delineable in the pattern; hence their higher contribution to S through the squaring of the number of them recognized in the two compared checkerboard patterns (Eq.1). This similarity metric is illustrated in Fig.8, showing one ■ shape and one ┐ shape in both patterns in Fig.8a. Since the former is composed of two ┐ shapes, $k = 3$ and $k' = 1$. Meanwhile, $l = 1$ and $l' = 1$ too, meaning that $S = 38$ as per Eq.1. Likewise, Fig.8b shows a hypothetic shape in which 4 different ┐ shapes could be discerned as well as one ■, three ⅃, one Π and three |. All of these shapes are immersed inside a more complex shape and only | can be discerned as a distinct one. Therefore, in a case where both compared patterns contain the same shape as that in Fig. 8b, $3k = 3\cdot3 + 4\cdot3 = 21$, $4l = 4\cdot1 + 4\cdot3 = 16$, and $5m = 5\cdot1$, while $3^2k' = 3^2\cdot1$, meaning that $S = 51$.

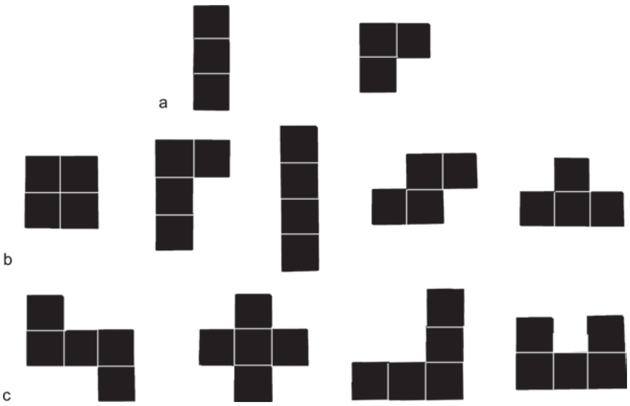


Figure 7. Basic shapes used to measure the similarity score between the digitalized cell distance matrix heat maps and the sona patterns. Each shape present in both of the compared patterns added the number of points equivalent to the number of blocks comprising it to the similarity score: 3 points (a), 4 points (b) or 5 points (c). Each shape that was not a part of a larger shape, but was distinct in both patterns added the squared number of blocks comprising it to the similarity score: 3^2 points (a), 4^2 points (b) or 5^2 points (c). Overall, the number of identical shapes in subfigures (a), (b) and (c) was used to derive the numbers k/k', l/l' and m/m', respectively, and compute the similarity score, S.

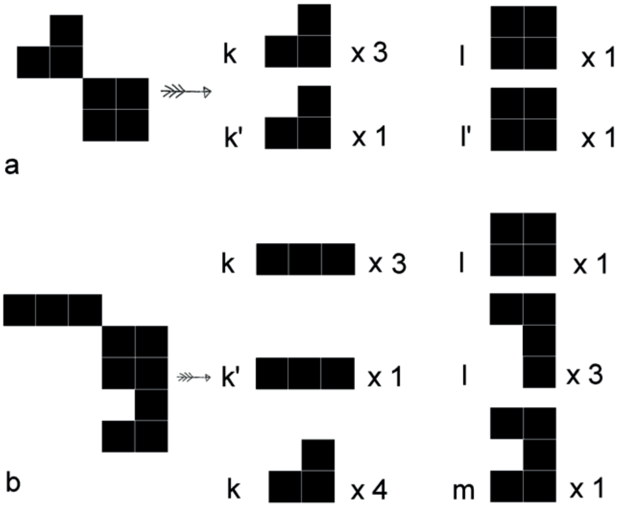


Figure 8. A schematic illustration of the assignment of k, k', l, l', m and m' values in two hypothetic cases corresponding to the identical shapes in (a) and (b) detected in both of the compared patterns, one derived from the pairwise cell distance matrices and another one derived from sona drawings.

The results of the comparison analyses demonstrated a significantly higher degree of similarity between the sona patterns and the cells subjected to the treatment than between the sona patterns and the control, untreated cells. Concordantly, Fig.9 shows that for 4 out of 10 sona patterns, namely a nest of doves, an antelope's paws, walking on tiptoes in a forest and the human being, the similarity with the treated cells was significantly higher than that with the control cells. The standard level of statistical confidence ($p < 0.05$) applied to the difference between these pairs of values. Simultaneously, for 3 more sona patterns, namely the story about the bat and the sun, fire and inside the belly of a lion, the similarity score was higher for the treated cells than for the untreated ones with a relatively high level of statistical confidence, averaging at $p = 0.087 \pm 0.030$, very near the conventional limit of $p = 0.05$. Furthermore, with respect to the 3 remaining sona patterns, namely a fairy swallow, a bridge made of sticks and lianas, and looking sick, the similarity score was higher for the treated cells than for the control ones, but with nil levels of statistical significance. These results unequivocally indicate that sona drawings could be a good early-stage predictor of the response of cells and tissues to a therapy administered in the form of magnetic nanoparticles.

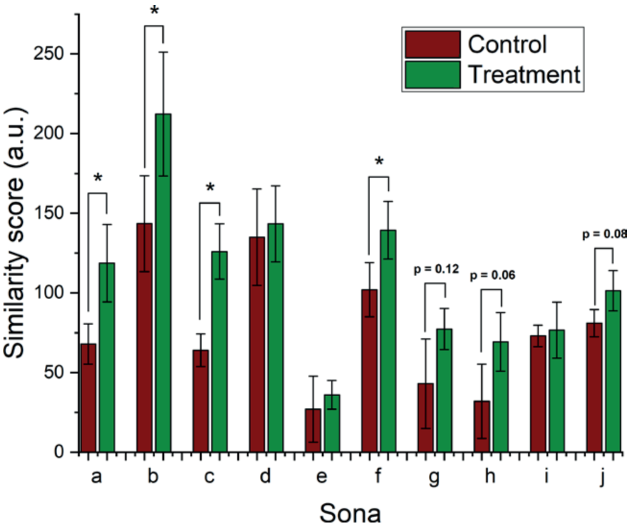


Figure 9. Comparison of similarity scores for different sona patterns and two cell populations, one that was left intact (control, wine) and one that was subjected to the nanoparticle therapy (treatment, green). Individual bars are data points representing averages (n = 3 x 3), while error bars represent the standard deviation. Data points statistically significantly different ($p < 0.05$) with respect to one another are connected with an asterisk. Comparative data points whose difference displayed moderate levels of statistical significance are connected with exact p-values. The reader is referred to Fig.6 for the identification of sona patterns marked with (a) through (j).

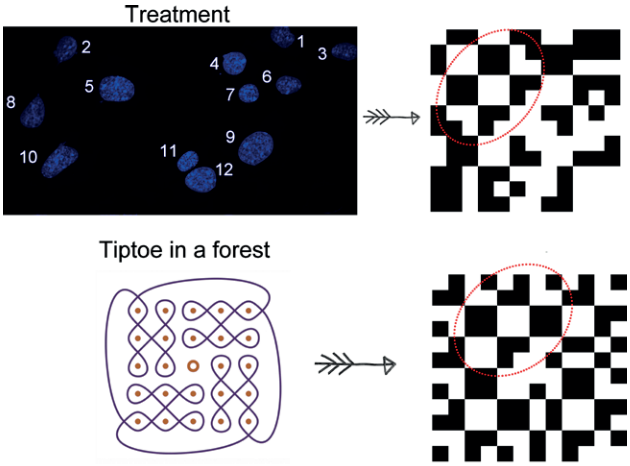


Figure 10. An example of a high level of similarity between two digital checkerboard patterns, one obtained from an image of the treated cells and another from a sona drawing representing tiptoeing in a forest. Adjacent squares and ┘ shapes are denoted as discernable in both the cell pattern and the comparative lusona pattern.

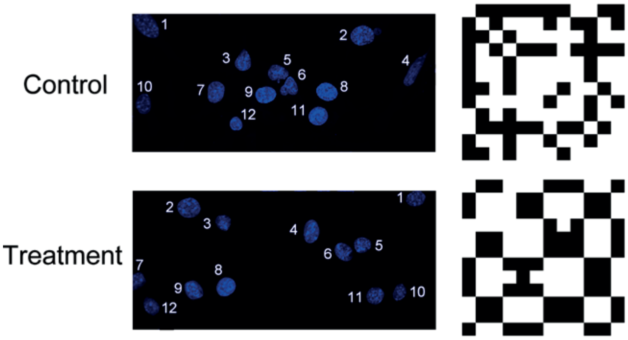


Figure 11. Two indistinct images of control and treated cell populations yielding drastically different cell distance checkerboard patterns.

Notwithstanding the apparently higher similarity of the sona patterns to the treated cell population than to the control one, the absolute similarity scores varied between different sona patterns within a single cell sample group. Thus, for example, the similarity with respect to the pattern representing a bridge made of sticks and lianas was very low, while the one with respect to the lusona illustrating tiptoes in a forest was very high. As for the latter, the simultaneous presence of distinct □ and ┐ shapes in both a representative cell pattern and the lusona pattern representing tiptoeing in the woods can be seen in Fig.10, serving as a testimony to this high level of similarity. Initially, it was hypothesized that certain sona patterns, such as those standing for feeling sick or finding oneself in the belly of a lion might bear higher levels of similarity with respect to cells thrown off the equilibrium by the treatment, while the control cells might be resonating more with sona representing the nest of doves or antelope's paws, but no such trends were noticed.

The overall trend noticed during the analysis, in fact, was that the treated cells, on average, gave considerably more ordered and symmetric digital cell distance patterns than the control cells (Fig.11), which is a significant insight in itself, unrelated to the similarity with the sona patterns. This is illustrated by the scores assigned to the total number of recognizable basic elements shown in Fig.7, which equaled 205 ± 30 per image for the control and 276 ± 42 per image for the treated population. Since sona drawings tend to be relatively symmetric, commonly possessing one or more mirror symmetry planes and rotational symmetry axes, including 2-fold (180 °), 3-fold (120 °) or 4-fold (90 °), it is thus logical to expect that their similarity with the treated cells would be higher than that with the untreated, control ones, and this was verified in the course of the similarity analyses. A completely different question is why the cells challenged by the treatment tend to adopt greater levels of translational orderliness, at least in the indirect metrics utilized here. An explanation referring to the contact between the nanoparticles and the cells as the one sending downstream signals that alter the gene expression, which in turn affects cell signaling and, thence, the intercellular distancing would be purely phenomenological and pines for a more detailed mechanistic insight. To start with, it could be noted that apoptosis and individuation are natural complements, given that cells embarking on the apoptotic path tend to estrange themselves from the cellular community. This tendency toward individuation may be one factor favoring the observed separation at more regular intervals among cells subjected to the nanoparticle therapy than among the control cells. As in concert with these findings, prior research has shown that imposition of an excessive translational order to the cells produces higher levels of mechanochemical stress, but leads to boosted metabolism[24]. The subtle effects of translational order in the environment around the cells and within the cell assemblies *per se* are largely unexplored and the developments of novel metrics, such as the one proposed here, are needed to understand these effects better.

4. DISCUSSION

Albeit merely scratched so far, the possibilities of interdisciplinary research appear to be limitless. One implicit objective of this study was to explore the cross-fertilization of two relatively distant disciplines, specifically cell biology and ancestral storytelling. The study emerged from an aspiration to prove the capacity of indigenous and largely forgotten sources of knowledge and perspectives at life to solve some of the most actual problems of the western science and medicine. In its course, the study demonstrated that the traditional drawings in the sand by the Tchokwe tribes and other indigenes of central Africa can be used to predict the evolution of the cells toward nonviable states based on a simple visual observation thereof and the processing of their spatial order using a primitivistic computational model devised and described here. Specifically, similarity between the digital pairwise distance patterns derived from the optical images of cells and analogous patterns derived from sand drawings representing a range of different objects and events was significantly higher for cell populations subjected to the nanoparticle therapy than for the control cells. The fact that this difference in similarity was observed at stages of the therapy when no apparent morphological deformities or cell density drops were noticeable suggests that this method may prove applicable for *in situ* monitoring of the response of tissues or organs to various therapies, in such a way that the slightest indications of adversities would be noticed timely and acted upon before their effects become irreversible. If so, we might witness a day when what African men and women drew in the sand to amuse and inspire their families and companions makes its way to the modern clinics as a part of tools capable of saving lives through avenues undreamt of by their humble originators.

By correlating the specifics of western cell biology with indigenous storytelling practices, this study has followed an authentically ethnoscientific approach. By definition, similarly to the way anticolonial strongholds provide a resistance to the urges to subdue the local cultures to the global trends and influences, ethnoscience has challenged the elitist views of universalistic science and its methods by engaging in the rediscovery of scientific worldviews and practices associated with local and largely underdeveloped cultural milieus not prominent on the map of the world. Sadly, however, how rare such ethnoscientific approaches have been so far is best illustrated by the fact that only a little over one hundred papers are currently listed in the Web of Science database satisfying the topic keyword “ethnoscience”, most of which, moreover, relate to the science of ethnography, which this term was first coined to describe in the 1960s. Phenotypic heterogeneities notwithstanding, the basics of cell biology, however, call attention to the malignancies associated with the rise of operational uniformity amongst cell aggregates in tissues, and it may just as well be the same with the social structures in which perspectives at science and other vital aspects of the society have grown exceedingly uniform. It is believed that engaging in studies aimed at discovering the practical epistemic ties between forgotten ethnic traditions and modern science would not only provide meaningful addenda to the repertoire of tools for predicting the cell response to therapy or any other form of physicochemical influence, but it may also provide an imperceptible contribution to healing the world at its broader, sociopolitical scales, which hover over our sciences interminably and, as it were, ominously.

It is understandable that cross-fertilization of disciplines, such as that accomplished here, will be met with disbelief by the bulk of the scientific community, but this is only because the ideal of interdisciplinarity has been approached very shyly throughout the past couple of decades. Bold embracement of this ideal and conception of models that transcend the boundaries of individual disciplines to reach out to some of their most distant counterparts may appear radical, but only because the stereotypical scientist of the modern day has not become accustomed to this approach. Another cause of disbelief is tied to the fact that out of all scientific disciplines, biomedical science has been possibly the most resistant one to intrusion of influence from humanities or related disciplines. Any transdisciplinary models that build on the current medical trends and technologies via cross-fertilization with arts or humanities have a high chance of being accused as pseudoscience, with the creator of such concepts better being ready to face the dire consequences of his free and reckless thought. The fact that the model devised here goes beyond a mere qualitative elaboration and employs a rigorous quantitative approach may not do it a favor either and it may only reinforce the aura of lunacy around its creator, and this only because of the unaccustomedness of the scientific community to the interdisciplinary crossroads like the one constructed here. However, if the heavily forested, wild and uncultivated path partially cleared with this work becomes followed by other authors and groups, the future may bring about similarly daring and provocative encounters of disciplines and they will no longer appear as outlandish as they do today. This is where the views endorsed here could be christened visionary, even though their genesis and dissemination have been purely accidental. Any of such views that may be deemed visionary have emerged from a simple desire to save the forgotten traditions of the world from their steady descents into oblivion. How this looking back fondly into a distant past paves way for the future is left for other progressive philosophers to untangle.

Another notion that may make many science mainstreamers cringe is that of analogy intrinsic at two places in this study: first at the point where McPhee's pioneering work in ethnomusicology inspired the ethnoscientific premises of this work and secondly, with respect to the correspondence between biological patterns of cell activity and patterns derived from the ancient African tradition of storytelling. Analogies, of course, are double-edged swords, as they underscore the mental reasoning of lunatics and of creative thinkers alike. Obviously, they can be misleading, as it can be exemplified by a number of cases from the history of science[25]. On the other hand, there are countless examples of revolutionary scientific concepts that stemmed from their originators' envisioning a lucid analogy between domains dominated by completely disparate phenomena[26]. In fact, the very notion of the model in the domain of scientific theory can be taken to imply that the description of a physical system has been modeled after something else, meaning that the role of analogies in establishing explanatory models in science must be nothing short of crucial. For this reason, reliance on analogies in science bears resemblance to holding a double-edged sword, which may turn into an effective tool one moment and the medium for self-destruction another. In defense of the use of analogies in this study, being inspired by it in terms of the advents in ethnomusicology is no sin. But neither is the attempt to compute conditions for a rigorous geometric correspondence between cellular patterns originating from a biomedical lab and sand patterns drawn by ancient African storytellers. The establishment of one such correspondence at the elemental geometric level in a statistically and methodologically solid way is sufficient to rid the model of any accusations of weak analogism. As ever, good science is the best remedy against its mediocre counterpart. However, the common fallacy of the mainstream scientific thought is to assign the epithet of superficiality or charlatanism to anyone enwrapping science with broader historical, geopolitical or aesthetic insights. What underlies one such fallacious deduction is nothing other but falling prey to the trap of analogy, namely between an unconventional breadth of the intellect expressed in unorthodox forms and the dilution of this intellect. To put it simply, the elaboration of the process of thought that inspired the ideas behind this study and the exposition of the broader social context that the study sprang from and that it could have an effect on should not cloud the analytical rigor and goodness of science that lies at its core.

Finally, it is worth reemphasizing that this has been primarily a conceptual study aiming to probe new interdisciplinary grounds and propose an entirely new methodology. Because of this conceptual nature of the study, an indefinite room for the perfection of the method is left for the up and coming pursuers of this "anticolonial" approach to science to probe. In that sense, should this study succeed in sparking the interest in further explorations of these and other ancestral drawings as sources of patterns usable in some of the most advanced spheres of science, it is conceivable that countless methods to process these geometries into codes applicable in a range of similarity algorithms would emerge. For example, instead of employing an intrinsically digital, 0/1 principle in the construction of the patterns, a series of equidistant values – for example, 1, 0.66, 0.33, 0, and over again - could be ascribed to successive squares covered by the finger as it draws the monolinear curve. Also, patterns could be constructed by mimicking the sliding of a quill pen down the paper and gradually decreasing the values assigned to squares traversed by the finger through the sand, as it is described in Fig.12. These amendments to the proposed model are explicated here because the goal, deep down, is not to force the reader into the adoption of a single computational worldview, but rather to inspire her to go beyond the model in ques-

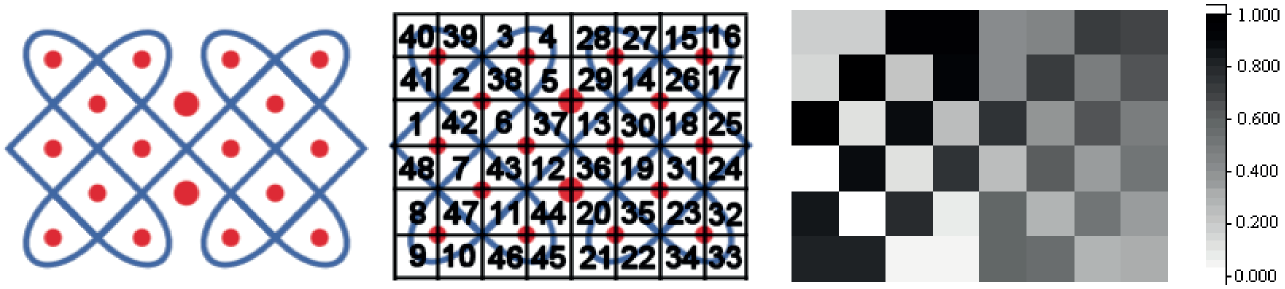


Figure 12. Transforming the Tchokwe lusona representing a nest of doves not into a digital (1010101…) heat map like the one used for comparison purposes, but to a heat map where the values assigned to the squares decline in direct proportion to the number of squares crossed by the monolinear line as it traces its way from the point of origin back to it.

tion and branch it into one or more out of infinite possible directions. This, needless to add, is in agreement with the broader aforementioned goal of incentivizing the global development in such a way that diversity of perspectives on knowledge creation is fostered rather than abolished in favor of intrinsically colonial, universalist approaches. Moreover, the matching of the patterns more complex than the digital patterns used in this work would likely require more sophisticated computer algorithms than the one employed here. In this study, such algorithms were deliberately stayed away from on the basis of the assumption that excessively complex computations would be at odds with the simplistic premises of the study. Namely, this is one out of many of the recent so-called backyard studies by this author, which fall under the umbrella of what has been christened as science of and for the poor[27],[28],[29]. As per this philosophy, the use of simplistic, inherently "poor" methods is favored over the use of their more exquisite opposites[30], whenever possible, lest the issues of disloyalty with respect to the essence of the philosophy of science of and for the poor become suspected. One such simplistic but strenuous method for pattern similarity scoring was created from scratch and proven feasible in this study, which is not to say that more computationally complex methods may not give even more reliable comparisons and predictions. Such more complex methods, however, are left for scientists better equipped with sophisticated computational skills and devices to devise and implement, which this author neither is nor aspires to be. Nevertheless, the seed of hope sown between these lines is that more computationally in-depth examinations of this model may eventually evolve into scale-invariant approaches for predicting phenomena in countless other existential domains, from biophysics to economy to human behavior. Hope also remains that the proof of concept presented here will arouse the interest of the scientific community in engaging in the discovery of similar models built on the interchange between natural sciences and unknown indigenous traditions of the world. One such path might gradually guide science away from a universalistic, neocolonial elitism and into humbler waters, more respectful of ancestral worldviews and indigenous practices of acquiring and sharing knowledge.

This study has demonstrated that ancestral drawings may appear simple on surface, but like all physical expressions, they can be converted to complex patterns with applicability in virtually every scientific discipline. As far as the Tchokwe sand drawings are concerned, they have inspired sparse research in graph theory[31], symmetry groups[32] and derivation of various elementary mathematical operations[33]. However, countless other similar visual forms, including constellations of stars, mandalas, kaleidoscope fractals, musical harmonies converted to geometric figures or ornaments in monasteries and temples could present similar sources of patterns that could be harnessed for practical uses in countless fields of science. Fishermen from tropical islands have had unique ways of counting and describing the aggregates of aquatic organisms, alongside pebbles, shells and notches[34], and such models can prove useful in predicting the outcomes of cellular populations challenged with various treatments at early time points based on their aggregation degrees or other interaction parameters. The distinct enumeration of coconuts and divination spirits native to the Caroline Islands in the Pacific, which influenced the measures used in the construction of habitats and social ranking[35], can be another source of numerical symmetries relatable to the ordering of cells in tissue culture. Yet another example may come from seafaring maps constructed by islanders from the South Seas[36], which could be tested with a little bit of imagination for their ability to predict cell fate in correspondence with asymmetries and other spatial inclinations exhibited by the cells. Implicitly, in fact, this study has aspired to spur the interest of researchers in such models that are made possible by the fractal nature of the physical reality, where similarities of patterns exist across all scales[37]. Such interests, if diligently explored, may bring about novel marriages between disciplines, which, just like progenies of genetically distant parents, may be freer of recessive degeneracies arising from stale reiterations of the same old paradigm. Many of such studies, as this author hopes, will move one step ahead of the most anticipatable scenario where the ethnic scientific component gets absorbed into the western scientific perspective, something that this study, itself, could be reprimanded for. Knowing that the criticisms of colonialism have been more often than not infected by that which they criticized[38],[39], this raises a cautionary viewpoint warning against biases that implicitly favor the epistemological grounds on which the viewer stands. Contrarily, here, the rarer and more outstanding approaches would be such that they absorb the western science points of view into a local, ethnic perspective and allow the latter to grow, producing along the way scientific methods, modalities of reasoning and forms of expression that are nothing like the conventional empirical settings, analytical frameworks and scientific papers produced in the West. Ultimately, such interdisciplinary crossings may reawake the renaissance interest in the imaginative association of hard science with arts, with positive repercussions on the viability of both not being able to be measured by all the grains of sand in this world. Until then, we should make sure that these lines in the sand, like Archimedes' circles drawn in the proximity of inconspicuous uniforms and boots exerting the misdemeanors of neocolonial elitism, are let be.

5. CONCLUSION

This study describes a primitivistic computational technique for predicting the response of cells and tissues to treatment at its early stages, when no gross morphological or cell density changes are detectable yet. The technique is based on measuring the degree of geometric correlation between transformed distance matrices of populations of cultured cells and digital checkerboard patterns derived from a set of traditional central African drawings in the sand. Experiments performed on osteosarcoma cells subjected to an anticancer therapy in the form of magnetic nanoparticles resulted in the detection of significant similarity between the patterns obtained by processing the pairwise distance matrices of treated cells and the patterns obtained from the indigenous sand drawings relative to the control. Based on these results, it can be concluded that the course of the therapy is predictable by comparing optical images of cells with indigenous ideographs processed using the metrics postulated by the model. The success of this method may provide an impetus for finding solutions to problems plaguing the modern medicine outside of the scope of its mainstream analytical frameworks and in the ancestral heritage of relatively obscure ethnic traditions.

6. REFERENCES

[1] R. Mueller. Bali, Tabuh-Tabuhan, and Colin McPhee's Method of Intercultural Composition. *Journal of Musicological Research* 10, 127–175, 1990.

[2] A. N. Whitehead. Science and the Modern World. Nolit, Belgrade, 1925.

[3] J. M. Galbraith, S. Shaik, D. Danovich, B. Braïda, W. Wu, P. Hiberty, D. L. Cooper, P. B. Karadakov, T. H. Dunning. Valence Bond and Molecular Orbital: Two Powerful Theories that Nicely Complement One Another. *Journal of Chemical Education* 98, 3617-3620, 2021.

[4] V. Uskoković. On Science of Metaphors and the Nature of Systemic Reasoning. *World Futures: The Journal of New Paradigm Research* 65, 241–269, 2009.

[5] U. D'Ambrosio. A Research Program and a Course in the History of Mathematics: Ethnomathematics. *Historia Mathematica* 16, 285–287, 1989.

[6] P. Gerdes. Re-Envisioning Mathematics Curricula in Higher Education. Presented at the 7th Annual Teaching & Learning Conference 2013, University of Kwazalu-Natal, 2013.

[7] E. F. Schumacher. Small is Beautiful: Economics as if People Mattered, 25 Years Later...with Commentaries. Hartley & Marks, Vancouver, BC, 1998.

[8] V. Uskoković, M. Ševkušić, D. P. Uskoković. Strategies for the Scientific Progress of the Developing Countries in the New Millennium: The case of Serbia in comparison with Slovenia and South Korea. *Science, Technology & Innovation Studies* 6 (1) 33–62, 2010.

[9] V. Uskoković. Ion-Doped Hydroxyapatite: An Impasse or the Road to Follow? *Ceramics International* 46 (8/B) 11443–11465, 2020.

[10] A. J. Bishop. Western Mathematics: The Secret Weapon of Cultural Imperialism. *Race & Class* 32, 51–65, 1990.

[11] V. Uskoković, S. Pernal, V. M. Wu. Earthicle: The Design of a Conceptually New Type of Particle. *ACS Applied Materials and Interfaces* 9 (2) 1305–1321, 2017.

[12] V. M. Wu, E. Huynh, S. Tang, V. Uskoković. Brain and Bone Cancer Targeting by a Ferrofluid Composed of Superparamagnetic Iron-Oxide/Silica/Carbon Nanoparticles (Earthicles). *Acta Biomaterialia* 88, 422–447, 2019.

[13] V. Uskoković, E. Huynh, S. Tang, S. Jovanović, V. M. Wu. Colloids or Powders: Which Nanoparticle Formulations Do Cells Like More? *Colloids and Surfaces B: Biointerfaces* 181, 39–47, 2019.

[14] V. Uskoković, T. A. Desai. Simultaneous Bactericidal and Osteogenic Effect of Nanoparticulate Calcium Phosphate Powders Loaded with Clindamycin on Osteoblasts Infected with Staphylococcus Aureus. *Materials Science and Engineering C: Materials for Biological Applications* 37, 210–222, 2014.

[15] J. Pijuan, C. Barcelo, D. F. Moreno, O. Maiques, P. Siso, R. M. Marti, A. Macia, A. Panosa. In vitro Cell Migration, Invasion, and Adhesion Assays: From Cell Imaging to Data Analysis. *Frontiers in Cell Developmental Biology* 7, 107, 2019.

[16] E. H. Gombrich. The Story of Art. Fifteenth Edition, Phaidon Press, London, pp. 601, 1989.

[17] M. Ascher. Graphs in Cultures–A Study in Ethnomathematics. *Historia Mathematica* 15, 201–227, 1988.

[18] P. Gerdes. Geometria Sona de Angola: Matematica duma Tradição Africana. Instituto Superior de Tecnologias e Gestão, Belo Horizonte, Mozambique, 2012.

[19] M. Fontinha, Desenhos na areia dos Quiocos do Nordeste de Angola, Instituto de Investigação Científica Tropical, Lisbon, 1983, pp. 163, 193, 263. Reproduced as Fig.39 in P. Gerdes. Geometria Sona de Angola: Matematica duma Tradição Africana. Instituto Superior de Tecnologias e Gestão, Belo Horizonte, Mozambique, 2012.

[20] P. Gerdes. Ethnomathematics and Education in Africa. Chapter 9. Exploration of the mathematical potential of 'Sona': An example of stimulating cultural awareness in mathematics teacher education. Instituto Superior de Tecnologias e Gestão, Belo Horizonte, Mozambique, pp. 145–174, 2014.

[21] S. Saha, M. Ghosh, S. Ghosh, S. Sen, P. K. Singh, Z. W. Geem, R. Sarkar. Feature Selection for Facial Emotion Recognition Using Cosine Similarity-Based Harmony Search Algorithm. *Applied Sciences* 10, 2816, 2020.

[22] J. B. Kho, A. B. J. Teoh, W. Lee, J. Kim. Bit-String Representation of a Fingerprint Image by Normalized Local Structures. *Pattern Recognition* 103, 107323, 2020.

[23] B. K. Patra, R. Launonen, V. Ollikainen, S. Nandi. A New Similarity Measure using Bhattacharyya Coefficient for Collaborative Filtering in Sparse Data. *Knowledge-Based Systems* 82, 163–177, 2015.

[24] V. Uskoković, T. A. Desai. Does Translational Symmetry Matter on the Micro Scale? Fibroblastic and Osteoblastic Interactions with the Topographically Distinct Poly(ε-Caprolactone)/Hydroxyapatite Thin Films. *ACS Applied Materials and Interfaces* 6 (15) 13209–13220, 2014.

[25] A. Sullivan-Clarke. Misled by Metaphor: The Problem of Ingrained Analogy. *Perspectives on Science* 27 (2) 153–170, 2019.

[26] V. Uskoković. Chemical Reactions as Petite Rendezvous: The Use of Metaphor in Materials Science Education. *Journal of Materials Education* 36 (1-2) 25–50, 2014.

[27] V. M. Wu, V. Uskoković. Waiting for Apataw: 250 Years Later. *Foundations of Science* 24 (4) 617–640, 2019.

[28] T. Uskoković, E. Uskoković, V. M. Wu, V. Uskoković. Calcium Phosphate and Senescence of Orange Jubilees in the Summertime. *ACS Applied Bio Materials* 3, 3770–3784, 2020.

[29] E. Uskoković, T. Uskoković, V. M. Wu, V. Uskoković …And All the World a Dream: Memory Effects Outlining the Path to Explaining the Strange Temperature-Dependency of Crystallization of Water, a.k.a. the Mpemba Effect. *Substantia* 4, 59–117, 2020.

[30] V. Uskoković. Celeste's Plight: What Can Film Teach Natural Science? *Film International* 17 (1) 69–88, 2019.

[31] S. Jablan, R. Sazdanović. LinKnot: Knot Theory by Computer, World Scientific, Singapore, 2007.

[32] D. Chavey. Strip Symmetry Groups of African Sona Designs, In: Proceedings of the 13th Annual Conference of BRIDGES: Mathematical Connections in Art, Music, and Science, Pécs, Hungary, July 24-28, 2010, pp. 111-118, 2010.

[33] P. Gerdes. Drawings from Angola: Living Mathematics, Research Centre for Mathematics, Culture and Education, Maputo, Mozambique, 2012.

[34] C. R. Hallpike. The Foundations of Primitive Thought. Clarendon Press, Oxford, UK, 1979.

[35] W. H. Alkire. Systems of Measurement on Woleai Atoll. *Anthropos* 65, 1–75, 1970.

[36] M. Ascher. Mathematics Elsewhere: An Exploration of Ideas across Cultures. Princeton University Press, Princeton, NJ, 2002.

[37] V. Uskoković. On the Relational Character of Mind and Nature. Res Cogitans: *Journal of Philosophy* 6 (1) 286–400, 2009.

[38] G. C. Spivak. Three Women's Texts and a Critique of Imperialism. *Critical Inquiry* 12, 243–261, 1985.

[39] D. Hanlon. The 'Sea of Little Lands': Examining Micronesia's Place in 'Our Sea of Islands'. *The Contemporary Pacific* 21, 91–110, 2009.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

OPEN ACCESS

Citation: Kragh H. (2022) Chemists Without Knowing It? Computational Chemistry and the Møller-Plesset Perturbation Theory. *Substantia* 6(2): 43-54. doi: 10.36253/Substantia-1564

Received: Feb 01, 2022

Revised: Apr 1, 2022

Just Accepted Online: Apr 15, 2022

Published: September 1, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Chemists Without Knowing It? Computational Chemistry and the Møller-Plesset Perturbation Theory

Helge Kragh

Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, Copenhagen, Denmark.
E-mail: helge.kragh@nbi.ku.dk

Abstract. This paper considers aspects of the chemistry-physics relationship from a historical perspective and with a focus on the entrance of quantum mechanics in twentieth-century chemistry. Traditionally, theoretical physics was widely regarded as epistemically superior to chemistry if also, from the chemists' point of view, of little practical relevance. With the emergence of quantum chemistry in about 1930, the gulf widened as it seemed that the new discipline was more physics than chemistry. One way of investigating theoretically many-electron atoms was by means of the Hartree-Fock approximation method. The Møller-Plesset perturbation theory introduced in 1934 by a Danish and an American physicist was a refinement to the Hartree-Fock method. Although the Møller-Plesset theory was initially neglected – and is still neglected in the historiography of quantum chemistry – it came to play a most important role in later studies. Indeed, it is a prime example of what in sociological studies of science is known as a "sleeping beauty." The paper discusses the historical context of the Møller-Plesset theory, concluding that, in a sense, its originators were "chemists without knowing it."

Keywords: quantum chemistry, chemistry-physics relations, Møller-Plesset theory, chemistry Nobel Prizes, sleeping beauties.

1. INTRODUCTION

Much has been written on the physics-chemistry relationship from both a historical and philosophical perspective.[1] In the first part of this paper I briefly discuss how physicists have often considered chemistry a science inferior to their own (Section 2). This somewhat condescending attitude was only reinforced with the advent of the so-called old quantum theory principally due to Niels Bohr and Arnold Sommerfeld (Section 3). The second part is devoted to the emergence and early development of quantum chemistry based on post-1925 quantum mechanics (sections 4-5). Rather than dealing with the chemical bond, a classical and well-researched area of the quantum-chemical revolution, the paper looks at a theory of many-electron systems from 1934 which seems to have been forgotten in the historical literature

(Section 6). This theory, the Møller-Plesset perturbation theory or method, has an interesting and little-known history. It illustrates in its own way how physicists unintendedly could make important contributions to theoretical chemistry. Moreover, it also illustrates the concept of a "sleeping beauty," a term used for scientific papers which hibernate for a long time until they are called alive and become highly influential (Section 7). In the final Section 8 I briefly reconsider the physics-chemistry relationship in the light of the history of the Nobel prizes awarded to either physicists or chemists.

2. CHEMISTRY VERSUS PHYSICS? THE EARLY PHASE

The relationship between the two sister sciences chemistry and physics has never been fixed but always in a state of flux. Traditionally, physics was considered the big brother, a much nobler and more scientific field than the supposedly primitive and empirical chemistry. This is an old image still widely shared in the public and also, regrettably, by many scientists. As far back as 1669 Bernard Fontenelle, the perpetual secretary of the Académie Royale des Sciences in Paris, wrote as follows:

> *Through its visible operations, chemistry resolves bodies into a certain number of crude tangible principles; salts, sulfurs, etc. while through its delicate speculations, physics acts on the principles as chemistry acts on bodies, resolving them into other even simpler principles, small bodies fashioned and moved in an infinite number of ways. … The spirit of physics is clearer, simpler, less obstructed, and, finally, goes right to the origins of things, while the spirit of chemistry does not go to the end.*[2]

More than a century later, after Newton's mechanical physics had been generally accepted, Immanuel Kant repeated Fontenelle's message of chemistry's lower epistemic status. Not only was chemistry inferior to Newtonian physics, according to Kant it was not even a genuine science and could never become one. The problem was that by its very nature laboratory-based chemistry was, or was claimed to be, intractable to the mathematical method and systematic deduction from higher laws or principles. As Kant expressed it in his *Metaphysische Anfangsgründe der Naturwissenschaft* from 1786:

> *Chemistry can be nothing more than a systematic art or experimental doctrine, but never a proper science, because its principles are merely empirical, and allow of no a priori presentation in intuition. Consequently, they do not in the least make the principles of chemical appearances conceivable with respect to their possibility, for they are not receptive to the application of mathematics.*[3]

At the time when Kant degraded chemistry to a non-science there were already a few attempts to apply mechanical physics to chemical phenomena, indeed to subjugate the latter under the former. Thus, in an ambitious work of 1758 with the characteristic title *Essai de Chymie Mécanique* the Swiss natural philosopher Georges-Louis Le Sage claimed to have explained chemical affinity and properties of matter purely in terms of mechanical physics. According to Le Sage, cohesion, affinity, and gravitation were all aspects of the same general law of mechanics.[4]

Later in the century, the Newtonian paradigm came to be highly regarded by chemists and physicists in the French tradition mainly due to Pierre-Simon Laplace and Claude Louis Berthollet. However, the dream of a Newtonian chemistry was more rhetoric than reality. It remained a dream and in the early part of the nineteenth century it was realized to be a dead end.[5] During the last quarter of the century the dream was shortly revived in the version of "vortex chemistry" based on the mathematically advanced theory of the vortex atom proposed by William Thomson, J. J. Thomson, William Hicks, and other British physicists. However, to the large majority of chemists this theory was too much physics and mathematics, and too little chemistry. Latest by the turn of the century vortex chemistry (and vortex atom theory generally) was abandoned.[6] Still, a few chemists in Britain and the United States responded favorably to J. J. Thomson's vision of a vortex chemistry. Harry C. Jones, a physical chemist at Johns Hopkins, referred positively to the theory in a textbook of 1902. And according to Francis Venable, professor of chemistry at the University of North Carolina, the vortex theory of matter offered a future theory of everything which included all chemical phenomena.[7]

When speaking about the chemistry-physics relationship over long periods of time it is important to avoid anachronisms and keep in mind that the terms "chemistry" and "physics" once had different meanings than they have today. What Fontenelle and Kant referred to with these terms cannot be directly translated into the sciences as known in the second half of the nineteenth century. This said, we shall first briefly consider some aspects of the relationship between chemistry and physics in the period from about 1880 and 1920. By that time Kant's claim of chemistry as a non-scientific art had long been contradicted by its explosive development. And yet the epistemic status of chemistry as compared to that of physics remained a matter of discussion.

With the emergence of chemical thermodynamics, a highly abstract yet empirically powerful theory developed principally by Josiah Willard Gibbs in the United

States and Hermann Helmholtz in Germany, it seemed for a while that chemistry had been solidly founded in universal laws of nature. However, it was a foundation laid by physicists with no experience in laboratory chemistry and no high opinion of the kind of work most chemists were engaged in. As Helmholtz arrogantly expressed it in a letter of 1891: "Thermodynamic laws in their abstract form can only be grasped by rigorously schooled mathematicians, and are accordingly scarcely accessible to the people who want to do experiments on solutions and their vapor tensions, freezing points, heats of solution, &c."[8] Helmholtz referred somewhat condescendingly to the new school of physical chemistry established by Svante Arrhenius, Jacobus van' t Hoff, Wilhelm Ostwald, and others.

Nevertheless, by the turn of the century a small group of chemists had become "rigorously schooled mathematicians" who mastered the abstract theory of thermodynamics. One of them was the Dutch chemist Johannes van Laar who in a series of works cultivated thermodynamics as the royal road to what he called "mathematical chemistry."[9] On the other hand, although thermodynamics was a major step toward integrating physics and chemistry, it was at most a partial integration. After all, thermodynamics is concerned only with the state functions and bulk matter, whereas it is not applicable to the chemical elements and compounds – or to atoms and molecules – that chemistry is first and foremost about.

Traditional chemists found the mathematics of physical theories to be incomprehensible as well as irrelevant for their science, and there were also other reasons why many of them resisted what they felt was an intrusion of physics into chemistry. One of the reasons was the discovery in the 1890s of radioactivity and the electron. The great Dmitri Mendeleev was in some respects a traditionalist who firmly believed that if the physicists' subatomic particle (the electron) and transmutation of elements (radioactivity) were accepted, chemistry would degrade into a pre-scientific state.[10] He thought to have found an alternative to the new physics in the form of the ether, which he, contrary to the physicists, conceived as an ultralight chemical element with a place in the periodic table. Referring to radioactivity and what he called the "metachemical" electron, he stated: "It is my desire to replace such vague ideas by a more real notion of the chemical nature of the ether."[11] Mendeleev wanted to establish the supremacy of chemistry over the new physics, but the large majority of scientists – whether physicists or chemists – ignored his grand project.

Although Mendeleev's proposal of incorporating the world ether as an essential part of chemistry failed, several contemporary chemists shared his skeptical or even hostile attitude to the new physics. One example is Arthur Smithells, professor of chemistry at the University of Leeds, who in a presidential address to the 1907 meeting in Leicester of the British Association for the Advancement of Science (BAAS) warned against what he called the "invasion" of chemistry by mathematics and physics:

> *With radioactivity, in relation to the ponderable, we seem almost to be creating a chemistry of phantoms … associated as it is with the exuberance of mathematical speculation of the most bewildering kind concerning the nature, or perhaps I should say the want of nature, of matter. … Though chemistry and physics meet and blend there is an essential difference between the genius of physics and the genius of chemistry. Apart from his manipulative skills, the latter is not given to elaborate theories and is usually averse to speculation; nor has he the usually an aptitude in mathematics. … Chemistry should not be invaded by mathematical theorists.*[12]

Henry Armstrong, like Smithells a chemical traditionalist, noted in another presidential address to the chemical section of BAAS two years later that "now that physicists are regular excursionists into our territory, it is essential that our methods and our criteria be understood by them." He found it "a serious matter that chemistry should be so neglected by physicists."[13]

3. QUANTUM THEORY ENTERS CHEMISTRY

The gulf between chemistry and physics only deepened with the advent of Niels Bohr's quantum atom which not only addressed physical problems but also chemical. After all, the title of his three seminal articles in the 1913 volume of *Philosophical Magazine* was "On the Constitution of Atoms and Molecules." Bohr applied the new theory to problems which traditionally belonged to chemistry, such as the heat of formation of molecular hydrogen, the covalent bond, and the periodic system of the elements.[14] In an address of 1920 given to the Royal Danish Academy of Sciences, he cautiously suggested that in the future theoretical chemistry might become a branch of atomic physics: "Since … a possibility has been opened up of interpreting chemical experiences with the aid of considerations originating in the so-called physical phenomena, a connection between physics and chemistry has been created which does not correspond to anything conceived of before."[15]

Other theoretical physicists were more direct in their reductionist attitude to the physics-chemistry relationship. Max Born in Göttingen tended to see chemistry as inferior to physics because chemistry – which he, contrary to Kant, after all admitted as a proper science – lacked a mathematical foundation. To illustrate his point, he made use of a military metaphor:

> *We realize that we have not yet penetrated far into the vast territory of chemistry, yet we have travelled far enough to see before us in the distance the passes that must be traversed before physics can impose her laws upon her neighbor science.*[16]

For a while, several physicists thought that Bohr, with his new quantum theory of atoms and molecules, might become the new Newton who succeeded in basing chemistry deductively on the higher principles of physics. The British physicist Oliver Lodge referred implicitly to the Kantian dream of a mathematized chemistry when he lyrically wrote about "The brilliant attempts at further analysis of the atoms of all the chemical elements, so as to deduce their properties – the full beauty of atomic astronomy which is now unfolding before the eyes of enthusiastic experts." He concluded that, "we are living in the dawn of a kind of atomic astronomy which looks as if it were going to do for Chemistry what Newton did for the Solar System."[17] Born's colleague in Göttingen, the great mathematician David Hilbert, was more explicit. According to him, the desired reduction of chemistry to physics required "a Newton of atomic theory, and this has been Niels Bohr, who on the basis of new physical ideas, namely, the quantum theory, made a deeper understanding of this area a possibility."[18]

However, not everyone, and as expected not the chemists in particular, agreed that Bohr was a new Newton or, for that matter, that chemistry needed to comply with the strange laws of quantum physics. In fact, Bohr's attempt to extend the quantum theory of atomic structure to the realm of chemistry was considered unconvincing by most chemists. Their dissatisfaction with the semi-classical Bohr atom was given voice by the American physical chemist Richard Tolman, who, contrary to many of his colleagues in the chemical sciences, was also an accomplished mathematician. In an address delivered in Toronto in 1921, he objected to Bohr's postulates of stationary states and the mechanism of light emission in terms of quantum jumps. With respect to the Bohr-Sommerfeld atom he said that it was "constructed by the physicists, like a solar system … partly because they were entirely unfamiliar with the actual facts concerning the behavior of atoms in chemical combination."[19] Moreover, pretending to represent the average chemist, he stated the chemist's point of view as

> *… extreme hostility to the physicists, with their absurd atom, like a pan-cake of rotating electrons, an attitude which is only slightly modified by a pious wish that somehow the vitamin "h" [Planck's constant] ought to find its way into the vital organs of their own, entirely satisfactory cubical atom. … In general I feel that the cubical atom of Lewis and Langmuir must be regarded as representing chemical facts better than anything proposed by the physicists.*

Without going into further detail, molecular structure remained an unsolved problem within the framework of the old quantum theory which was unable to explain even the simplest molecules such as H_2 and H_2^+. The result was that the majority of chemists disregarded the Bohr-Sommerfeld theory and instead adopted the "cubical atom" with fixed electrons such as proposed by Gilbert N. Lewis and Irving Langmuir in particular. Although this kind of atomic model was pure nonsense according to the quantum physicists, from the point of view of the chemists it was useful and of great heuristic value.[20] The objections of the physicists were summarized by Edward Andrade, professor of physics at the Artillery College, Woolwich, who wrote about Langmuir's model of the atom: "It is scarcely necessary to insist on the artificiality of this picture… The electrons in Langmuir's atom have, in fact, so few of the known properties of electrons that it is not immediately clear why they are called electrons at all."[21]

4. THE EMERGENCE OF QUANTUM CHEMISTRY

Chemical considerations played no role in the establishment of quantum mechanics as the theory was formulated 1925-1926 principally by Heisenberg, Born, and Jordan (matrix mechanics), Dirac (q-number algebra), and Schrödinger (wave mechanics).

As seen in retrospect, modern quantum chemistry took its beginning with a seminal paper of 1927 written by two German physicists, 27-year-old Fritz London and the four years younger Walter Heitler. The title of their paper in *Zeitschrift für Physik* was "Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik" (The Interaction of Neutral Atoms and Homopolar Bond According to Quantum Mechanics).[22] The basic approach of the Heitler-London theory of the H_2 molecule was to consider separately one of the electrons in each of the combining atoms and then, by means of approximation methods and taking into regard the recently discovered resonance effect, to construct a wave function representing the paired-electron bond between them. Apart from explaining the formation of H_2 from two hydrogen atoms, they also explained why two helium atoms cannot form a He_2 molecule. Without making use of any empirical data, Heitler and London estimated from their *ab initio* calculations the dissociation energy of the molecule to be about 2.4 eV.

In a more general sense the main result of the Heitler-London paper was its deductive argument that the covalent bond can be explained purely in terms of spin quantum mechanics and is therefore outside the reach of classical chemistry. The pioneering work of the two physicists suggested a mathematization of chemistry more real and thorough than what had previously been dreamt of. It can be regarded as yet another example of the "invasion" of a chemical territory by theoretical physicists with basically no background in or knowledge of laboratory chemistry. In a letter to Heitler from 1935, London indicated his lack of respect for the culture of chemistry:

> *The word "valence" means for the chemist something more than simply forces of molecular formation. For him it means a substitute for these forces whose aim is to free him from the necessity to proceed, in complicated cases, by calculations deep into the model. … The chemist is made out of hard wood and he needs to have rules even if they are incomprehensible.*[23]

Given Linus Pauling's background in chemistry and later reputation as a "chemical translator," it is noteworthy that in his younger days he subscribed to the reductionist view expressed by some quantum theorists. Thus, in a lecture from 1928 to the American Chemical Society, he stated that chemistry was a kind of by-product of theoretical physics:

> *We can now predict with a considerable measure of confidence the general nature of the future advances [in theoretical chemistry]. We can say, and partially vindicate the assertion, that the whole of chemistry depends essentially upon two fundamental phenomena: these are (1) the one described in the Pauli Exclusion Principle; and (2) the Heisenberg-Dirac Resonance Phenomenon.*[24]

Pauling's assertion had more than a little in common with Dirac's better-known claim from the following year (see below).

The approach of Heitler and London formed the backbone of what came to be known as the valence bond (VB) method, which in the version developed by Pauling and others dominated quantum chemistry during the 1930s. The alternative molecular orbital (MO) method can be traced back to a paper that yet another German physicist, Friedrich Hund, published in 1927. Hund assumed that an individual electron moved in the field from all the nuclei and the other electrons in the molecule. His approach was paralleled by works done by Robert Mulliken, who contrary to Hund had a solid chemical training and a PhD in physical chemistry. At around 1931 the two methods, valence bond and molecular orbitals, reached immaturity and quantum chemistry entered as a social and scientific reality.

The concept of "chemical physics" was well known in the nineteenth century, when it typically referred to chemical agents such as heat, light, and electricity.[25] It now re-emerged in a different form which indicated the growing autonomy of quantum chemistry. The *Journal of Chemical Physics* was founded in 1933 with Harold Urey as its first managing editor. In the first issue Urey reflected on the old theme of the relationship between physics and chemistry, which he thought had entered a new and fruitful phase of symbiotic cooperation:

> *At present the boundary between the sciences of physics and chemistry has been completely bridged. Men who must be classified as physicists on the basis of training and of relations to departments or institutes of physics are working on the traditional problems of chemistry; and others who must be regarded as chemists on similar grounds are working in fields which must be regarded as physics.*[26]

The new journal was aimed for contributions too mathematical for *Journal of Physical Chemistry*, too physical for *Journal of the American Chemical Society*, and too chemical for *Physical Review*. Although it contained many articles on quantum chemistry, *Journal of Chemical Physics* was not devoted to this branch of science. As many or more articles were on molecular spectroscopy, kinetics of reactions, materials science, and more traditional areas of physical chemistry. Only in 1967 did the community of quantum chemists get its own journal, the *International Journal of Quantum Chemistry* created by Per-Olov Löwdin, the influential Swedish expert and a major force in establishing quantum chemistry as a proper scientific discipline.[27]

Most but not all of the post-World War II generation of chemists came to realize that theoretical chemistry is essentially based on the laws of quantum mechanics. "The whole of chemistry is one huge manifestation of quantum phenomena," wrote Carl Johan Ballhausen, a professor of chemistry at the University of Copenhagen best known for his important contributions to so-called ligand field theory. According to Ballhausen:

> *Without a background in quantum theory it is impossible to possess an "in depth" understanding of chemistry. The elucidation of chemical phenomena by means of the quan-*

> *tum laws is now left to the chemists; the solid state physicists do not have the necessary chemical background and the high energy physicists are not interested in electrons. Let us therefore think in deep gratitude and admiration to those pioneering physicists who opened the doors to modern chemistry.*[28]

Ballhausen and his contemporaries had no problem with recognizing quantum chemistry as based on work done by physicists and only subsequently developed by chemists.

5. EARLY MANY-ELECTRON THEORIES

Although not originally recognized to belong to the domain of quantum chemistry, in the years around 1930 several physicists dealt with the problem of calculating systems with many electrons. One of them was Paul Dirac, who in a paper of 1929 studied the exchange interaction of several identical particles such as electrons belonging to the same atom. When Dirac's paper is still cited today it is not so much because of its scientific content but rather because of its introductory remarks concerning the hypothetical reduction of chemistry to physics:

> *The general theory of quantum mechanics is now almost complete … The underlying physical laws necessary for the mathematical theory of a large part of physics and* the whole of chemistry *are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed.*[29]

Dirac referred in his paper not only to the Heitler-London theory but also to an important theory by the Cambridge mathematical physicist Douglas R. Hartree, who in a series of papers published 1928-1929 in the *Proceedings of the Cambridge Philosophical Society* introduced the so-called self-consistent field approximation method for calculation of many-electron atoms.[30] The general idea of this method was to reduce the many-electron problem to a one-electron problem, which was done by representing the effect of an electron on other electrons by a sort of average field corresponding to a central non-Coulomb field of force. In this way Hartree could obtain an approximate solution to the Schrödinger equation even for fairly complicated atomic systems (such as Na^+ and Cl^-) that agreed well with observed values.

However, as was realized early on by John Slater and a few other physicists, the Hartree method was in some respects flawed as it disregarded the spin states and the associated Pauli exclusion principle.[31] In 1930 the Russian physicist Vladimir Fock published a mathematically complex paper in *Zeitschrift für Physik* in which he improved the method by taking into consideration that the indistinguishability of electrons give rise to exchange forces.[32] The result was what soon became known as the Hartree-Fock approximation method, which since then has played an important role in quantum-chemical calculations. However, initially the method was applied exclusively to small and medium-sized atoms, and even in these cases calculations based on the Hartree-Fock theory were laborious. To extend the method to molecules required computational resources that were available only in the post-World War II era.

Whereas the papers of Hartree and Fock attracted critical interest among physicists, as indicated by citations to them they were largely ignored by the chemists. Thus, during the period 1928-1932 Hartree's first paper on the self-consistent field method received 27 citations, all of them in *Physical Review* or other physics journals (Google Scholar). The picture is the same with Fock's paper, which during 1930-1934 was cited 9 times. There were no citations to either of the papers in journals of chemistry or physical chemistry.

The Hartree-Fock method was not the only approach to many-electron calculations developed in the 1930s. In calculations based on this method, the interaction between electrons of opposite spins was taken into account only by means of an average interaction. To remedy for this deficiency various so-called electron correlation methods were developed, the first and arguably most important of which was the Møller-Plesset perturbation theory dating from 1934. According to a review paper of 2011:

> *In 1934, Møller and Plesset described in a short note of just five pages how the Hartree-Fock (HF) method can be corrected for electron pair correlation by using second-order perturbation theory. This approach is known today as Møller–Plesset perturbation theory, abbreviated as MPPT or just MP in the literature. MPPT, although in the beginning largely ignored, had a strong impact on the development of quantum chemical ab initio methods in the past 40 years.*[33]

Although thousands of papers have been written on this widely used perturbation method, many of them referring to the original paper from 1934, it and its two authors are nearly invisible in the historical literature on quantum and computational chemistry.[34] Most likely, very few of the modern scientists using the method and referring to the 1934 paper have any idea of whom

Møller and Plesset were. So, who were they and what was the context of their contribution to what retrospectively can be identified as the history of quantum chemistry? If the Møller-Plesset theory had a strong impact only from about 1970, what about the earlier history?

6. MØLLER-PLESSET THEORY IN THE 1930S

As Christian Møller and Milton Plesset stated in their abstract, "A perturbation theory is developed for treating a system of *n* electrons in which the Hartree-Fock solution appears as the zero-order approximation." And later in the paper: "Thus, the perturbation method shows that the theory of the self-consistent field is accurate in the determination of energy to the second approximation."[35] In other words, Møller and Plesset used the Hartree-Fock theory as a starting point but added a small perturbation given by the deviation of the Hartree-Fock Hamiltonian (energy operator) from the exact Hamiltonian. The perturbation term of the second order corresponded to the electron-electron interaction neglected in the Hartree-Fock theory.

Møller is described in a Wikipedia article on him as "a Danish chemist and physicist," which is a gross mistake given that neither he nor his coauthor Plesset ever worked in or published on chemistry.[36] They were quantum physicists with no interest whatsoever in chemistry or even recognizing that their short paper in *Physical Review* belonged to the new fields of chemical physics and quantum chemistry. The two authors did not offer any application or calculation, as for instance Hartree did. They considered their work to be a contribution to theoretical quantum mechanics and no more than that. In fact, the words "atom" and "molecule" did not appear in the article, which also did not mention "chemistry" or related terms. Nor did it refer to experimental data of any kind.

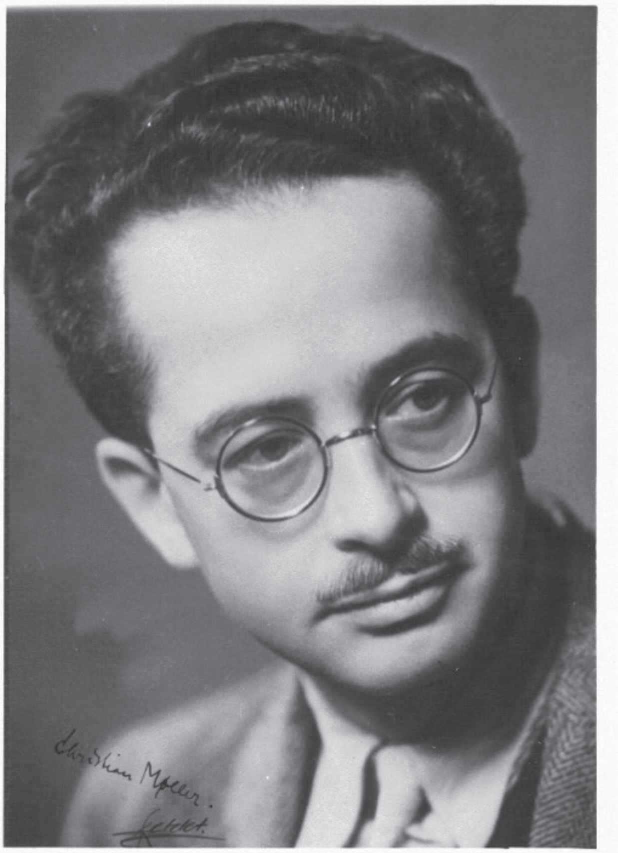


Figure 1. Portrait photography of Christian Møller, 1936. Credit: Niels Bohr Archive, Photo Collection, Copenhagen.

Christian Møller (1904-1980) was a 30-year-old Danish physicist at Bohr's institute of theoretical physics in Copenhagen (Figure 1). At the time he was best known for an important quantum theory of relativistic electron-electron scattering, a phenomenon which became known as Møller scattering.[37] By 1934, when he collaborated with Plesset, he had begun working on Enrico Fermi's new theory of beta-radioactivity. Shortly after having completed the paper with Plesset, he went to Rome on a Rockefeller stipend to work with Fermi' group. Later in his career Møller turned to the meson theory of nuclear forces, to which he contributed with several papers in the period from 1939 to 1946. In 1943 he was appointed professor of mathematical physics at the University of Copenhagen. Since the early 1950s Møller focused increasingly on the theory of general relativity on which subject he became internationally recognized as a leading expert. His authoritative and much-used textbook from 1952 titled *The Theory of Relativity* played an important role in the so-called renaissance of general relativity.

Milton Spinoza Plesset (1908-1991) earned his PhD at Yale University in 1932 and subsequently moved to Caltech, where he worked with J. Robert Oppenheimer on positron theory and problems of quantum electrodynamics. In 1933 he went to Bohr's institute in Copenhagen on a National Research Council fellowship. In September that year he participated in the annual institute conference with, among others, Møller, Dirac, Heisenberg, and Heitler (Figure 2). While in Copenhagen he also accompanied Bohr and his wife on a visit to the Soviet Union in May 1934, where he met Fock and other Russian physicists. And then he found time to collaborate with Møller on the many-electron theory which came to bear their names.

Figure 2. Conference at Bohr's institute, September 1933. On the first row: N. Bohr, P. Dirac, W. Heisenberg, P. Ehrenfest, M. Delbrück, L. Meitner. Milton Plesset is on the second row, number four from the right, and Christian Møller seats behind him, number three from the right on the third row. Credit: Niels Bohr Archive, Photo Collection, Copenhagen.

Figure 3. Milton Plesset (left) with N. Bohr, F. Kalckar, E. Teller, and O. Frisch at the Copenhagen institute in 1934. Credit: Niels Bohr Archive, Photo Collection, Copenhagen.

Bohr valued Plesset as a very promising physicist, such as evidenced in a letter he wrote to an American colleague:

> *Surely he is one of the best of young American theoretical physicists and especially he has as you know a great insight in the relativistic quantum theory of the electron… He hopes soon to publish an account of some of his work together with [John] Wheeler, and you may perhaps has seen a recent paper in the Physical Review on the many-electron problem, which he published a few months ago together with Møller.*[38]

After having worked in Copenhagen and elsewhere in Europe, Plesset returned to Caltech where he was appointed professor in engineering science in 1963. His scientific work after World War II was mostly concerned with fluid dynamics and nuclear physics (Figure 3).[39]

The paper by Møller and Plesset was predominantly mathematical, with no indication at all of the areas of physics and chemistry to which the theory might be applied. Apparently they did not care. None of the two Copenhagen physicists considered their work to be important and they never returned to it or related fields of science. When Plesset was interviewed in 1981, he did not even mention Møller and the work he did with him.[40]

As shown by the number of citations to the 1934 paper, for a long period the Møller-Plesset theory was ignored. During the 1930s the paper received only 5 citations, all of them in *Physical Review* except one in *Journal de Physique et le Radium* written by the Swedish theoretical physicist Oskar Klein. While two of the citing papers considered the theory in relation to nuclear structure, none of them referred to the electron structure of atoms. By 1962 the cumulative number had increased to 22, less than one citation per year in average. In other words, the Møller-Plesset perturbation method was scarcely visible in the scientific literature. However, the poor record changed drastically from about 1980 – the year that Møller died – and today the total number of citations to the Møller-Plesset paper has exploded to about 16,600 according to Google Scholar or 13,600 according to Web of Science (Figure 4). Of course, when evaluating such citation curves one has to take into consideration the general growth in the numbers of publications in the period.[41] But even then the Møller-Plesset citation curve is highly anomalous.

7. A SLEEPING BEAUTY

The fate of the Møller-Plesset theory only changed when computers began to be widely used to solve problems in chemical physics and quantum chemistry. According to Dieter Cremer, a German specialist in computational chemistry, perturbation theory as a tool in theoretical chemistry was rediscovered in the 1960s and from the mid-1970s onwards this kind of theory developed rapidly.[42]

So-called MP*n* methods – meaning Møller-Plesett theories of perturbation order *n* – were developed by, among others, the British-American theoretical chemist and later Nobel laureate John Pople.[43] Due to the works of Pople and his collaborators, the old Copenhagen paper of 1934 became much better known. Pople concluded that the original Møller-Plesset method carried

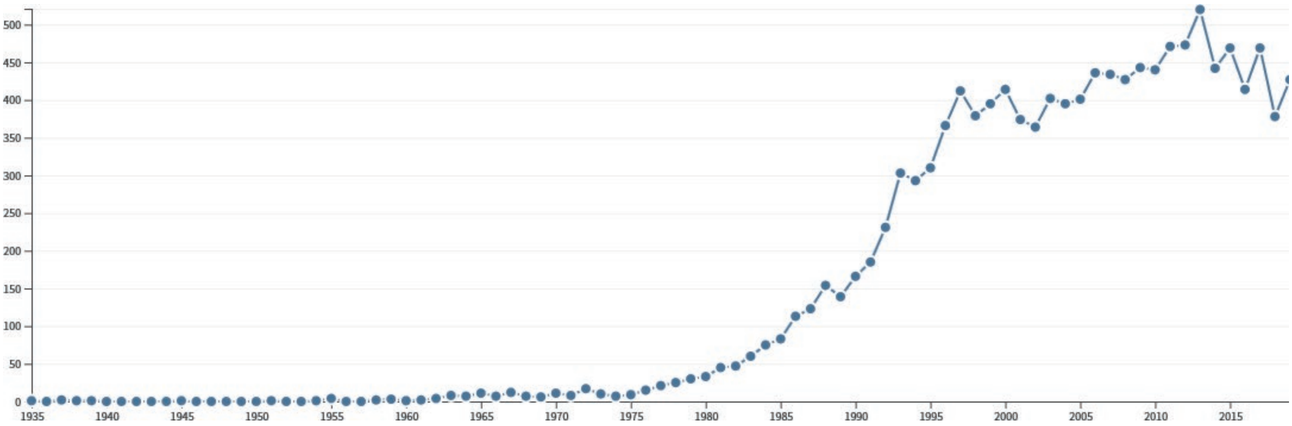


Figure 4. Number of citations per year 1935-2020 to the Møller-Plesset paper according to Web of Science.

to second and third order (MP2, MP3) had advantages over other methods and for small atoms and molecules agreed satisfactorily with experimental data. When Pople in 1998 gave his Nobel lecture in Stockholm, he praised the Møller-Plesset theory as an important step in the history of computational chemistry.[44]

The growth in visibility is illustrated by the number of citations (Google Scholar) to the Møller-Plesset paper in the six decades between 1962 and 2021:

1962-1971: *91.* 1972-1981: *222.* 1982-1991: *1070.*
1992-2001: *3570.* 2002-2011: *5330.* 2012-2021: *6100.*

By far most of the many citations to the 1934 paper are in journals devoted to chemical physics and quantum chemistry.

The paper by Møller and Plesset is a prime example of what in the sociology of science is known as a "sleeping beauty." This is a scientific paper whose relevance has not been recognized for a long time and then, more or less suddenly, becomes highly influential and cited.[45] Such sleeping beauties are of obvious interest from a historical and sociological point of view. Why were they initially ignored? Why did a sleeping beauty wake up at a particular, much later date?

A recent large-scale study of citation histories in all branches of natural and social sciences suggests that sleeping beauties are not exceptional and particularly not so in chemistry and physics.[46] The authors define a parameter called the "beauty coefficient" (B) which expresses the number of citations a paper has received and how long after publication it gained them. It is so constructed that a paper which accrues citations linearly over time has $B = 0$, whereas one which languishes for 100 years before rising to fame can have $B = 10{,}000$ or even more (for the full definition of B, see ref. 45). The study in question lists the top fifteen sleeping beauties in science since 1900, seven of which it classifies as chemistry and five as physics. One of them is the Møller-Plesset paper of 1934, for which $B = 2{,}584$ and the "awakening time" is found to be 1982. Another and much better known sleeping beauty, but with a beauty coefficient ($B = 2{,}258$) a little less than that of the Møller-Plesset paper, is the famous Einstein-Podolsky-Rosen (EPR) paper of 1935 on the completeness of quantum mechanics.[47]

8. NOBEL PERSPECTIVES

It is well known that a large number of Nobel chemistry prizes have been awarded to scientists who were either physicists or whose work would be normally classified as physics.[48] On the other hand, no Nobel Prize in physics has ever been awarded to a chemist. Consider as an early example the 1908 chemistry prize to Ernest Rutherford for his contributions to radioactivity including "the chemistry of radioactive substances." Bemused to have transformed so quickly from a physicist to a chemist, he wrote to Otto Hahn: "I must confess that it was very unexpected and I am very startled at my metamorphosis into a chemist."[49] Many years later the 1951 chemistry Nobel Prize was awarded to Glenn Seaborg and Edwin McMillan for their discoveries of the first transuranic elements. While Seaborg was trained in chemistry under G. N. Lewis, McMillan was a nuclear physicist, such as he pointed out in his Nobel lecture. "In spite of what the Nobel Prize Committee may think, I am not a chemist," he said.[50]

As mentioned, quantum chemistry was originally created by physicists rather than chemists and has to this day continued as an interdisciplinary field in which physicists play an important role. This is reflected in several of the more recent Nobel Prizes. Thus, one-third of the 2013 prize was awarded to Michael Levitt, an Israeli trained in physics and molecular biology but not in chemistry. When Levitt was informed about the honor, he reportedly said, "I never studied chemistry, actually I'm a physicist. But that's okay."[51] The Nobel Prize awarded to John Pople and Walter Kohn in 1998 was the first and so far only one explicitly motivated in quantum chemistry. While Pople was a quantum chemist (or perhaps a chemical physicist), Kohn's background was purely in theoretical physics. He wrote his doctoral dissertation under Julian Schwinger, one of the founders of modern quantum electrodynamics, and later changed to theoretical condensed matter physics. It was in this context that he developed the so-called density functional theory, a very successful approach to the many-particle problem which was widely considered an alternative to Møller-Plesset theory.[52]

Much like Møller and Plesset, Kohn was a theoretical physicist whose work unintendedly came to play a crucial role in quantum chemistry. Like Plesset had stayed at Bohr's institute in the 1930s, where he met Møller, so 27-year-old Kohn came on a fellowship to Copenhagen to work in the same institute, where Møller was appointed his supervisor. In a report of 1953, Bohr and Møller wrote that, "In all his work Dr. Kohn has proved himself a highly qualified theoretical physicist with great knowledge of a wide field of problems."[53] During Kohn's stay at the Bohr institute, which lasted from July 1951 to September 1952, he participated in a large conference on problems of quantum physics attended not only by Møller and Plesset, but also by Heisenberg, Bethe, Pauli, and other quantum luminaries. However, he did not enter a collaboration with Møller, whose research interests at the time were quite different from his. When Kohn developed his density functional theory in the mid-1960s, Møller and Plesset had almost forgotten about their earlier work and none of them showed any interest in Kohn's new approach or any other approaches to quantum chemistry.

9. CONCLUSION

The relations between physics and chemistry have changed significantly over time, often with physicists entering the field of chemistry with theories that most chemists found to be difficult and of no relevance to what chemistry is really about. The pioneers of quantum chemistry were theoretical physicists who had but little interest in traditional chemical problems and did not at all consider themselves to be chemists. The main result of the present study is a slight reevaluation of the standard history of early quantum chemistry, or at least a supplement to it. While this standard history covers in considerable detail Hartree's work and its extension to the Hartree-Fock theory, it has little to say about Fock's contribution and even less about the one of Møller and Plesset. The two contributions, the one from 1930 and the other from 1934, had in common that they were one-time mathematical investigations not originally related to chemical problems.

The Møller-Plesset theory exemplifies to some extent how quantum theorists acted as "chemists without knowing it" insofar that much later the theory came to be seen as an important contribution to computational chemistry. Because the Møller-Plesset theory was a "sleeping beauty" with very little initial impact, it is perhaps understandable that it does not figure in historical writings on the early period of quantum chemistry. After all, it would be anachronistic to let our knowledge of the theory's later development, say after the 1970s, influence the historical account of quantum chemistry in the 1930s. It is less understandable and harder to justify that the Møller-Plesset method has also been neglected in the writings on the more recent era, where extensions of this method have undeniably played a very significant role.

REFERENCES

[1] For an overview of the debate citing many references, see H. Chang in *Relocating the History of Science: Essays in Honor of Kostas Gavroglu* (Eds.: T. Arabatzis, J. Renn, A. Simões), Springer, New York, **2015**, pp. 193-210. See also H. Kragh, *Between the Earth and the Heavens: Historical Studies in the Physical Sciences*, World Scientific, London, **2021**, pp. 151-172, on which parts of the present paper rely.

[2] Quoted in B. Bensaude-Vincente, *Ber. Wissenschaftsgesch.* **2009**, *32*, 365-378.

[3] I. Kant, *Metaphysical Foundations of Natural Science* (Ed. M. Friedman), Cambridge University Press, Cambridge, **2004**, p. 6.

[4] J. S. Rowlinson, *Notes Rec. Roy. Soc.* **2003**, *57*, 35-45.

[5] F. Gregory, *Arch. Int. H. Sci.* **1984**, *34*, 108-123; A. M. Duncan, *Laws and Order in Eighteenth-Century Chemistry*, Clarendon Press, Oxford, **1996**.

[6] H. Kragh, *Centaurus* **2002**, *44*, 32-114, esp. pp. 60-69.

[7] H. C. Jones, *Elements of Physical Chemistry*, Macmillan, New York, **1902**; F. P. Venable, *The Study of the Atom*, American Chemical Society, Easton, PA, **1904**.

[8] Quoted in H. Kragh in *Hermann von Helmholtz and the Foundations of Nineteenth-Century Science* (Ed. D. Cahan), University of California Press, Berkeley, **1993**, pp. 401-431, on p. 429.

[9] J. J. van Laar, *Lehrbuch der mathematischen Chemie*, J. A. Barth, Leipzig, **1901**; H. A. M. Snelders, *Centaurus* **1986**, *29*, 53-71.

[10] H. Kragh, *Ambix* **1989**, *36*, 49-65. M. D. Gordin, *A Well-Ordered Thing: Dmitrii Mendeleev and the Shadow of the Periodic Table*, Basic Books, New York, **2004**, pp. 217-224.

[11] D. I. Mendeleev, *An Attempt Towards a Chemical Conception of the Ether*, Longmans, Green & Co., London, **1904**, p. 17.

[12] A. Smithells, *Proc. Brit. Assoc. Adv. Sci.* **1907**, 469-479, and *Nature* **1907**, *78*, 352-357.

[13] H. E. Armstrong, *Proc. Brit. Assoc. Adv. Sci.* **1909**, 420-454, on p. 477. See also A. Simões, K. Gavroglu in *Chemical Sciences in the 20th Century* (Ed. C. Reinhardt), Wiley-VCH, Weinheim, **2001**, pp. 51-74.

[14] H. Kragh, *Niels Bohr and the Quantum Atom: The Bohr Model of Atomic Structure 1913-1925*, Oxford University Press, Oxford, **2012**; H. Kragh, *Phys. Today* **2013**, *66* (5), 36-41.

[15] *Niels Bohr: Collected Works*, vol. 3 (Ed. J. Rud Nielsen), North-Holland, Amsterdam, **1976**, p. 240.

[16] M. Born, *Naturwissenschaften* **1920**, *8*, 373-382, on p. 382.

[17] O. Lodge, *Atoms and Rays: An Introduction to Modern Views on Atomic Structure and Radiation*, George H. Doran Co., New York, **1924**, p. 203.

[18] Quoted in *David Hilbert's Lectures on the Foundations of Physics 1915-1927* (Eds. T. Sauer, U. Majer), Springer, Dordrecht, **2009**, p. 209.

[19] R. C. Tolman, *J. Opt. Soc. Am. Rev. Sci.* **1922**, *6*, 211-228.

[20] A. N. Stranges, *Electrons and Valence: Development of the Theory, 1900-1925*, A&M University Press, College Station, TX, **1982**. See also H. Kragh, *Riv. Storia Scienza* **1985**, *2*, 463-486.

[21] E. N. de Andrade, *The Structure of the Atom*, G. Bell and Sons, London, **1923**, p. 239. See also T. Arabatzis, *Representing Electrons: A Biographical Approach to Theoretical Entities,* University of Chicago Press, Chicago, **2006**.

[22] *Quantum Chemistry: Classic Scientific Papers* (Ed. H. Hettema), World Scientific, London, **2000**, pp. 140-155. On the early development of quantum chemistry, see J. Mehra, H. Rechenberg, *The Historical Development of Quantum Theory*, vol. 6, Springer, New York, **2001**, pp. 521-571. A full history is given in K. Gavroglu, A. Simões, *Neither Physics nor Chemistry: A History of Quantum Chemistry*, MIT Press, Cambridge, MA, **2012**.

[23] Quoted in ref. 22 (Gavroglu, Simões), p. 100.

[24] Unpublished lecture of 6 April 1928 quoted in B. S. Park, *Brit. J. Hist. Sci.* **1999**, *32*, 21-46.

[25] T. Pynchon, *Introduction to Chemical Physics*, Van Nostrand, New York, **1874**.

[26] H. C. Urey, *J. Chem. Phys.* **1933**, *1*, 1-2. On this journal and the disciplinary boundaries between chemical physics and physical chemistry, see M. Jo Nye, *From Chemical Philosophy to Theoretical Chemistry: Dynamics of Matter and Dynamics of Disciplines 1800-1950*, University of California Press, Berkeley, **1993**, pp. 227-261, and J. W. Servos, *Physical Chemistry from Ostwald to Pauling: The Making of a Science in America*, Princeton University Press, Princeton, **1990**, pp. 251-298.

[27] A. Simões, K. Gavroglu, *Int. J. Quantum Chem.* **2014**, *114*, 116-127; B. S. Park, *Ann. Sci.* **2003**, *60*, 219-247.

[28] C. J. Ballhausen, *J. Chem. Educ.* **1979**, *56*, 357-361; C. J. Ballhausen, *Introduction to Ligand Field Theory*, McGraw-Hill, New York, **1962**.

[29] P. A. M. Dirac, *Proc. Roy. Soc. A* **1929**, *123*, 714-733, emphasis added. For the response of chemists to Dirac's claim, see A. Simões, *Phys. Perspect.* **2002**, *4*, 253-266.

[30] C. F. Fischer, *Douglas Rayner Hartree: His Life in Science and Computing*, World Scientific, New Jersey, **2003**; B. S. Park, *Hist. Stud. Nat. Sci.* **2009**, *39*, 32-62.

[31] J. Slater, *Phys. Rev.* **1928**, *32*, 339-348.

[32] V. Fock, *Z. Phys.* **1930**, *61*, 126-148.

[33] D. Cremer, *Comp. Mol. Sci.* **2011**, *1*, 510-530.

[34] Møller-Plesset theory is not mentioned in ref. 22 (Gavroglu, Simões) nor in other of the many works written on the history of quantum and computational chemistry.

[35] C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618-622. Received 14 July 1934 and published 1 October the same year.

[36] https://en.wikipedia.org/wiki/Christian_M%C3%B8ller

[37] H. Kragh, *Arch. Hist. Exact Sci.* **1992**, *43*, 299-328.

[38] N. Bohr to L. Page, 23 December 1934. Bohr Scientific Correspondence, Niels Bohr Archive, Copenhagen.

[39] See the obituary by T. Y. Wu in *Memorial Tributes: National Academy of Engineering*, vol. 6, National Academies Press, New York, **1993**, pp. 172-174.

[40] Interview with Plesset by C. Bugé of 8 December 1981, online as https://oralhistories.library.caltech.edu/127/

[41] R. Sinatra, *Nature Phys.* **2015**, *11*, 791-796; D. B. Baker, *Chem. Eng. News* **1981**, *59*, 29-34.

[42] Ref. 33.
[43] Ref. 22 (Gavroglu, Simões), pp. 224-229; L. Radom in *New Dictionary of Scientific Biography*, Scribner's, New York, **2008**, pp. 129-133, which includes a rare reference to the Møller-Plesset perturbation theory.
[44] J. S. Binkley, J. A. Pople, *Int. J. Quantum Chem.* **1975**, *9*, 229-236; J. A. Pople, J. S. Binkley, R. Seeger, *Int. J. Quantum Chem.* **1976**, *10*, 1-19; J. A. Pople, *Ang. Chem. Int. Edit.* **1999**, *38*, 1894-1902.
[45] A. van Raan, *Scientometrics* **2004**, *59*, 467-472.
[46] Q. Ke et al., *P. Natl. Acad. Sci. USA* **2015**, *112*, 7426-7431. https://doi.org/10.1073/pnas.1424329112
[47] A. Einstein, B. Podolsky, N. Rosen, *Phys. Rev.* **1935**, *47*, 777-780.
[48] For a list of chemistry prizes awarded to physicists 1908-1977, see H. Kragh, *Quantum Generations: A History of Physics in the Twentieth Century*, Princeton University Press, Princeton, **1999**, p. 432.
[49] Rutherford to Hahn, 29 November 1908, quoted in A. S. Eve, *Rutherford: Being the Life and Letters of the Rt. Hon. Lord Rutherford*, Cambridge University Press, Cambridge, **1939**, p. 183. See also H. Kragh, "Chemical and other aspects of Rutherford's nuclear atom," *J. Roy. Soc. New Zeal.* **51** (2021): 513-527.
[50] E. McMillan, in *Nobel Lectures, Chemistry 1942-1962*, Elsevier, Amsterdam, **1964**, pp. 314-322, on p. 318. Online: https://www.nobelprize.org/prizes/chemistry/1951/mcmillan/lecture/
[51] Tweet from Stanford University, 9 October 2013, see https://twitter.com/stanford/status/387913130673979392
[52] A. Zangwill, *Arch. Hist. Exact Sci.* **2014**, *68*, 775-848.
[53] Bohr to P. R. Wallace, 5 March 1953 (Møller Papers, Niels Bohr Archive, Copenhagen).

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

OPEN ACCESS

Citation: Kenndler E. (2022) Capillary Electrophoresis and its Basic Principles in Historical Retrospect. Part 4. Svante Arrhenius´ Electrolyte Dissociation. From 56 Theses (1884) to Theory (1887). *Substantia* 6(2): 55-77. doi: 10.36253/Substantia-1769

Received: Jun 06, 2022

Revised: Jul 20, 2022

Just Accepted Online: Jul 21, 2022

Published: September 1, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Capillary Electrophoresis and its Basic Principles in Historical Retrospect. Part 4. Svante Arrhenius´ Electrolyte Dissociation. From 56 Theses (1884) to Theory (1887)

Ernst Kenndler

Institute for Analytical Chemistry, Faculty of Chemistry, University of Vienna, Währigerstrasse 38, A 1090, Vienna, Austria
E-mail: ernst.kenndler@univie.ac.at

Abstract. Since the main interest of Svante Arrhenius, a student at Uppsala University, was the electrical conductivity of highly dilute electrolyte solutions, which had not yet been determined at the beginning of the 1880s, he decided to determine experimentally the molecular conductivities of aqueous solutions of about fifty electrolytes and their dependence on the dilution. In his dissertation, which he began in the winter of 1882/1883, he summarized his results and considerations in 56 "theses". He observed that strong acids had a high molecular conductivity, which increased only slightly with increasing dilution. Weak acids, in contrast, had low molecular conductivities, but these increased abruptly above a certain dilution. Arrhenius' innovative hypothesis was that electrolyte molecules are composed from two parts, "*an active (electrolytic) and an inactive (non-electrolytic) part*," with the proportion of the active part increasing with increasing dilution at the expense of the inactive part. Moreover, the electrically active part, which conducts electricity, was also the chemically active part. Arrhenius introduced the activity coefficient, later quoted as the degree of dissociation, which indicated the proportion of active molecules to the sum of active and inactive molecules. He tentatively related activity coefficient to molecular conductivity. He assumed that the higher the activity coefficients of different acids at the same equivalent concentrations, the stronger they are. Arrhenius tested his hypothesis taking the heat of neutralization of acids with a strong base measured by Thomsen and Berthelot. Strong acids developed the highest neutralization heats, i.e., the activation heat of water, since they consisted entirely of active H^+ and OH^- ions, which combined to inactive H_2O. Weak acids developed correspondingly less. The established parallelism between the molecular conductivities of acids and their heats of neutralization was the first proof of Arrhenius' hypothesis. He relied on thermochemistry and completed his dissertation. He presented his dissertation in June 1883 and published it in 1884 to obtain his doctorate. At that time, Wilhelm Ostwald was investigating the affinities of acids to bases, i.e. the intensity of the effects of acids on the rates of reactions they cause. He took the rate constants as a measure of the relative strength of the acids. After receiving Arrhenius' thesis, he measured the acid´s molecular conductivities and found a remarkable proportionality to the reaction rate constants of the hydrolysis of methyl acetate and the inversion of cane sugar caused by them. This was the second proof of Arrhenius' hypothesis, based on the results of chemical kinetics. A memoir presented in 1885 by J. H. van 't Hoff on the analogy between the osmotic pressure of a highly dilute solution separated from the pure solvent by a semipermeable membrane and the pressure of an ideal gas containing the same number of particles as the solution led to probably the most convincing proof of the Arrhenius hypothesis. This analogy corresponded to Avogadro's well-known law, which is $PV=RT$. He found that the pressure for non-conductors such as glucose followed this law, but was higher for electrolytes. This deviation was accounted for by the van 't Hoff factor i, which indicates into how many particles the solute – at least partially – has dissociated, so that the modified law is $PV=iRT$. The factor i could be deduced from Raoult's freezing point depression, and could also be calculated using Arrhenius' degree of dissociation α. The degree of dissociation, in turn, was determined from the ratio of the conductivity of a dilute electrolyte solution and that under limiting conditions. The agreement found between the factors i determined by the two independent methods was the third proof of the Arrhenius hypothesis. There was a fourth proof, namely the additivity of physical properties. With these four nonelectrical and independent proofs, the 56 theses of Arrhenius' dissertation became the groundbreaking theory of *dissociation of substances dissolved in water*, which he published in 1887. In 1903 the Nobel Prize in Chemistry was awarded to him "*in recognition of the extraordinary services he has rendered to the advancement of chemistry by his electrolytic theory of dissociation*".

Keywords: Dissociation Theory, Arrhenius, Electrolyte, Solution, Activity Coefficient.

INTRODUCTION

The 1880s, to which this article refers, were remarkable years, for in that decade the prevailing view that the molecules of electrolytes consist in their solutions of oppositely charged ions which are tightly bound together and can be separated only by an electric force acting on them was replaced by a radically different theory. This conventional view dated back nearly eight decades to the time of Theodor von Grotthuß and Humphry Davy. This view was challenged but by no means generally replaced by the free-ion hypothesis of Rudolf Clausius in 1857.[1-3] We have described its history in our earlier articles[4, 5] and do not repeat it here. Instead, we will focus on the pioneering theory of Svante Arrhenius.

Arrhenius' scientific interest was in the conductivity of highly dilute electrolyte solutions. This was the subject of his dissertation, which ultimately consisted of a collection of 56 theses or propositions. He presented his dissertation to the Royal Swedish Academy of Sciences in June 6, 1883, and published it in 1884.[6, 7] After a break of about three years, he published his theory of electrolyte dissociation in 1887.[8, 9]

Remarkably, Arrhenius managed to form his hypotheses into a solid theory by relying on results obtained by other scientists in the fields of non-electrical phenomena in physics and chemistry. In this article, the chronology of this successive confirmation of his theory is traced.

The historical overview, starting from a collection of 56 theses of a dissertation up to the groundbreaking theory "*Ueber die Dissociation der in Wasser gelösten Stoffe*"[8] (On the dissociation of substances dissolved in water) is the subject of the present article.

SVANTE ARRHENIUS: BIOGRAPHY

Svante August Arrhenius, born in 1859 in Vik (also Wik or Wijk) near Uppsala, Swedish Kingdom, attended the cathedral school in Uppsala, matriculated in chemistry and mathematics at Uppsala University in 1876, graduating as early as 1878. His photo from that year is shown in Figure 1.

He then began studies of chemistry with lectures in mathematics and laboratory work in autumn of 1878 at the Swedish chemist, mineralogist and oceanographer Per Teodor Cleve, professor of chemistry from 1874 to 1905 at the University of Uppsala. Arrhenius, however, realized that his interests lay more in the overlapping area between physics and chemistry. Since the opportunities to work on such topics were not favorable at Uppsala University, he changed to the Physical Institute of the Swedish Academy of Sciences in Stockholm in 1881,

Figure 1. Photo of Svante Arrhenius from 1878, the year when he began his studies at the University in Uppsala. Provided by courtesy of the archives at the Royal Swedish Academy of Sciences.

where Erik Edlund (1819 – 1888), professor of physics, agreed to supervise him for his dissertation on the condition that he works independently and deals with topics of his own choosing.

Edlund had a very great diversity of interests from which Arrhenius profited considerably. Edlund explored, for example, electric sparks,[10, 11] atmospheric electricity and aurora borealis,[12] streaming potentials and dealt with theories of electrical phenomena.[13, 14]

After Arrhenius had initially tried to find a method for determining the molecular weight of chemical compounds in solution – albeit unsuccessfully and without knowledge of Raoult's paper from 1882[15] – a topic he had begun at Uppsala, and which he continued in Edlund´s laboratory, he turned his interest to the electric conductivity of dilute electrolyte solutions.

Arrhenius´ first publication

Most readers probably do not know that Arrhenius' very first publication was not his dissertation, but his own article published shortly before. In the spring of 1882 he had the opportunity to carry out independently in Edlund's physical laboratory a study of the polarization of electrodes, an undesirable effect of great importance in determining the resistances of electrolyte solutions. He measured the time course of the decrease in polarization current at the electrodes after the direct current had been switched off. From the results he derived some general conclusions which, however, need not be further discussed in the present context. Arrhenius paper "*Undersökning med Rheotom* övfer *den Galvaniska Polarisationens Försvinnande i ett Polarisationskärl, hvars Plattor* äro *förbunda genom en metallisk Ledning*"[16] (Examination of the disappearance of galvanic polarization in a polarization vessel with a rheotome, the plates of which are connected by a metallic conductor) was dated October 11, 1882. It is probable that Arrhenius did not choose the subject of this study by chance, for knowledge of the operation of the device was of advantage for the measurements of the conductivities of dilute electrolyte solutions in his dissertation, which he began in the winter of 1882/1883, only a few weeks after the publication of this first paper.

THE DISSERTATION: RECHERCHES SUR LA CONDUCTIBILITÉ GALVANIQUE DES ÉLECTROLYTES

When in the winter of 1882/1883 Arrhenius began with a dissertation studying highly diluted solutions, he was faced with the unpleasant situation that only a few publications had appeared on the subject up to that time, mainly those published by Friedrich Kohlrausch[17-25] and Rudolf Lenz.[26-29] Although Kohlrausch and Lenz had published reliable values for conductivities in recent years, they were of little help to Arrhenius. Kohlrausch reported comprehensive studies of the conductivities of a large number of dilute solutions from 1874 to 1879,[17-24, 30, 31] but their electrolyte concentrations were too high for Arrhenius' objective. They were usually greater than about 5% by weight and reached for some solutes nearly 100%. In 1878 Lenz did investigate solutions with lower molecular electrolyte concentrations in the range of 10^{-2} equ.L^{-1}, but his work was limited to the small group of potassium, sodium and ammonium salts and hydrogen acids.[28] Moreover, the concentration ranges were still too high for Arrhenius' purposes.

Arrhenius emphasized on pages 3 and 4 of his dissertation "*Recherches sur la conductibilité galvanique des* électrolytes"[6, 7] – the dissertation was written in French – that Kohlrausch had announced several times a work with conductivities in highly dilute solutions. But since this had not yet appeared – Kohlrausch´s article "*Ueber das Leitungsvermögen einiger Electrolyte in* äusserst *verdünnter wässeriger Lösung*" (On the conductivity of some electrolytes in extremely dilute aqueous solution) was not published until 1885[32] – Arrhenius decided to begin the investigation he had planned even without knowing Kohlrausch's results.

Arrhenius carried out the experimental part of his dissertation in Edlund's laboratory in Stockholm in the winter of 1882/1883 and spring of 1883. In May 1883 he wrote the second, theoretical part of his dissertation at his parents' home in Uppsala. Initially Arrhenius wrote his dissertation in one part, but it was rather complicated to read and somewhat confusingly structured. Sven Otto Pettersson, then docent of physics in Uppsala, who was asked by Arrhenius to read the dissertation in advance, found the work worthy of submission, but advised Arrhenius to restructure it; so Arrhenius divided it into two parts. He also recommended to avoid the term *dissociation*, since it was understood to mean the decomposition of a compound into its smaller constituents under the influence of heat.

According to Pettersson´s advice Arrhenius divided the initial version into Part 1[6] which consisted from 63 pages, and into Part 2[7] with 89 pages.

It may be confusing that Arrhenius did never use the term *dissociation* in his dissertation. *Dissociation* for thermal decomposition was introduced in 1857 by the French chemist Henri Étienne Sainte-Claire Deville (1818 – 1881) in his paper "*Sur la dissociation ou la décomposition spontanée des corps sous l'influence de la* *chaleur*" (On the spontaneous dissociation or decomposition of bodies under the influence of heat).[33] Examples are the decomposition of ammonium chloride on heating into *HCl* and NH_3, or of molecular iodine J_2 into its atoms, and the reversible recombination of the gaseous particles to the parent molecules on cooling. Pettersson foresaw that the term dissociation for the cleavage of electrolyte molecules in solutions would meet with disapproval from the chemistry professors at the university. Therefore, Arrhenius used in his dissertation the terms electrically *active* and *inactive* for the ionic and the non-ionic particles.

We emphasize that we rely on the two-part dissertation submitted to the Royal Swedish Academy of Sciences in June 1883 and published in 1884.[6, 7]

PREMIÈRE PARTIE. LA CONDUCTIBILITÉ DES SOLUTIONS AQUEUSES EXTREMEMENT DILUÉES DÉTERMINÉE AU MOYEN DU DÉPOLARISATEUR

In Part 1, "The Conductivity of Extremely Dilute Aqueous Solutions Determined by Means of the Depolarizer",[6] Arrhenius first described in detail the apparatus and the devices, the experimental conditions and the chemicals.[1] He paid special attention to the control of the correct function of the "*depolarisateur*" Edlund had constructed,[34] and by which he could transform constant into sinusoidal alternating current.[2] It is schematically depicted in Fig. 3 of Arrhenius´ dissertation[6], and explained there and in Edlund´s paper from 1875.[34]

The measuring cell which Arrhenius used is shown in Figure 2. It served for both, the preparation of the different dilute samples and the measurement of their resistances in a rather unconventional, but economical manner.

The measurement procedure was as follows. First, a weighed amount of electrolyte was placed in the glass vessel where it was dissolved in 35 cm^3 distilled water. The quantity of water and therefore the concentration of the electrolyte were determined by weighing the vessel before and after filling. After measuring the resistance, part of the solution was removed and replaced by distilled water, and the two liquids were mixed directly in the vessel. The dilution was determined again by weighing the vessel after replacing a part of the solution and after refilling with solvent. This procedure was carried out several times in succession, preparing solutions with three to six different concentrations, usually four to five. Solutions were diluted till the measured resistance approached that of pure water (in fact, the lowest concentration Arrhenius measured was in the range of about 10^{-4} normal.[3])

The resistances of the dilute electrolyte solutions were between several hundred and about 120000 Ohm. The reproducibility was of the order of 1% or less. However, the chosen procedure had the disadvantage that the concentrations were not determined by quantitative analysis and were therefore of low accuracy, which consequently affected the molecular conductivities, a systematic error that Arrhenius readily accepted.[4]

Arrhenius measured the electrical resistances of 47 compounds, i.e. salts, bases, and acids at various dilutions[5], and from these their molecular conductivities[6]. He found that the conductivity in dilute solutions

---

[1] We give here a more detailed description of his experimental conditions, because Arrhenius performed his experiments in a very contrarian manner compared to Kohlrausch, who measured the conductivities in about the same years. In part 3 of our series, we described how much importance Kohlrausch attached to measuring values that were as accurate as possible. He then derived empirical laws from them. Arrhenius, on the other hand, as he himself said, was satisfied with less accurate values. He was, however, able to recognize in them an existing tendency from which he deduced his hypotheses.

[2] Edlund's depolarizer, incidentally, was severely criticized by E. Dorn,[35] [36]. but his criticism was rejected by Edlund.[37]. It is, however, remarkable that even in 1886, two years after the completion of Arrhenius´ dissertation the problem of measuring accurate conductivities with alternating current was still a matter of question, at least for Oliver Lodge, to whom we will return below. He added a critical comment in ref. [38], footnote 1 on p. 384 in "*Translation of a letter received from Dr. Arrhenius respecting the above Criticism*" which read: "*This opens a large question, viz., how far it is advisable to depend on the use of alternating currents as a device for avoiding polarisation difficulties …. Unless the question of electro-chemical capacity be thus considered, and either eliminated by calculation or proved to be negligible by experiment, the presumed advantage of alternating currents in dealing with electrolytic resistance is illusory*." Nowadays, *depolarizer* is the obsolete synonym for an electroactive substance (see *IUPAC,* https://doi.org/10.1351/goldbook.D01599, *Compendium of Chemical Terminology, 2nd ed. (the "Gold Book")*.

[3] Arrhenius often used the term "normal" as a measure of the concentration. It was introduced by Karl Friedrich Mohr in 1855 in his textbook about titrimetric analysis.[39]

[4] Arrhenius pointed out this low accuracy in ***§ 13.*** *Conductibilité moléculaire*. p. 38. "From formula (3), the molecular conductivity can be calculated in each special case. The named formula contains the amount of dissolved electrolyte (P). It was not possible to determine this quantity analytically, but only by weighing the electrolytes. As the substances examined are hygroscopic and also somewhat impure, this determination cannot possess great accuracy." In his reply from 1886 to Oliver Lodge´s criticism on his low accuracy (ref. [38], p. 386), he wrote: "6. That the last work of Kohlrausch contains as you say incomparably better experimental data (especially more accurate) is true enough. But without my data I could not have formed a coherent picture of the whole."

[5] The apparent paradox that the conductivity of an electrolyte solution decreases with increasing dilution, but its molecular conductivity increases, has already been clarified in our previous article[5] in connection with Kohlrausch´s law of independent ion migration.

[6] The molecular conductivity k/m is the ratio of the specific conductivity, k, and m, the concentration of the solution in g.equiv. per liter. It is the equivalent conductivity, as Rudolf Lenz emphasized in his study of

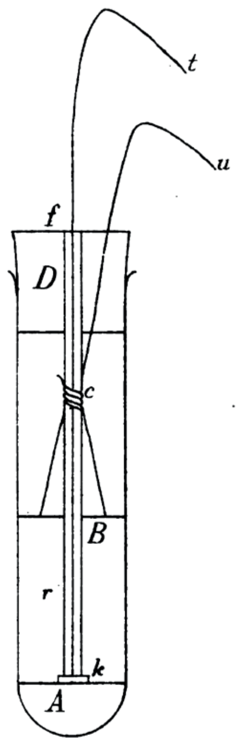


Fig. 4.

Figure 2. Scheme of the cell Arrhenius used for the measurement of the resistance. Glass vessel, height, ca. 150 mm; diameter, 25 mm. *A*, *B*: platinized platinum electrodes, 2/7 mm thickness. Lower electrode *A* welded to a thick platinum wire, fused into a narrow glass tube *K*, which passed through the center of the conical cork or rubber plug *D*, attached to tube *r* at the top at *f*. Upper electrode *B*: perforated in center to allow tube *r* to pass through. Two long platinum wires, riveted to plate *B*, attached to the tube at point *c*. One of these wires formed with *t* one of the poles. *k*, small rubber plate to hermetically seal vessel at the bottom. The test liquid (30 to 40 cm^3) filled the vessel to a point slightly above *B*. Reproduced from Fig. 4 of Arrhenius´ dissertation.[6]

for most electrolytes is proportional to the number of electrolyte molecules. If this is not the case, it must be assumed that chemical reactions take place upon dilution, but he had no conclusive explanation at first. Yet it was these effects that attracted his particular interest in his work.

Arrhenius introduced a quantity he called "dilution exponent" and measured its dependence on the dilution for a given electrolyte. The dilution exponent was the ratio by which the resistance of a salt solution increases when it is diluted by water to twice its volume. In the best case, the exponent is 2.[7]

In § 13. he calculated the molecular conductivities, actually the conductivity at an electrolyte concentration in equivalent per liter of solution, as function of the concentration in the series of dilutions prepared from of one and the same electrolyte solution. He found from his own data, and from those he has taken from publications of Kohlrausch[23, 24] und Lenz,[28] that the molecular conductivity increases with increasing dilution for all 47 electrolytes he investigated (in the summary of Part 1 he mentioned only 45 electrolytes). The electrolytes were salts, and only five bases and acids[8] each. Their molecular conductivity approached a certain limiting value; the same result found Kohlrausch.[23] For most electrolytes, especially for simple ones like the alkali halogenides, molecular conductivity increased slightly and linearly with increasing dilution. However, some solutes, especially ammonia and acetic acid, exhibited deviating properties. Their conductivity remained at low values at lower dilution, but abruptly increased when dilution increased. We have discussed these effects already in Part 3 of our series;[5] see the lower part of Figure 5 there. The complete plot was published by Kohlrausch in 1885.[32]

To find an explanation for this deviating shape of the curves, Arrhenius resorted to Hittorf's observation of the formation of complex salts of cadmium and iodine.[40] He stated on p. 60 in *Thesis 11.* (the entire dissertation comprised 23 paragraphs and 56 theses, in French "*propositions*"): "Aqueous solutions of all electrolytes contain the dissolved electrolyte at least in the form of molecular complexes", and attributed the observed irregularities to possible chemical decomposition of complex molecules with increasing dilution.

He postulated in *Thesis 2*: "If two or more salts are dissolved in the same non-conducting solvent, the conductivity of the solution is equal to the sum of the conductivities which the solution would possess if only one salt were dissolved at one time and the other salt alone at another time." In § 16. *Chemical action*, he formulated the following important two theses:

> *Thesis 4.* If, when diluting any solution, the conductivity does not change in proportion to the amount of electro-

---

haloïd compounds in 1877.[26]

[7] To be precise: As Arrhenius measured and tabulated resistances, he defined the dilution exponent as the ratio by which the resistance increases when a given electrolyte solution of the series is diluted. It is obvious that this indicates the corresponding decrease in conductivity. If the solution is diluted exactly to twice its volume, the dilution exponent should be 2, because the number of the initial conducting molecules is exactly bisected. Arrhenius found that most electrolytes had values of 1.95 to 2, but some deviated with higher, others with lower values.

[8] Arrhenius measured the resistances of hydrochloric, nitric, sulphuric, phosphoric and boric acid.

> lyte, a chemical change must have occurred in the solution due to the addition of the solvent.
> *Thesis 5.* If two substances are dissolved in water at the same time and thesis 2. is not applicable, a chemical process must have taken place between the two.

In § 17. he discussed the peculiarities of hydrates,[9] and in § 20. the nature of the resistance of electrolytes. He attributed the large conductivity of acids to the fact that they have hydrogen as a cation which, because of its small molecular volume, penetrates most easily into the solvent.

In the last chapter, § 21: *Properties of solutions of normal salts*, he makes a whole series of conjectures and speculations. The most important hypotheses, however, relate again to the decay of complex molecules, such as those of Hittorf quoted above. He argued in *Theses 11*, *12* and *13* that complexes probably decay upon the addition of solvents, so that the complexity approaches asymptotically a lower limit. From this follows the important hypothesis that the more completely complexes decay, the more molecular conductivity increases. Important for the further development of his theory is his conviction (p. 60, 61)

> *Thesis 11.* Aqueous solutions of all electrolytes contain the dissolved electrolyte, at least in part, in the form of molecular complexes.
> *Thesis 12.* The limit, to which the complexity of a normal salt dissolved at extreme dilution tends to approach, is of the same degree for all normal salts. (Probably this thesis applies for all electrolytes.)
> *Thesis 13.* It is likely that this limit is reached only when the salts are divided into simple molecules as represented by the chemical molecular formula.

From today's perspective, this could be seen as a pre-stage to his dissociation theory, as it transfers the idea of the decay of molecular complexes into individual molecules to the decay of molecules into their ions. However, the remaining problem was why the molecular conductivities of most electrolytes regularly increase with increasing dilution, and why some of them increase in deviating manner.

SECONDE PARTIE. THÉORIE CHIMIQUE DES ÉLECTROLYTES

Part 2, "Chemical Theory of Electrolytes"[7] of Arrhenius´ dissertation began in §1 with his considerations of "Ammonia as Electrolyte". It was well known by him that compounds such as liquid, pure and anhydrous ammonia – its boiling point is -33°C -do not transport electricity, but they become conductors when water is added. In Volume II. p. 147, of Marcellin Berthelot´s[10] *Mécanique chimique,* § 2. *DissoIution des gaz*,[51] he found

> *Par exemple, la solution du gaz ammoniac, saturée à basse température se trouve contenir l'eau et le gaz suivant des proportions définies: soit* $AzH^3+H^2O^2$*, très sensiblement. Cet hydrate cristallise d'ailleurs dans un mélange refrigerant.* (For example, the solution of ammonia gas, saturated at low temperature, contains water and gas in defined proportions: i.e. $AzH^3 + H^2O^2$, very substantially. This hydrate crystallizes in a refrigerant mixture.) [Symbol *Az* stands for French *azote*, i.e., nitrogen. Author´s note].

Since, according to the prevailing hydrate theory, hydrates were generally assumed to be dissolved in the same composition in which they crystallize from their solution, Arrhenius thus had every reason to believe that ammonia hydrates exist at least partially in solution as AzH_40H. Berthelot further wrote

> *Mais tous ces composés sont peu stables, et susceptibles de subsister seulement en présence des produits de leur décomposition; c'est-à-dire que le gaz dissous [*H_3Az*], le liquide dissolvant et leur combinaison forment un système en* équilibre, équilibre *dont les conditions varient avec la température et la pression.* (But all these compounds are not very stable, and likely to remain only in the presence of the products of their decomposition; that is to say, the dissolved gas, the dissolving liquid and their combination form a system in equilibrium, an equilibrium whose conditions vary with temperature and pressure.)

This means that the equilibrium $NH_3+H_2O = NH_4OH$ is established. From these three components water and ammonia are non-electrolytes, only the weak

---

[9] By hydrates Arrhenius meant acids and bases, using the prevailing term for them. The theory of hydrates of compounds, not necessarily only of electrolytes, will be the motive of a vigorous rejection of Arrhenius' theory of dissociation of electrolytes right after its publication. This earlier hydrate theory should not be confused with the modern theory of solvation or hydration of charged particles. This interesting chapter on this controversy, usually not even mentioned in textbooks and therefore widely unknown, will be covered in a future part of our review series.

[10] Marcellin Pierre Eugène Berthelot (1827 – 1907) was the most influential and respected chemist in France in his time. He was famous for his innovations in many scientific areas, among others in organic synthesis,[41],[42][43] and in biochemistry and pharmacology. He introduced thermochemistry, where he coined the terms exothermic and endothermic, before thermodynamics took over the dominant role. His standard books about thermochemistry (e.g. refs. [44],[45],[46]) with numerous data were used by Arrhenius to support his theory. Since midst 1880 Berthelot´s interest turned to the history and philosophy of sciences,[47] in particular to alchemy.[48],[49],[50]

base NH_4OH has the capability to contribute to electric conductivity by splitting off the cation NH_4^+ and the anion OH^-. According to the law of mass action, the equilibrium shifts towards NH_4OH when water is added, which is reflected in *Thesis 14.* "The conductivity of an ammonia solution is caused by a small amount of AzH_4OH, which is contained in it and which increases with the dilution of the solution."

Kohlrausch observed the same effect upon dilution of acetic acid we have already mentioned above, and an analogue relation between conductivity and dilution with other acids he had investigated.[21] Based on these facts Arrhenius came to the main conclusion that the amount of the conducting part of the electrolyte increases with dilution at the expense of the non-conducting part, and formulated one of the most important theses of his dissertation, which turned out to be the nucleus of his later dissociation theory. Literally it reads (and stated that *Thesis 15.* also applies to bases).

> *Thesis 15. La solution aqueuse d'un hydrate quelconque se compose, hors l'eau, de deux parties, l'une actife (*électrolytique*), l'autre inactive (non-*électrolytique*). Ces trois parties constituantes, l´eau, l' hydrate actif, et l' hydrate inactif, forment un* équilibre *chimique, tel qu´à une dilution la partie active augmente et la partie inactive diminue. (L'actlvité* électrolytique *se confond avec l'activité chimique).* (The aqueous solution of any hydrate is composed, apart from water, of two parts, one active (electrolytic), the other inactive (non-electrolytic). These three constituent parts, water, active hydrate, and inactive hydrate, form a chemical equilibrium, such that on dilution the active part increases and the inactive part decreases.) (The electrolytic activity is equal to the chemical activity).

Then Arrhenius introduced the term "*activity coefficient*"[11] and defined it as

> *Le coëfficient d'activité d'un électrolyte est le nombre exprimant le rapport du nombre d´iones qu'il y a reéllement dans l'électrolyte, au nombre d´iones qui y seraient renfermés, si l´électrolyte était totalement transformé en molécules electrolytiques simples. (Ces molécules sont nécessairement constituées d'une manière analogue à celle des sels.)* (The activity coefficient of an electrolyte is the number that expresses the ratio of the number of ions that are present in the electrolyte to the number of ions that would be present if the electrolyte were totally transformed into simple electrolytic molecules. (These molecules are necessarily constituted in a manner similar to that of salts.)

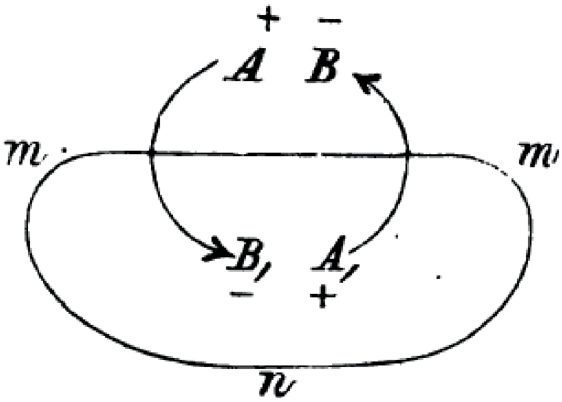


Figure 3. Schematic drawing of the flow of the circular current. For explanation, see main text.

In § 3. *Hypothesis of Clausius and Williamson*, Arrhenius addressed the question of the number of free dissociated ions in a dilute electrolyte solution, resorting to the free-ion hypothesis of Clausius. Arrhenius tried to draw conclusions from this hypothesis, which are exemplified in Figure 3. Although this figure was used for a different conclusion, it is fully suitable for the present one.

In this figure, two molecules of the same electrolyte are shown (we negate the curve labeled *mm,n*). The upper molecule consists of cation *A* and anion *B*, the other one of A_1 and B_1. All ions are charged with the same amount of electricity. Cation A_1, for example, can split off and combine with anion *B* of the next molecule. As a result, cation *A* is released and can unite with anion B_1. In solutions, of course, not only these two molecules, but any larger number A_2B_2, A_3B_3, ... A_nB_n … are present, so that this process extends over several molecules. However, it will not end until the cation of the last molecule has united with the anion of the first.

Though Arrhenius believed that this process was very fast, there were still free ions existing for short time during this exchange. These have moved in a closed loop, as well as the electricity charged with them. This transport of electricity was called by Arrhenius a *circular current*. It is the natural state of electrolyte ions in a solution – mind you – without an applied electric potential, that is, under the condition that no current flows through the solution. This means that circulating currents occur permanently in an electrolyte solution.

Because of its connection with the circular flows just discussed, § 6 shall now be treated prior to § 4. However, before continuing with this paragraph, we point out that this concept of the circular current was already introduced by Grotthuß as early as 1819 who called it *infinite circular molecular exchange*, but was hardly taken note of.

---

[11] It is equal to the *dissociation ratio* Oliver Lodge defined in 1885 on p. 756 of ref. [52] for dissociated molecules in solutions.

Although an excerpt of Grotthuß´ work appeared in Gilbert's *Annalen der Physik,*[53] his entire work was published in the scarcely accessible *Annalen der Curländischen Gesellschaft für Litteratur und Kunst.*

Grotthuß wrote in his article in reference to Figure 5, shown below as Figure 4

> *"In den Flüssigkeiten, die aus heterogenen Elementartheilchen bestehen, ..., muss zwischen diesen Elementartheilchen ein beständiger Galvanismus, und dadurch ein beständiger wechselseitiger polarischer Molekularaustausch unterhalten werden, den man durch das auf Taf. III in Fig. 5 dargestellte kreisförmige Schema ausdrücken kann. Jede Wasserzersetzung, die man mit dem Namen chemische belegt, ist daher nur eine Störunung des fortwährenden Molekulargalvanismuis, oder eine Ausgleichung des unendlichen kreisförmigen Molekularaustausches zu einem endlichen linienförmigen.*" (In liquids consisting of heterogeneous elementary particles …. a constant galvanism must be maintained between these elementary particles, and thus a constant mutual polaric molecular exchange, which can be expressed by the circular scheme shown on Plate III in Fig. 5. Any water decomposition, which is called chemical, is therefore only a disturbance of the continuous molecular galvanism, or a leveling of the infinite circular molecular exchange to a finite linear one.)

After this retrospect to the unexpectedly early concept of the *infinite circular molecular exchange* by Grotthuß, which resembles also Williamson's 1850 hypothesis[54] on the permanent exchange of "radicles" between molecules by double decomposition we continue with § 6. It addressed Arrhenius' conclusion about the consequences of this current when a second electrolyte, *CD*, is dissolved in the solution instead of a single electrolyte, e.g., *AB* as described above.

§ 6. *The double decomposition*. If *AB* and *CD* move through the inactive solvent by the circular current, again without being driven by an applied electric force, consequently the pairs *AD* and *CB* also form from *AB* and *CD* by double decomposition. Thus, all four electrolytes are present in the solution at the same time. Their amounts depend on the respective activity coefficients of the electrolytes (the present degree of dissociation). If these are equal, the amounts of all four substances are also equal. If initially one equivalent each of *AB* and *CD* are dissolved, half an equivalent of each of the four electrolytes *AB, CD, AD* and *CB* will be found at equilibrium. For unequal affinity coefficients, a corresponding equilibrium is reached by double decomposition, and *Thesis 29.* applies: "Every salt dissolved in water partially decomposes into acid and base. The amount of these decomposition products is the more substantial the weaker the acid and the base are and the greater the amount of water is." This thesis leads inevitably to the chief question of the strength of acids and bases.

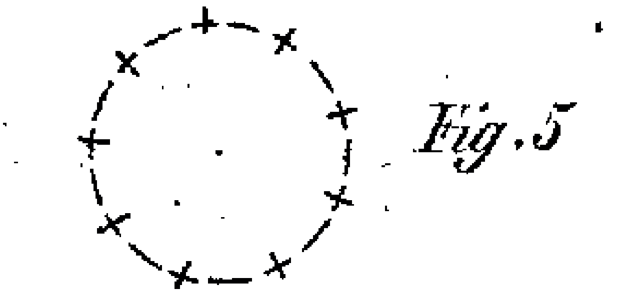


Figure 4. Circular current in a liquid which consists of positively and negatively charged particles. Grotthuß did not call them ions in his text, because this term was unknown in his time. Note that no electric potential is applied, and therefore no current flows through the liquid. Taken from ref. [53].

Kohlrausch differentiated two groups of acids (the same applied for bases). Strong acids show a linear dependence of the molecular conductivity on the number of dissolved molecules, weak acids do not. For both groups some examples of the molecular conductivity, λ, are listed in Table 1.

It should be recalled that Arrhenius had previously equated the molecular conductivity with the activity coefficient. Considering the molecular conductivities of the acids and bases from Table 1, he formulated *Thesis 33.* "The greater the activity coefficient (molecular conductivity) of an acid, the stronger it is. This thesis also applies to bases".

With this thesis he indirectly formulated his acid-base theory. It stated that acids are compounds that dissociate into H^+ ions and negatively charged anions in an aqueous solution. Analogously, bases dissociate in water to OH^- ions and to cations. Remarkably, he did not define it literally at any point in his dissertation, but it can be clearly understood as such because of the multiple mentions of the active parts of acids and bases.

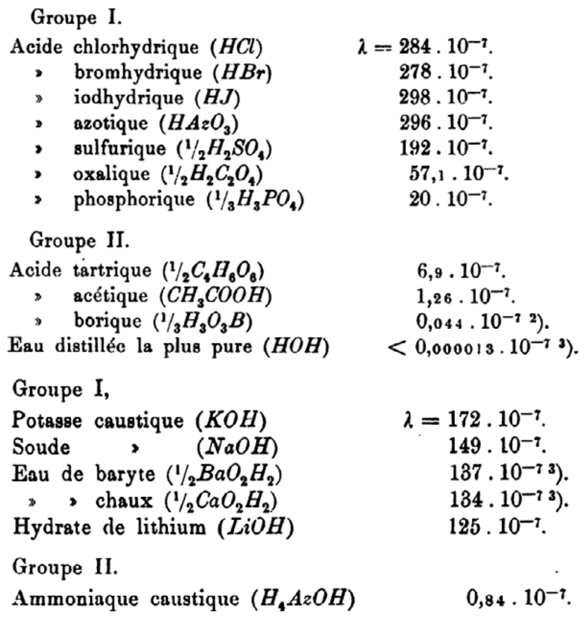
The values of the molecular conductivities in Table 1 indicate that the hydrohalic and nitric are strong acids, acetic acid is weak, boric acid is the weakest. In the group of the bases, sodium and potassium hydroxide are strong, ammonium hydroxide is a weak base.[12]

Arrhenius, however, was not satisfied with the fact that the ranking of the strength was based only on the conductivity, and wanted to evaluate it by an independent method which was not based on electric

---

[12] Water is a special case, it has the lowest conductivity of all compounds; it is listed in the group of acids, but it as was generally known that it can also act as a base.

Table 1. Molecular conductivities, λ, of strong (Group I) and weak (Group II) acids and bases. Concentrations of electrolyte solutions: 1 g equiv. solute per liter solution. The conductivities were taken or calculated from ref. [21], that of distilled water see Première partie, § 8.[6] The table is reproduced from Arrhenius´ dissertation, Seconde partie,[7] p. 14 and 15.



Groupe I.	
Acide chlorhydrique (HCl)	$\lambda = 284 . 10^{-7}$.
» bromhydrique (HBr)	$278 . 10^{-7}$.
» iodhydrique (HJ)	$298 . 10^{-7}$.
» azotique ($HAzO_3$)	$296 . 10^{-7}$.
» sulfurique ($^1/_2H_2SO_4$)	$192 . 10^{-7}$.
» oxalique ($^1/_2H_2C_2O_4$)	$57{,}1 . 10^{-7}$.
» phosphorique ($^1/_3H_3PO_4$)	$20 . 10^{-7}$.
Groupe II.	
Acide tartrique ($^1/_2C_4H_6O_6$)	$6{,}9 . 10^{-7}$.
» acétique (CH_3COOH)	$1{,}26 . 10^{-7}$.
» borique ($^1/_3H_3O_3B$)	$0{,}044 . 10^{-7}$ 2).
Eau distillée la plus pure (HOH)	$< 0{,}000013 . 10^{-7}$ 3).
Groupe I,	
Potasse caustique (KOH)	$\lambda = 172 . 10^{-7}$.
Soude » ($NaOH$)	$149 . 10^{-7}$.
Eau de baryte ($^1/_2BaO_2H_2$)	$137 . 10^{-7}$ 3).
» » chaux ($^1/_2CaO_2H_2$)	$134 . 10^{-7}$ 3).
Hydrate de lithium ($LiOH$)	$125 . 10^{-7}$.
Groupe II.	
Ammoniaque caustique (H_4AzOH)	$0{,}84 . 10^{-7}$.

properties. In his Nobel Lecture, held December 11, 1903 on "*Development of the theory of electrolytic dissociation*"[55] he admitted that he himself felt that a single line of evidence was not sufficient for a well-founded theory. He said

> "If this concept had only been applicable to accounting for the phenomenon of electrical conductivity, its value would not have been particularly great."
> "An examination of the numerical values adduced by Kohlrausch and others for the electrical conductivity of acids and bases as compared with Berthelot's and Thomson's measurements of their relative strengths in terms of their chemical effect showed me that the acids and bases with the greatest conductivity are also the strongest. I was thus led to the assumption that the electrically active molecules are also chemically active, and that conversely the electrically inactive molecules are also chemically inactive, relatively speaking at least. *etc.*"

For the sake of completeness, we think it appropriate to include a chapter, albeit brief, on thermochemistry, which, incidentally, was later replaced by the emerging field of thermodynamics.

1ST PROOF. THERMOCHEMISTRY, HEAT OF NEUTRALIZATION

Thermochemistry, a branch of physical chemistry, deals with the exchange of heat energy at changes of state of a chemical system, e.g., at phase transformations and chemical reactions. It hypotheses that all chemical changes involve the generation of heat, that the heats of reaction were direct measures of the chemical affinity, and that those processes take place where the most heat is generated. Such processes are dissolving, mixing, diluting of substances, the decomposition and recombination of compounds, the decomposition of salts by acids, the determination of the stoichiometry of acids and bases, the reaction of acids with bases, i.e., neutralization reactions, the change of the aggregate state such as melting or evaporation, etc.

Thermochemical experiments were first described 1838 in a paper from the estate of Pierre Louis Dulong (1785 – 1838),[56] followed by the comprehensive "*Thermo-chemische Untersuchungen*" (Thermochemical studies) by Germain Henri Hess[13] from 1840 to 1842,[57-65] and between 1844 and 1850 by the investigation of Abria,[66] by Pierre Antoine Favre (1813 – 1880) and Johann Theobald Silberman (1806 – 1865),[67, 68] and by Thomas Andrews (1813 – 1885).[69-72]

However, the main contributions to the recent theory were published by the Danish chemist Hans Peter Jørgen Julius Thomsen[14] as early as 1853 and 1854 in four papers on "*Die Grundzüge eines thermo-chemischen Systems*" (The principles of a thermo-chemical system) in Poggendorff´s *Annalen der Physik und Chemie*.[73-76] Nevertheless, the French chemist Marcellin Berthelot quoted above, who did not formulate a slightly modified form until 1865 as a lecture (in lessons at the Collège de France in 1865 and published it the same year in the *Revue des cours scientifiques*, see also ref. [77]) and also not until 1869 as a journal article in *Annales de chimie et de physique*[77] claimed the priority of thermochemistry, which led to a lifelong fierce controversy between the two scientists.

After a hiatus of fifteen years, Thomsen continued from 1869 to 1871 with extended studies on "*Wärmetönung*" ("heat toning"), as he called it, that is, the heat released or absorbed by the change of state of a sys-

---

[13] Hess formulated the law named after him,[57] also known the law of constant heat summation. It states that *the heat evolved or absorbed in a chemical process is the same whether the process takes place in one or in several steps.*

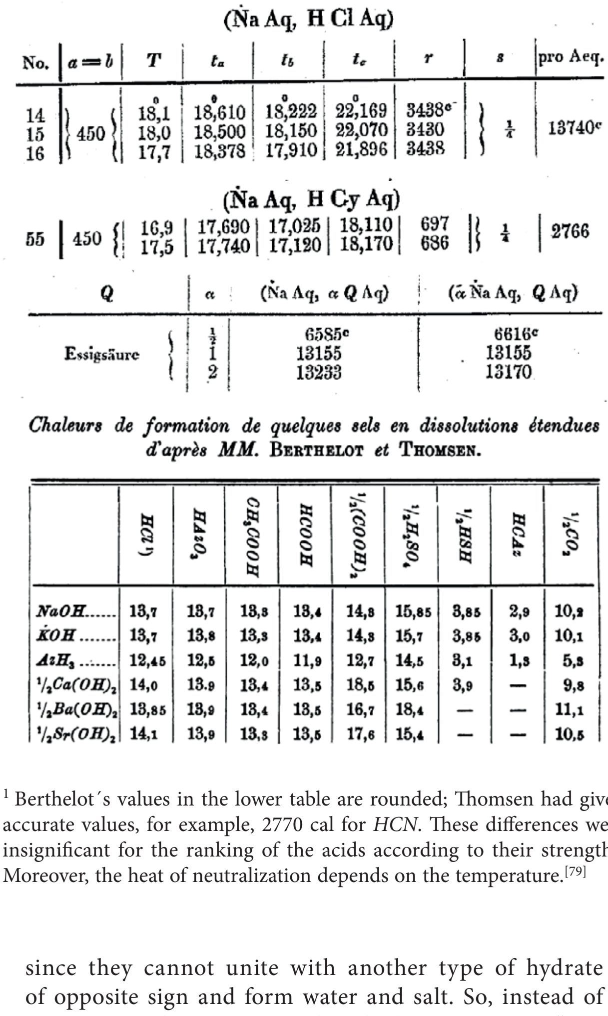
[14] Hans Peter Jørgen Julius Thomsen (1826 – 1909) was professor for chemistry at the University of Copenhagen from 1866 till 1891, and from 1883 till 1902 director of the *Polytekniske Læreanstalt*, the later Danish Technical University.

tem. Of all the heats of reaction measured by Thomsen, Arrhenius was most interested in the heat of neutralization of an acid with a strong base in order to classify acids according to their strength on the basis of the amounts of heat they released. Between 1869 and 1870 Thomsen published a series of six paper on "*Thermochemische Untersuchungen*"[78-83] (Thermochemical studies) with heats of reaction of numerous acids, including those of neutralization. He collected the results from 1100 calorimetric experiments and 600 chemical reactions in "*Thermochemische Untersuchungen. I. Neutralisation und verwandte Phänomene*" from 1882.[84] It was the first volume of a series of four.[15] Cutouts from the tables with the values of monobasic hydrochloric, acetic and hydrocyanic acid which he published in 1869[79] and 1870[82] in *Annalen der Physik und Chemie* are shown in the upper panel of Table 2.

The main source Arrhenius quoted in his work was Marcellin Berthelot´s 1879 book "*Essai de mécanique chimique fondée sur la thermochimie.*" Tabulated values of the heat of neutralization from *Tome 1. Calorimétrie*",[44] *Livre III (Données numeriques), Chapitre VI (Chaleurs de formation des sels)*, p. 383 ff., are shown in the lower panel of Table 2. These rounded values were taken by Arrhenius in his dissertation in Part 2, p.68.

In his dissertation Arrhenius discussed the reasons for the greater heat of neutralization of the strong acids and the lower heat of the weaker acids. He formulated the principle in § 20, p. 67, of Part 2. It read

Table 2. Heats of neutralization for different acids and bases. Upper panel: Thomsen´s values from 1869 and 1870 for the monobasic hydrochloric, hydrocyanic and acetic acid. T; temperature in °C. α; number of equivalent base added to 1 equivalent acid. Concentrations: varying, typically 1 eq. in 200 or 400 L water. Taken from refs. [79, 82]. Lower panel: Table from Berthelot´s 1879 book which was reproduced by Arrhenius in his dissertation, Part 2, p. 68.[1]. Temperature 15°C. Numbers are in kcal. For details, see main text.



($\dot{N}$a Aq, H Cl Aq)

No.	a = b	T	t_a	t_b	t_c	r	s	pro Aeq.
14	450	18,1°	18,610°	18,222°	22,169°	3438c	¼	13740c
15		18,0	18,500	18,150	22,070	3430		
16		17,7	18,378	17,910	21,896	3438		

($\dot{N}$a Aq, H Cy Aq)

No.	a = b	T	t_a	t_b	t_c	r	s	pro Aeq.
55	450	16,9	17,690	17,025	18,110	697	¼	2766
		17,5	17,740	17,120	18,170	686		

Q	α	($\dot{N}$a Aq, α Q Aq)	(α $\dot{N}$a Aq, Q Aq)
Essigsäure	½	6585c	6616c
	1	13155	13155
	2	13233	13170

Chaleurs de formation de quelques sels en dissolutions étendues d'après MM. Berthelot *et* Thomsen.

	$HCl^{1)}$	$HAzO_3$	CH_3COOH	$HCOOH$	$^1/_2(COOH)_2$	$^1/_2H_2SO_4$	$^1/_2HSH$	$HCAz$	$^1/_2CO_2$
$NaOH$......	13,7	13,7	13,8	13,4	14,8	15,85	3,85	2,9	10,2
$\dot{K}OH$	13,7	13,8	13,3	13,4	14,3	15,7	3,85	3,0	10,1
AzH_3	12,45	12,5	12,0	11,9	12,7	14,5	3,1	1,3	5,3
$^1/_2Ca(OH)_2$	14,0	13.9	13,4	13,5	18,5	15,6	3,9	—	9,8
$^1/_2Ba(OH)_2$	13,85	13,9	13,4	13,5	16,7	18,4	—	—	11,1
$^1/_2Sr(OH)_2$	14,1	13,9	13,3	13,5	17,6	15,4	—	—	10,5

[1] Berthelot´s values in the lower table are rounded; Thomsen had given accurate values, for example, 2770 cal for *HCN*. These differences were insignificant for the ranking of the acids according to their strengths. Moreover, the heat of neutralization depends on the temperature.[79]

> § 20. *Degagement de chaleur aux reactiones chimiques.*
> "As we know, Mr. THOMSEN claims that all bases, when in the form of dissolved hydrates, release the same amount of heat by neutralizing the same amount of an acid. This simple fact is called "saline thermoneutrality". In contrast, not all acids release the same amount of heat when they combine with a base. This fact seemed very bizarre to the savants of thermochemistry. However, according to the above, it seems easy to explain it. It is obvious that from the thermochemical point of view, two hydrates can be equal only when they are both in the *active state*. In the inactive state, the analogous compounds do not play the role of hydrates (acids or bases), since they cannot unite with another type of hydrate of opposite sign and form water and salt. So, instead of assuming with THOMSEN that the hydrates are in "dissolved state", we assume that they are in "active" state. Then we express the following, very plausible hypothesis:
> *Le procédé chimique à cause duquel un système d´un équivalent l'acide (actif) et d'un équivalent de base (aussi active) se transforme en un nouveau système, consistant d´un sel (non compliquè) et de l'eau, est accompagé d'un* même *dégagement de chaleur indépendant de la nature des hydrates*. (The chemical process of converting a system consisting of one equivalent of an (active) acid and one equivalent of a (also active) base into a new system consisting of a (non-complex) salt and water is accompanied by the same heat evolution, which is independent of the nature of the hydrates).

---

[15] Thomsen´s motivation to publish this series from 1882 and 1886[84],[85],[86],[87]. was due to reasons of priority. In the Introduction of the first volume in 1882 (p. 12, ref. [84].) Thomsen wrote "*Schon im Jahre 1853-54 hatte ich in den Annalen der Physik und Chemie, Bd. 88, 90, 91 und 92* [73],[74],[75],[76] *die fundamentalen dynamischen Gesetze der chemischen Processe entwickelt. Da man aber mir die Priorität hierfür hat streitig machen wollen, werde ich im Folgenden einige Hauptpunkte dieser Abhandlungen wieder hervorheben.*" (Already in 1853-54 I had developed the fundamental dynamical laws of chemical processes in *Annalen der Physik und Chemie*, vol. 88, 90, 91, and 92. Since, however, one has wanted to dispute my priority for this, I will again emphasize some main points of these treatises in the following.)

Arrhenius then described the sequence in which the neutralization reactions occurs: *1.* Neutralization through the active parts of the hydrates. *2.* New formation of active from inactive parts. *3.* Neutralization of the newly formed active parts. *4.* Formation of complex molecules of the resulting salt. *5.* Possible solidification of the salt.

He concluded that the sum of the generation of heat in processes *Nr. 1* and *3* must be constant at the formation of one equivalent of salt. This is the reason why the heat of neutralization of strong acids with strong bases is about constant, and agrees with the experimental values shown in Table 2. Process *Nr. 2*, the new formation of active from inactive parts, in contrast, is accompanied by the absorption of heat. He expressed it in *Theses 52, 53,* and *56*:

> *Thesis 52. La transformation de l´état inactif en l´état actif* d´un hydrate (*faible) est accompagnée par une absorption de chaleur.* (52. The transformation from the inactive to the active state of a (weak) hydrate is accompanied by heat absorption.)
> *Thesis 53. A la neutralisation, un acid faible dégage en général moins de chaleur qu´un acide fort. Une proposition analogue est valable pour les bases* (On neutralization, a weak acid generally releases less heat than a strong acid. An analogous proposition is valid for bases).
> *Thesis 56. La chaleur de neutralisation dégagée par la transformation d´une base et d'un acide parfaitement actifs en eau et en sel non-compliqué, n'est que la chaleur d'activité de l'eau.* (The heat of neutralization released at the conversion of a perfectly active base and acid into water and a simple salt is nothing but the heat of activation of the water.)

The heat of neutralization is therefore a measure for the strength of an acid or a base. For strong acids and bases it is given by the reaction of their active parts, the H^+ and the OH^- ions of the completely dissociated electrolytes, under formation of the inactive water molecule. This amount corresponds to the heat of activation of water, which is necessary to completely split water into H^+ and OH^-. This amount is about 13700 cal mol^{-1} at 18 °C,[16] and corresponds to the experimentally found neutralization heat of a strong acid with a strong base. When a weak acid is neutralized with a strong base such as $NaOH$, the inactive part of the acid must be converted to the active, which requires heat absorption. The total heat balance of the reaction is thus less than 13700 cal, and its magnitude is the measure Arrhenius had assumed for the strengths of acids. Since the heats of neutralization correlate with the molecular conductivities, Arrhenius succeeded in confirming his original hypothesis, that is, the results of thermochemistry were the first proof which he had mentioned as a necessity in his Nobel lecture.

He concluded his dissertation with the summary of Part 2, of which we reproduce the key passage. It read

> *Dans la partie présente de cet ouvrage, nous avons d'abord fait voir la vraisemblance·de ce que les* électrolytes *peuvent se rencontrer sous deux formes différentes, l'une active, l'autre inactive, de sorte que la partie active est toujours, dans les memes circonstances extérieures (température et dilution), une certaine fraction de ln quantité totale de l'électrolyte. La partie active conduit l'électricité et est ainsi en réealité* électrolytique*, mais non pas la partie inactive.*

In the present part of this work, we have first shown the verisimilitude that electrolytes can occur in two different forms, one active, the other inactive, so that the active part is always, under the same external circumstances (temperature and dilution), a certain fraction of the total quantity of the electrolyte. The active part conducts electricity and is therefore actually electrolytic, but the inactive part is not. [Boldface by the author].

This was the first step towards what later became known as the dissociation theory of electrolytes in solutions. He completed his dissertation and submitted it to the Swedish Academy of Sciences.

PRESENTATION OF DISSERTATION, ITS LOW GRADING, AND THE CONSEQUENCES

Arrhenius presented his hypotheses in a lecture held at the Royal Swedish Academy of Science on June 6, 1883, and submitted them at Uppsala University in a slightly modified form as dissertation in 1884.[6, 7] However, Robert Thalén, the physics professor, and Per Teodor Cleve, the chemistry professor at the university who evaluated the dissertation were not convinced, and Arrhenius' doctoral thesis was given a poor grade. It was awarded a forth class, its defense a third class. This rating was too low to lecture as a docent, and the Uppsala faculty offered him only an unpaid position, what he refused.

Immediately after the disappointing assessment in 1884, Arrhenius sent copies of his dissertation to several European physical chemists, most of whom did not respond,[17] except Wilhelm Ostwald, professor at the

---

[16] The heat of neutralization depends on the temperature. Thomsen reported that it decreases by 43 cal °C^{-1}, see footnote 16.

[17] We would like to comment that although the English physicist Sir Oliver Joseph Lodge (1851 – 1940) mentioned above did not contact Arrhenius directly, he subjected the two parts of Arrhenius' dissertation separately to an extraordinary detailed critical analysis of almost every

Riga Polytechnic.[18] Ostwald described his ambivalent judgment to reply or not in Chapter 11 of his "*Lebenslinien: Eine Selbstbiographie*"[90] (Lines of Life: A Self-Biography; also Lifelines: An Autobiography) after reading Arrhenius's dissertation. He found that "the work contained obvious weaknesses (which were also exaggeratedly emphasized by other critics afterwards); so that I still had to reckon with the possibility that those correct results had only turned out that way by chance".

Ostwald had already executed experiments about the specific affinities of acids and bases. He conducted them during the time Arrhenius was working on his dissertation, but in contrast to him, using various non-electrochemical methods. His investigations, which led to the second proof of Arrhenius´ hypotheses, were a good fortune for Arrhenius. He said retrospectively in his Nobel lecture "*Development of the theory of electrolytic dissociation*" held on December 11, 1903 (we continue with this lecture which we have cited above up to "*etc.*")

> ". ….. *etc.* The Norwegian research scientists Guldberg and Waage had developed a theory according to which the strength of different acids could be measured …… by their capacity of increasing the speed of certain chemical reactions. In conformity with this we can suppose that the speed of a reaction produced by an acid is proportional to the number of active molecules in it. I had only a few experiments by Berthelot to demonstrate this law, but in 1884 Ostwald published a large number of observations which proved that this conclusion was correct."

These observations of Wilhelm Ostwald, to which Arrhenius referred, will be the subject of the following chapter.

2ND PROOF. OSTWALD´S "STUDIEN ZUR CHEMISCHEN DYNAMIK" (STUDIES ON CHEMICAL DYNAMICS): MOLECULAR CONDUCTIVITY AND AFFINITY COEFFICIENT

Wilhelm Ostwald, born in 1853 in Riga, now capital of Latvia, studied from 1872 chemistry at the University of Dorpat, now Tartu, Estonia. In 1882, at the age of 29, he became a professor at the Riga Polytechnic, where he continued the work he had made in Dorpat.[91-95]

From 1882 to 1884 Ostwald, already professor in Riga, published the first three parts of the series on "*Studien zur chemischen Dynamik*" (Studies on Chemical Dynamics) but without knowing about Arrhenius' dissertation. A photograph of Ostwald taken at the time when he began with these studies is shown in Figure 5.

The goal of Ostwald's "Studies in Chemical Dynamics" was to determine the relative strengths of acids based on the intensity of an acid's effect on a base. His aim was to express this intensity by a number which he quoted as "Affinitätsgröße" (affinity value, affinity, in some papers affinity coefficient) of the acid. He determined this affinity from the reaction rate of the acid with the corresponding reactant. In other words, the greater the affinity of an acid in a reaction under unchanged conditions, e.g., at constant temperature and concentrations, the stronger the acid. He chose acids for his new research because he had already done earlier work with these compounds in Dorpat between 1877 and 1881 on other questions and with other methods.[93-97]

In his first paper of this series,[98] dated December 1882, Ostwald determined the rate of complete conversion of acetamide by different acids, all under the same conditions. For "*Zweite Abhandlung*"[99] Ostwald hydrolyzed methyl acetate to acetic acid and methyl alcohol. This reaction was analogous to that described by M. Berthelot and L. Péan de St. Gilles in 1862 for the formation of esters from acids and alcohols.[100-103] The subject of "*Dritte Abhandlung: Die Inversion des Rohrzuckers*" (Third Treatise: The Inversion of Cane Sugar)[104] was the influence of acids on the rotation of polarized light of cane sugar. Investigations on the course of inversion as a function of reaction time had already been successfully carried out in 1850 by Ludwig Wilhelmy[19] but remained

---

paragraph, thesis and equation in a report of the British Assiciation for the Advancement of Science.[88]. The entire analysis contained no less than 27 closely printed pages (Part I, p. 357- 362, Part II, p. 362-384.) Lodge pioneered radiotelegraphy and discovered electromagnetic radiation earlier than Heinrich Hertz. He held patents for radio and the moving coil loudspeaker. He was familiar with electrophoresis as well. In 1886, he invented the moving boundary method,[89] an electrophoretic method that replaced Hittorf´s method for determining transference numbers. We have described the latter method in detail in Part 3 of our historical review.[5] The moving boundary method had the great advantage that the transference numbers could be measured in much less time, with much less effort and without chemical analysis of the constituents in the electrode compartments, at lower cost and, above all, with much higher accuracy. In 1881 he became professor of physics and mathematics at University College Liverpool. From 1900, Lodge was the first rector of the new University of Birmingham. In the late 1880s Lodge developed a keen interest in spiritualism and telepathy and became a member of the Ghost Club. Founded in London in 1862, the Ghost Club is a paranormal research organization. The club is concerned with ghosts and hauntings, but has also studied UFOs, dowsing and cryptozoology.

[18] In the following we will tell more about Ostwald, since he played an eminently important role in Arrhenius' life.

[19] Ludwig Ferdinand Wilhelmy (1812 – 1864) was a German physicist and physical chemist. He is considered the first to publish quantitative studies on chemical kinetics. In 1850, using a polarimeter, he measured the conversion of sucrose into fructose and galactose after the addition of acid (we have quoted it in the main text; it is the reaction Ostwald studied in his *Third Treatise*). He formulated the kinetics of the reaction in complex differential equations and published in 1851 the book "Versuch einer mathematisch-physikalischen Wärme-Theorie"[105] (Attempt of a mathematical-physical theory of heat).

Figure 5. Photo of Wilhelm Ostwald from 1882, the year he became a professor at the Riga Polytechnic. Unknown author. Reprinted with permission. © Copyright Gerda and Klaus Tschira Foundation, 2022.

largely unnoticed.[106, 107] In 1862 J. Löwenthal and E. Lenssen[108, 109] also described the influence of acids on the inversion of sugars.

Ostwald added the corresponding acid solution to a solution of cane sugar and determined the inversion rate from the optical rotation measured with the aid of a thermostated polarization tube. The rate constants were calculated by Wilhelmy´s formula.[106, 107] This method was the most accurate and the least labor-intensive and time-consuming of the methods used, since it did not require chemical analysis to keep track of its progress. The affinities of the acids obtained by this third treatise agreed sufficiently well with those from the hydrolysis of methyl acetate. This result definitively confirmed the conclusion drawn from the previous treatises that the order of the affinity corresponds to the order of the strengths of the acids. Ostwald thus succeeded in compiling such a list by three independent methods.

Ostwald had just completed the third treatise – it was dated April 1884 – when he received in June 1884 the dissertation Arrhenius has sent to him. He reported in Chapter 11 of his autobiography [90] that at this day he had a tooth ulcer, his wife gave birth to their daughter, and, last but not least, he received the dissertation. The tooth ulcer healed, mother and daughter were well, but the dissertation gave him a headache and a restless night. At first he thought the work was nonsense. After closer study, however, he realized that the author had treated and partly solved the great problem of the proportionality of the affinity coefficients of acids and bases and their electric conductivity in a much more comprehensive way than he had.

We quote a passage from his autobiography which we regard as typical of the integrity of Ostwald's character. He describes with admirable sincerity the moral dilemma he faced. Should he, who was himself working on a new branch of research, prevent a possible competitor, or should he include him? We reproduce in the footnote [20] what Ostwald reported on this.

Immediately thereafter, Ostwald began to measure the molecular conductivities of the acids whose affinity coefficients he had determined in his "Studien zur chemischen Dynamik" described above, and meanwhile informed Arrhenius by letter about his plan. To measure the conductivities, he aimed to apply alternating current according to Kohlrausch's method,[25] but did not yet have any equipment available in Riga. He therefore borrowed a resistor box from the Riga telegraph office for a short time, copied it himself in the Polytechnic's workshop, and performed the measurements on his extensive collection of acids, which he possessed from earlier investigations. The measurements were completed in a few days, and he immediately submitted the results for publication to *Journal für praktische Chemie* as "*Notiz* über *das elektrische Leitungsvermögen der Säuren*" *(*Note on the Electrical Conductivity of Acids), dated July 1884.[110]

In this publication he held that the rate of chemical reactions depends on the velocity and thus on the conductivity of the ions which are involved. He mentioned that in his "*Studies in Chemical Dynamics*" of 1883

---

[20] Ostwald wrote: "One can easily imagine what a confusion of feelings such a realization must arouse in a young researcher [Ostwald] who has only to make his future and suddenly finds himself confronted with a highly energetic co-worker in the field which he had chosen so lonely and remote. In addition, the work contained obvious weaknesses ......, so that I still had to reckon with the possibility that those correct results had only turned out this way by chance. For a few days, as in Bürger's ballad, the black and the white companion [Gottfried August Bürger (1747 – 1794) was a popular German poet; author´s note] fought over my soul. It was certainly not difficult to keep this sudden competitor in the background by silence, since at present only a few professional colleagues cared at all about such questions. Then, because of the existing errors, one could condemn the whole thing and, besides, the publication in the writings of the Swedish Academy of Sciences was an obstacle for the dissemination anyway, since these hardly came into the hands of the chemists.
So all I had to do was ignore the writing to keep the competitor at bay, if not for good, then for the foreseeable future. ... I did not learn the details of the technique of fighting unwelcome coworkers as well as competitors until later. ... On the other hand, the scientific idealism that I had acquired as a self-evident prerequisite for all work in this highest field of human progress ... was active. ... In addition, the joyful feeling asserted itself to be able to plow a virgin soil shoulder to shoulder with a new fellow worker [Arrhenius], ..., especially since I found him equipped with intellectual working means which I had not used before and which, in combination with those familiar to me, ensured an all the more effective progress. … "

and 1884 he showed that the rates of reactions affected by acids are proportional to the affinity of the acids as determined by him.[98, 99, 104] Therefore, there is also a proportionality between the reaction rates and the electrical conductivity. He pointed out that Svante Arrhenius arrived at the same result by a different route, by that which he published in his dissertation in 1884.

In contrast to the many controversies about the priority of theories by other researchers, Ostwald emphasized with remarkable righteousness on p. 93

> ”*Dem Autor dieser Abhandlungen, die zu dem Bedeutendsten gehört, was auf dem Gebiet der Verwandtschaftslehre publicirt worden ist, kommt nicht nur die Priorität der Publikation, sondern auch die der Idee zu*“ (To the author of these treatises, which belong to the most important what has been published in the field of affinity theory, comes not only the priority of the publication, but also that of the idea.)

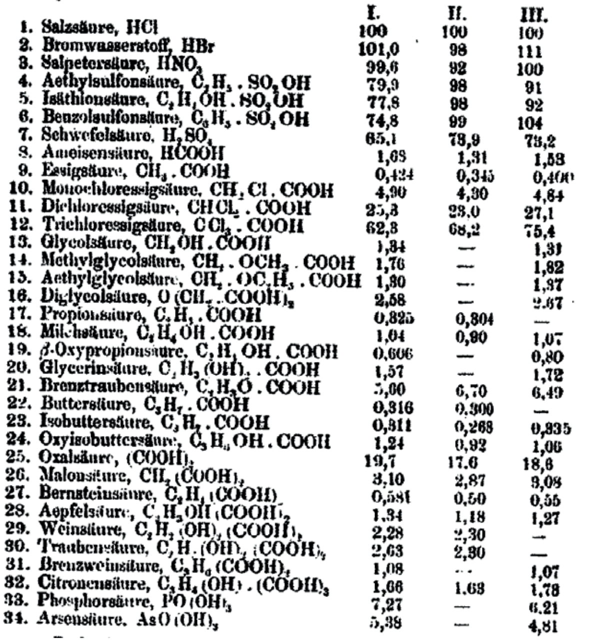
Since Arrhenius' dissertation was submitted to the Swedish Academy on June 6, 1883, and published in May 1884, he received it directly from him not until June 1884. He emphasized this chronological order by noting: "I give these details in order, in stating the independence of my efforts in this field from Arrhenius' work, not to fall into the appearance of an unmotivated priority reclamation."

Ostwald then gave in a table (Table 3) the comparative values obtained by three independent methods. Column I of this table lists the conductivities of the acids measured by him, column II the rate constants of the catalysis of methyl acetate and column III the rate constants for the inversion of cane sugar.

Comparing these values Ostwald concluded

> *Eine Übereinstimmung, wie sie die drei Reihen bieten, habe ich selbst nicht erwartet; diesselbe ist wohl geeignet, jeden Zweifel an der Bedeutung der Affinitätsgrössen zu heben. Bedenkt man, dass weder die Temperatur, noch die Verdünnung bei den drei verglichenen Versuchsreihen diesselbe war, so darf man die Uebereinstimmung der drei Reihen, deren Unterschiede im Uebrigen ganz gesetzmässig verlaufen wohl befriedigend nennen. In Bezug auf die weitgehenden Consequenzen, welche aus diesem Ergebnis gezogen werden können, muss ich auf die oben citirten Arbeiten von S. Arrhenius verweisen. Riga, Juli 1884.* (I myself did not expect an agreement such as that offered by the three series; this is probably suitable to remove any doubt about the significance of the affinity coefficients. Considering that neither the temperature nor the dilution was the same in the three series of experiments compared, one may quote the agreement of the three series, whose differences are otherwise quite lawful, as satisfactory. With regard to the far-reaching consequences which can be drawn from this result, I must refer to the above cited work of S. Arrhenius. Riga, July 1884).

Table 3. Column I. records the conductivities of the acids measured by Ostwald after he received the dissertation Arrhenius has sent to him, column II. the rate constants of the hydrolysis of methyl acetate, column III. the rate constants for the inversion of cane sugar. The values were related to *HCl*=100. Taken from ref. [110].



	I.	II.	III.
1. Salzsäure, HCl	100	100	100
2. Bromwasserstoff, HBr	101,0	98	111
3. Salpetersäure, HNO_3	99,6	92	100
4. Aethylsulfonsäure, $C_2H_5 . SO_2OH$	79,9	98	91
5. Isäthionsäure, $C_2H_4OH . SO_2OH$	77,8	98	92
6. Benzolsulfonsäure, $C_6H_5 . SO_2OH$	74,8	99	104
7. Schwefelsäure, H_2SO_4	65,1	73,9	73,2
8. Ameisensäure, HCOOH	1,68	1,31	1,53
9. Essigsäure, $CH_3 . COOH$	0,424	0,345	0,400
10. Monochloressigsäure, $CH_2Cl . COOH$	4,90	4,30	4,84
11. Dichloressigsäure, $CHCl_2 . COOH$	25,3	23,0	27,1
12. Trichloressigsäure, $CCl_3 . COOH$	62,3	68,2	75,4
13. Glycolsäure, $CH_2OH . COOH$	1,34	—	1,31
14. Methylglycolsäure, $CH_2 . OCH_3 . COOH$	1,76	—	1,82
15. Aethylglycolsäure, $CH_2 . OC_2H_5 . COOH$	1,80	—	1,87
16. Diglycolsäure, $O(CH_2 . COOH)_2$	2,58	—	2,67
17. Propionsäure, $C_2H_5 . COOH$	0,825	0,304	—
18. Milchsäure, $C_2H_4OH . COOH$	1,04	0,90	1,07
19. β-Oxypropionsäure, $C_2H_4OH . COOH$	0,606	—	0,80
20. Glycerinsäure, $C_2H_3(OH)_2 . COOH$	1,57	—	1,72
21. Brenztraubensäure, $C_2H_3O . COOH$	5,60	6,70	6,49
22. Buttersäure, $C_3H_7 . COOH$	0,316	0,300	—
23. Isobuttersäure, $C_3H_7 . COOH$	0,311	0,268	0,335
24. Oxyisobuttersäure, $C_3H_6OH . COOH$	1,24	0,92	1,06
25. Oxalsäure, $(COOH)_2$	19,7	17,6	18,6
26. Malonsäure, $CH_2(COOH)_2$	3,10	2,87	3,08
27. Bernsteinsäure, $C_2H_4(COOH)_2$	0,581	0,50	0,55
28. Aepfelsäure, $C_2H_3OH(COOH)_2$	1,34	1,18	1,27
29. Weinsäure, $C_2H_2(OH)_2(COOH)_2$	2,28	2,30	—
30. Traubensäure, $C_2H_2(OH)_2(COOH)_2$	2,63	2,30	—
31. Brenzweinsäure, $C_3H_6(COOH)_2$	1,08	—	1,07
32. Citronensäure, $C_3H_4(OH) . (COOH)_3$	1,66	1,63	1,78
33. Phosphorsäure, $PO(OH)_3$	7,27	—	6,21
34. Arsensäure, $AsO(OH)_3$	5,38	—	4,81

Immediately after the publication of this Note, Ostwald traveled via Stockholm to Uppsala where he met Arrhenius, and spent three days with lively discussions, and borne of great personal sympathy. They agreed that Arrhenius should come to Riga as soon as possible for joint work. For this, however, Arrhenius first needed his habilitation. Through Ostwald's offer of a position in Riga, Arrhenius was appointed docent in Uppsala. Moreover, Arrhenius received in 1886, with Edlund's support, a three-year traveling scholarship from the Royal Swedish Academy of Sciences to meet leading physicists in Europe.

He first used it for cooperation with Ostwald in Riga in 1886, visited Friedrich Kohlrausch in Würzburg in 1886 and 1887, Ludwig Boltzmann in Graz in 1887 (see the photo in Figure 6), van 't Hoff in Amsterdam in 1888,[21] and again in 1888 Ostwald in Leipzig, where Ostwald became professor of physical chemistry in 1887.

---

[21] Where he published by support of van ´t Hoff ”*Theorie der isohydrischen Lösungen*“,[112] concluded March 1888.

Figure 6. Photograph taken during Arrhenius´ visit of Ludwig Boltzmann in Graz in 1887.[111] Date: October 1887. Source Universität Graz. Author unknown. Arrhenius is the fourth from the right, standing behind Boltzmann. Public domain. The photograph shows, standing from the left: Walther Nernst,. Heinrich Streintz, Svante Arrhenius, Richard Hiecke. Sitting, from the left: Eduard Aulinger, Albert von Ettingshausen, Ludwig Boltzmann, Ignaz Klemenčič, Victor Hausmanninger.

3RD PROOF. VAN ´T HOFF´S THEORY OF OSMOSIS

It was a beneficial coincidence that in the midst 1880s Jacobus Henricus van ’t Hoff[22] investigated the osmotic pressure between the dilute solution of a solute and the pure solvent, which were separated by a semipermeable membrane. He was referred by the Dutch botanist Hugo de Vries to the work of Wilhelm Pfeffer (1845–1920), then professor of botany in Basle. In 1877 Pfeffer had measured the osmotic pressure, P, with membranes[23] that were permeable only to water, but not to solutes.[115] He found that, at a given concentration, P is proportional to the absolute temperature, T; at a given temperature, P is inversely proportional to the volume, V, or proportional to the concentration. Pfeffer describes this dependence by $P=kt/V$ or $PV=kT$, where k was a constant of proportionality.

Applying the laws of Boyle, Henry, Gay-Lussac and Avogadro, van ´t Hoff deduced that the constant k is equal to the gas constant R. The equation for the osmotic pressure read thus $PV=RT$. It is expressing the analogy between the osmotic pressure of a compound in solution and the pressure of the compound when it is in the gaseous state under the same conditions, i.e., at the same temperature and in the same volume.

van ´t Hoff presented his theory at l´Académie Royal des Sciences de Suède on October 14, 1885 as memoir entitled "*Lois de l`équilibre chimique dans l`état dilué, gazeux ou dissous*" and published it in *Kongliga Svenska Vetenskaps-Akademiens Handlingar*.[116] On p. 43 he emphasized this analogy by stating

> *La pression exercée par les gaz à une température déterminée si un méme nombre de molécules en occupe un volume donné, est* égale *à la pression osmotique qu'exerce dans les mémes circonstances la grande majorité des corps, dissous dans les liquides quelconques.* (The pressure exerted by the gases at a given temperature determined if the same number of molecules occupy a given volume, is equal to the osmotic pressure exerted by the great majority of bodies, dissolved in any liquids under in the same circumstances).[24]

Transformed to modern terminology: the osmotic pressure, Π, is directly proportional to the concentration, c, of the solute, or to the number of dissolved particles per unit volume, respectively. This relation is formally equal to the equation of state of an ideal gas, and reads $\Pi V=nRT$ (n is the number of moles) or $\Pi=cRT$.[118]

However, van ‘t Hoff found in experiment that the relationship held good for non-electrolytes as solutes, e.g., for glucose, but for electrolytes the osmotic pressures were higher. For such solutions the equation had to be modified to $PV=iRT$ (now $\Pi=icRT$, with i being the van ’t Hoff factor.

For strong electrolytes these factors, i, were integer numbers, and were equal to the number of ions upon complete dissociation. For electrolytes, e.g., with formula AB, the van ’t Hoff factor was 2, it was 3 for A_2B, etc. Remarkably, non-integer numbers were measured for weak electrolytes, indicating partial dissociation of molecules into ions and uncharged fractions. It was shown that van ‘t Hoff factor i and Arrhenius's degree of dis-

---

[22] The Dutch physical chemist Jacobus Henricus van ’t Hoff (1852 – 1911) was one of the most renowned scientists of his time. He made important contributions in various fields of chemistry. In addition to the subject treated here, namely the theory of the osmotic pressure of dilute solutions, he proposed the tetrahedral structure of the bonds of the carbon atom, explained optical activity, and developed stereochemistry. He contributed significantly to chemical kinetics and thermodynamic issues. He was the first Nobel Laureate in Chemistry in 1901 "*for his discovery of the laws of chemical dynamics and osmotic pressure in solutions*."

[23] Pfeffer produced membranes by placing two different solutions into contact in a cell made of clay. One solution contained copper acetate, the other potassium ferrocyanide. The detailed properties of numerous membranes and the conditions for their formation were described by Moritz Traube in 1867.[113] Those used by van ‘t Hoff are found as experiment No. 145 on page 244 in ref. [114].

[24] van ´t Hoff published in 1866 the very similar, but not the same paper "*L`équilibre chimique dans les Systèmes gazeux ou dissous à l`état dilué*" in Archives Néerlandaises.[117]

sociation α are related to each other.[25] For Arrhenius, the results of van 't Hoff's osmotic pressure theory were the most convincing confirmation of the theses of his dissertation.[26] This was the third proof and the decisive step towards a well-founded dissociation theory.

Ostwald was highly dissatisfied with the fact that the journals established in the1880s published scientific papers in the field of physical chemistry mostly intermingled with publications covering a wide variety of fields. In his opinion, papers on physical chemistry topics did not receive the attention they deserved. He therefore decided to found a new journal focusing on the field of physical chemistry, and published the first issue of this "*Zeitschrift* für *physikalische Chemie, Stöchiometrie und Verwandtschaftslehre,*[27] in 1887 with van ´t Hoff as co-editor and with the collaboration of the most prestigious European scientists. It was, like the following issues, a resounding success.

This journal offered Arrhenius a platform for the distribution of his theory, including explanatory statements for the unconvinced readers. Ostwald published Arrhenius´ paper on electrolyte dissociation[8] in this first volume.[28]

THE THEORY: "*UEBER DIE DISSOCIATION DER IN WASSER GELÖSTEN STOFFE*" (ON THE DISSOCIATION OF COMPOUNDS DISSOLVED IN WATER)

Arrhenius´ seminal publication on the dissociation of dilute electrolytes[8] essentially contained the results of his dissertation and the proofs we have already discussed above. So we will only briefly repeat these parts.

It is noticeable that he did not present the theory chronologically, that is, not with the theses of his dissertation that electrolyte molecules are divided into an electrically active and an inactive part even when no current passes the solution. The reason for this may have been that many chemists were skeptical of Clausius' hypothesis of free ions, which Arrhenius considered one of the precursors of his theory. Arrhenius began his publication with van 't Hoff's theory of the osmotic pressure exerted by dilute solutions of any substance on semi-permeable membranes[116, 117] which is described above in the 3rd proof. Its decisive result was that a factor, the van ´t Hoff factor *i*, has to be introduced. It was an integer or a fractional number, and depended on the number of all dissolved particles in the solution which, as it was believed also by Ludwig Boltzmann,[121] "bombard" the membrane.

Arrhenius went on to consider the activity coefficient as defined after *Thesis 15* in his dissertation, symbolized by α and called degree of dissociation in his dissociation theory. α can be determined from the ratio of the molecular conductivity Λ_d of the diluted solution and the limiting molecular conductivity, Λ_o as $\alpha=\Lambda_d/\Lambda_o$. He showed, as also proved by Kohlrausch, that at the limit of infinite dilution there is complete dissociation. In this case α=1, at lower dilution α<1. Arrhenius

---

[25] van 't Hoff wrote about these deviations of the osmotic pressure for solutions of electrolytes from Avogadro's law, that Arrhenius pointed out by letter to him the connection of factor *i* with the degree of dissociation (ref. [118] , p. 501.)

[26] van ´t Hoff´s theory of the osmotic pressure offered the theoretical explanation of the observations of the lowering of the freezing point of solutions with different solutes by François-Marie Raoult in 1882.[15] Raoult measured this lowering with six different solvents and about two hundred compounds as solutes and found empirically (transl. from French)

(i) All bodies, on dissolving in a definite liquid compound which can solidify, lower the freezing point. (ii) In all liquids, the molecular lowering of the freeing point due to the different compounds approaches two values, invariable for each liquid, of which one is double the other.) (iii) The normal molecular lowering of the freezing point varies with the nature of the solvent.) (iv) One molecule of any compound dissolved in 100 molecules of any liquid of a different nature lowers the freezing point of this liquid by a nearly constant quantity, close to 0.62 degrees.) If water is the solvent, and molar concentrations are taken, Raoult´s law can be expressed as: When one mole of particles is dissolved in one kilogram of water, its freezing point decreases by 1.86°C (if dissolved in 100 g water, the freezing point decreases by 18.6°C). The corresponding temperature difference is called the molar freezing point depression. This effect is independent of the type of the solute. It holds good only for very dilute solutions. Note that Arrhenius will make use of it in his dissociation theory which he publishes in 1887. We clarify the depression by an example, because the relation between concentration and temperature is sometimes given confusingly. Strong electrolytes consisting of two univalent ions give the molecular depression 2x1.85 =3.7, those consisting of one bivalent ion with two univalent ions give 3x1.85 =5.55, those of one tetravalent ion with three univalent ions give 4x1.85=7.4, etc. Non-electrolytes give 1.85°C depression (the difference between 1.85 and 1.86 is insignificant in practice).

[27] In 1928 the title was shortened to *Zeitschrift für physikalische Chemie* and a new numbering was assigned, in 1954 the numbering was changed again, since 1979 the journal is published with the subtitle *International journal of research in physical chemistry and chemical physics.*

[28] Remarkably, Max Planck published in the same volume a theory of dissociation independent of Arrhenius and based on thermodynamics, entitled "Über die molekulare Beschaffenheit verdünnter Lösungen" (On the molecular nature of dilute solutions), in which he dealt mainly with Raoult's law of freezing point depression.[119]. He introduced a constant *i*, the *decomposition coefficient*, of dissolved molecules, which was identical with van 't Hoff's constant *i* calculated for aqueous solutions of various substances, but had a different physical meaning there. Unlike Arrhenius, Planck did not link his theory to electrical conductivity. It was therefore less general and received little attention. To remain in the context of what are now called colligative properties, it should be mentioned that Planck, in the paper "Über den osmotischen Druck"[120] from 1890, showed that both the existence and the magnitude of osmotic pressure follow directly from the same general thermodynamic principles that underlie the laws of vapor pressure and freezing point of a solution, without reference to molecular conceptions.

showed that if α is known, the factor *i* could be deduced as follows.

If *m* is the number of inactive molecules, *n* the number of active molecules, and *k* the number of ions into which each active molecule splits, then $i=(m+kn)/(m+n)$. Consequently, since $\alpha=n/(m+n)$, the relation between *i* and α reads $i=1+\alpha(k-1)$, and can be calculated from the known value for *k* and the measure value for α. Second, the van ‘t Hoff factor *i* can be determined from Raoult’s law of freezing point depression, i.e., from the freezing temperature *t*, as $i=t/18.5$. Arrhenius thus had two independent methods to obtain the van ‘t Hoff factor and to compare their agreement. The values for 24 compounds out of a total of 86 which were calculated by the above methods are shown in Table 4.

By comparing the values of *i* for the individual compounds in the 4th and the 5th columns of the table, Arrhenius observed indeed a pronounced agreement of most of them (the deviations of copper and cadmium salts were explained by their tendency to form complex molecules). He thus concluded that the following hypotheses to calculate the figures were correct. *(1)* van ‘t Hoff’s law is valid not only for the majority, but for all compounds, that is to say, also for electrolytes in aqueous solutions. *(2)* Each electrolyte in aqueous solutions consists partly of active molecules and partly of inactive molecules. The inactive molecules are converted into active ones when diluted, with the effect that only active molecules are present in solutions under the limiting condition of infinite dilution. This led to the 3rd proof of his hypothesis.

After a discussion of possible objections of chemists, which essentially concerned the effect of molecular complexes, Arrhenius gave further evidence of the validity of his theory, which he summarized under the concept of additivity of physical properties, and which we have not explicitly discussed as 4th proof in the previous part of this text.

Arrhenius considered the properties of dilute salt solutions as additive when the sum of the properties of the parts of the solution, that is the solvent and the parts of the molecules which actually coincide with the ions, is equal to the property of the solution. These additive properties are physical in nature and can be expressed in numerals. Arrhenius cited the following additive properties in favor of his theory.

1. The conductivity. Its additive property was described in 1879 by Kohlrausch in his law of independent ion migration.[23] The conductivity of a salt solution is therefore the sum of the conductivities of the positive and negative ions. That of water as a solvent is usually negligible, except in extremely dilute solutions (see ref. [122] and the discussion about pure water in Part 3 of our series[5]). However, this applies only to so-called strong electrolytes, that is, completely dissociated compounds such as the salts of monobasic acids and the strong acids and bases. For weak acids and bases, e.g. for acetic acid, hydrocyanic acid and ammonia, the additivity does not apply, except for extremely dilute solutions for the reasons already mentioned above.

Table 4. Comparison of the van ´t Hoff factors, *i*, for dilute aqueous solutions of non-conductors, bases, acids and salts, derived from Raoult´s freezing point depression, *t* (column 4) and from its relation to the degree of dissociation, α, as given in the above text (column 5). *α* was measured from the ratio of the conductivities of the dilute solution and those at limiting conditions. Temperature *t* is in °C. The paper lists a total of 90 solutes.

SUBSTANCE	FORMULA	α	$i=t/18.5$	$i=1+(k-1)\alpha$
NON-CONDUCTORS				
Methyl alcohol	CH_3OH	0.00	0.94	1.00
Phenol	C_6H_5OH	0.00	0.84	1.00
Acetone	C_3H_6O	0.00	0.92	1.00
Acetamide	$C_2H_3ONH_2$	0.00	0.96	1.00
BASES				
Lithium hydroxide	$LiOH$	0.83	2.02	1.83
Sodium hydroxide	$NaOH$	0.88	1.96	1.88
Ammonia	NH_3	0.01	1.03	1.01
Methylamine	CH_3NH_2	0.03	1.00	1.03
ACIDS				
Hydrochloric acid	HCl	0.90	1.98	1.90
Nitric acid	HNO_3	0.92	1.94	1.92
Sulphuric acid	H_2SO_4	0.60	2.06	2.19
Hydrogen sulphide	H_2S	0.00	1.04	1.00
Boric acid	$B(OH)_3$	0.00	1.11	1.00
Hydrocyanic acid	HCN	0.00	1.05	1.00
Formic acid	$HCOOH$	0.03	1.04	1.03
Acetic acid	CH_3COOH	0.01	1.03	1.01
SALTS				
Potassium chloride	KCl	0.86	1.82	1.86
Ammonium chloride	NH_4Cl	0.84	1.88	1.84
Potassium cyanide	KCN	0.88	1.82	1.82
Sodium acetate	CH_3COONa	0.79	1.73	1.79
Ammonium sulphate	$(NH_4)_2SO_4$	0.59	2.00	2.17
Copper sulphate	$CuSO_4$	0.35	0.97	1.35
Mercuric chloride	$HgCl_2$	0.03	1.11	1.05
Cadmium iodide	CdJ_2	0.28	0.94	1.56
Cadmium nitrate	$Cd(NO_3)_2$	0.73	2.32	2.46

2. The heat of neutralization in dilute solutions. This has already been discussed in the chapter about the 1st proof above.

3. The specific volume and the specific gravity of dilute salt solutions. Claude-Alphonse Valson reported

1871 that the specific gravity (which is the ratio of the density of an object, and a reference substance), is an additive property.[123]

4. The specific refractive power of solutions. It was shown by John Hall Gladstone in 1863 that the so-called refractive equivalent is an additive property.[124] This additivity also applies to dilute aqueous solutions of dissociated electrolytes.

5. The capillarity phenomena. They are, according to Valson[125] additive properties of solutions of salts. However, since they can be traced back to the specific gravity they required no further justification.

6. The freezing point depression. Its additivity by salts in water was discussed in 1885 by Raoult.[126] Some properties are proportional to the freezing point depression. Guldberg[127] and van ´t Hoff[116] showed this for the lowering of the vapor pressure, and the osmotic pressure, Hugo de Vries for the isotonic coefficient.[128]

With this additional evidence for his theory, Arrhenius completed his seminal paper.

Ostwald promoted Arrhenius´ dissociation theory, nonetheless many chemists initially disapproved it. Their objections were mainly the same which were already directed against Clausius' theory, and which Arrhenius had tried to present as untenable in his dissertation in Part 2 on pages 6 and 31. In passing, it should be mentioned that Ostwald first actively defended the theory in 1888 in his articles in volume 2 of the same journal,[129, 130] in 1889 with W. Nernst in the joint paper "Über *freie Jonen*" (On free Ions) in volume 3,[131] and in many other papers and at various occasions. We have described in Part 3, p. 97-98, of the present series[5] that in this 1889 paper capillary electrophoresis was performed in one of four occasions in the entire 19th Century.

Beside his pioneering theory, Arrhenius contributed with numerous publication on this subject. In the years from 1887 to 1889, he published eight papers.[112, 132-139] in addition to the following attacking "*Electrolytic Dissociation versus Hydration*"[140] from 1889. It was addressed to the proponents of the prevailing hydrate theory, English chemists who vehemently rejected his theory.[29] This "early" hydrate theory dates back to the 1810s and is probably not known in detail to today's chemists. It is not to be confused with the modern hydrate theory, which is now known to every chemist and physical chemist.

The followers of the early hydrate theory, most notably Henry Edward Armstrong and Spencer Umfreville Pickering, considered Arrhenius' assumption that anhydrous ions exist in aqueous solutions to be incompatible with their theory.[30] They suspected that, especially in highly dilute solutions with their extreme excess of water, the dissociation theory was in conflict with the law of mass action.

In 1903, about fifteen years after the publication of his theory, Arrhenius received the Nobel Prize in Chemistry "*in recognition of the extraordinary services he has rendered to the advancement of chemistry by his electrolytic theory of dissociation*."[31] It is therefore somewhat surprising that some advocates of the hydrate theory held to their strict rejection of electrolyte dissociation decades after the Nobel Prize was awarded. We will not finish thus the story about this rejection without mentioning that, as reported in ref. [141], p. 1559, even in the 1930s Prof. Louis Albrecht Kahlenberg[32] taught electrochemistry at the University of Wisconsin disregarding the existence of ions in solutions.

Early hydrate theory and transformation of its parts into the modern one arguably deserved its own historical overview because of the decades-long importance of the former and its role in combating dissociation theory, especially since it has fallen into oblivion.

SUMMARY

When in the early 1880s the Swedish doctoral student Svante Arrhenius decided to investigate the electrical conductivity of highly diluted electrolyte solutions as the subject of his dissertation there were hardly any studies on this topic in the literature. Most of them were published by his 18 years older contemporary Friedrich Kohlrausch, who had been measured conductivities at

---

[29] The paper was communicated by himself in English; an exception, since he usually published in German, and preferably in Swedish in Kongl. Svenska Vetenskapsakademiens Handlingar, Bihang tili Kongl. Vetenakapsakademiens Handlingar and in Öfversigt öfver Kongl. Vetenskapsakademiens Förhandlingar.

[30] Indeed Arrhenius stated verbatim on p. 32 of his 1889 pamphlet [author´s note: *n* is the number of water molecules that form complexes of defined stoichiometry with a salt molecule]: "But as we have no ground for attributing any particular value to *n*, and as it is besides probable that many salts (e. g. most of those of potassium) exist only in the anhydrous state, the simplest and likeliest assumption is that the ions of the salts, and consequently the salts themselves, exist in solution without water of hydration."

[31] To Wilhelm Ostwald the Nobel Prize in Chemistry 1909 was awarded "*in recognition of his work on catalysis and for his investigations into the fundamental principles governing chemical equilibria and rates of reaction.*"

[32] Louis Albrecht Kahlenberg (1870 – 1941), a US American chemist, was professor of physical chemistry at the University of Wisconsin, where he taught and carried on research for forty-seven years until his retirement in 1940. Beside others, his scientific area was ion conductivity in non-aqueous solutions. Politically, he was opponent of America's entry into World War I, which was unnecessary in his opinion. He was a doctoral student of Wilhelm Ostwald at the University of Leipzig, where he received a PhD, but rejected Arrhenius's dissociation theory.

higher concentrations, and extended his measurements to highly dilute solutions in these years.

At first glance, the completely different ways in which Arrhenius and Kohlrausch conducted their experiments and interpreted their results are striking. Kohlrausch was known for his extremely careful execution of his experiments, for the search and elimination of possible sources of error, and the associated time-consuming and labor-intensive investigation of the solutions with the intention of obtaining the most accurate measurement results possible. These exceptionally accurate data allowed him to derive empirical equations from them (see, e.g., ref. [5]) Arrhenius carried out his conductivity measurements in a contrary manner. He did not attach any particular importance to the high accuracy of his measurements, but rather intuitively detected a certain tendency from the data obtained, from which he derived hypotheses rather than empirical laws.[33]

As part of his dissertation Arrhenius measured the electrical resistances of dilute solution from about fifty different electrolytes, and calculated their molecular conductivities as a function of their dilution. He summarized his observations and his conclusions in 56 theses, of which we quote some of the most important. He recognized that the strongest acids – he examined only 5 – were those with the highest molecular conductivity. He observed that the molecular conductivity increased with increasing dilution, and he distinguished two different groups of electrolytes. In the first group, the strong electrolytes, the molecular conductivity increased very little and almost linearly with dilution, approaching the maximum value under limiting conditions. It was assumed that this increase was due to decreasing frictional resistance between the electrolyte molecules as their distances increased with dilution. The second group, the weak electrolytes, behaved completely differently. Their molecular conductivity initially remained at a low level with increasing dilution, but increased abruptly at sufficient dilution. Arrhenius could not initially explain this deviant behavior because there was no reason to attribute it to the frictional effects mentioned above.

Kohlrausch remained essentially with the investigation of the strong electrolytes. However, the said deviant behavior aroused the special of interest of Arrhenius. To explain this effect, he made the bold hypothesis that the number of conducting molecules had to increase by splitting the electrolyte molecules into two parts. One part is electrically active, it conducts electricity, the second, electrically inactive part is non-conducting. He further hypothesized that with increasing dilution, the proportion of active parts increases at the expense of the inactive ones. At infinite dilution, all inactive parts are completely dissociated, a verb Arrhenius avoided to use in his dissertation. Arrhenius introduced the activity coefficient (later called degree of dissociation), which indicates the proportion of active molecules to the number of all molecules, if these were completely dissociated. The active parts must be capable of double decomposition, since a permanent exchange with those of other molecules takes place, which – in the absence of an electric potential – leads to a circular current of the ions in the solution. Arrhenius' hypothesis explained both, the high molecular conductivity of the strong acids, as they consist of the electrically active molecules, and the behavior of the weak acids with their initially small activity coefficient that increases sharply with increasing dilution.

To confirm his hypothesis, Arrhenius borrowed results from thermochemistry, especially those of the Dutch physicist Julius Thomsen. Thomsen had determined the heats of neutralization of acids in their reaction with a strong base. He found that strong acids evolve the largest heats of neutralization, weak acids develop smaller heats. These results were consistent with the Arrhenius hypothesis, since strong acids, like strong bases, already consist of the active H^+ and OH^- ions (their counterions do not contribute to the neutralization), which combine directly to form the inactive water and release its activation energy. In the case of weak acids, in contrast, the inactive part must be transferred into the active part for neutralization, a process that absorbs heat and reduces the heat balance accordingly. This was the first proof of the Arrhenius hypothesis, which was based on a non-electrical thermochemical phenomenon.

Arrhenius, however, was of the opinion that the conductivities measured by him and the heats of neutralization were not sufficient for a solid theory. Without conducting any further experiments himself, he found three more proofs to confirm his hypothesis. The next, the second proof resulted from Wilhelm Ostwald's studies of the affinities of different acids on bases. Ostwald's measurements concerned the kinetics of chemical reactions and the influence of the strength of the acids,

---

[33] This difference was emphasized by the above mentioned English inventor of the moving boundary method[89]. whom we also mentioned in Part 3 of our previous historical article. In a critical commentary in ref. [115] Lodge began his analysis – not very encouragingly for Arrhenius – with "*Whatever may have been the importance of the first part of this memoir at the date of its appearance (1883), the publication last October in Wiedemann's Annalen of a masterly memoir by Prof. F. Kohlrausch on the same subject throws it into the shade; for there can he no doubt that while the ground covered by both is similar, the Kohlrausch memoir is greatly superior, both in the experiments made and in the discussion upon them.*")

which determines their reaction rate. Among others, he investigated the effect of about forty different acids on the reaction rates of hydrolysis of methyl acetate, and on the inversion of cane sugar. After reading Arrhenius dissertation which he received in 1884, he concluded that the stronger an acid is, the more pronounced its affinity should be, as indicated by its molecular conductivity. Thus, Ostwald measured the molecular conductivities of diluted solutions of his acids, and found an astonishing high parallelism with the rate constants. This was the second proof of Arrhenius´ hypothesis: it was based on chemical kinetics.

Perhaps the most important piece of evidence was the analogy found by J. H. van ´t Hoff between the osmotic pressure exerted on a semipermeable membrane by a dilute solution and the pressure of a gas composed of the same number of particles. This pressure depends only on the number of dissolved particles, but is independent of their electric charge. The deviation of the osmotic pressure from that of Avogadro's law is expressed by van 't Hoff factor i, which is the number of all solute particles. On the one hand, i can be determined experimentally by Raoult's freezing point depression. On the other hand, it can be calculated with the aid of Arrhenius´ degree of dissociation, α. If the values of i obtained from both methods agree, the value for α is correct, and Arrhenius´ dissociation hypothesis is confirmed. The actual agreement found was the third proof. In addition, the additivity of physical properties, such as the specific gravities of solutions confirmed the hypothesis.

Based on these four proofs, Arrhenius had successfully developed the seminal theory "*Ueber die Dissociation der in Wasser gelösten Stoffe*" in 1887 from the collection of the 56 theses of the 1884 dissertation.

Finally, we would like to address an apparent discrepancy between the subject of our article series – the history of the basic principles of capillary electrophoresis – and that of the present Arrhenius dissociation theory. Although the most important properties studied by Arrhenius were the conductivities of electrolyte solutions, neither his dissertation nor his dissociation theory gives any indication of the associated mobility of ions in free solution under the influence of an electric field. Since this ion mobility is a central property in electrophoresis, one could assume at first glance that the dissociation theory has only little significance for electrophoresis. However, the opposite is the case, as it allows to express the conductivity and thus the mobility over the whole range from zero to its maximum value by means of the degree of dissociation, which can take all values between zero and unit. For this reason, the dissociation of weak electrolytes described by Arrhenius will certainly be of great relevance in the practice of ion electrophoresis. It will be indispensable for electrophoresis as a future separation method. In this respect, its detailed treatment in this article was justified.

REFERENCES

[1] R. Clausius, *Ann. Phys. Chem. (Pogg.)* **1857**, *101*, 338-360.
[2] R. Clausius, *Phil. Mag.* **1858**, *15*, 94-109.
[3] R. Clausius, *Ann. chim. phys.* **1858**, *53 [3]*, 252-256.
[4] E. Kenndler, *Substantia* **2021**, *5(2)*, 95-118.
[5] E. Kenndler, *Substantia* **2022**, *6(1)*, 77-105.
[6] S. Arrhenius, *Bihang Till K. Svenska Vet.-Akad. Handlingar* **1884**, *8 (13)*, 3-63.
[7] S. Arrhenius, *Bihang Till K. Svenska Vet.-Akad. Handlingar* **1884**, *8 (14)*, 3-89.
[8] S. Arrhenius, *Z. phys. Chem.* **1887**, *1*, 631–649.
[9] S. Arrhenius, in *Memoirs on the Modern Theory of Solution* (Ed.: H. C. Jones), Harper & Brothers, New York, London, **1888**, pp. 45-66.
[10] E. Edlund, *Ann. Phys.* **1868**, *210*, 337-355.
[11] E. Edlund, *Phil. Mag,* **1869**, *38*, 169-178.
[12] E. Edlund, *Phil. Mag.* **1878**, *6*, 289-306.
[13] E. Edlund, *Ann. Phys. Chem. (N. F., Wied.)* **1877**, *1*, 161-199.
[14] E. Edlund, *Ann. Phys. Chem. (N. F., Wied.)* **1878**, *3*, 489-494.
[15] F.-M. Raoult, *Compt. rend.* **1882**, *95*, 1030-1033.
[16] S. Arrhenius, *Bihang Kongl. Vetenskapsakademiens Handlingar* **1883**, *7*, 1-36
[17] F. Kohlrausch, O. Grotrian, *Götting. Nachr.* **1874**, *17*, 405-418.
[18] F. Kohlrausch, O. Grotrian, *Ann. Phys. Chem. (Pogg.)* **1875**, *154*, 1-14.
[19] F. Kohlrausch, O. Grotrian, *Ann. Phys. Chem. (Pogg.)* **1875**, *154*, 215-239.
[20] F. Kohlrausch, O. Grotrian, *Phil. Mag.* **1875**, *49*, 417-425.
[21] F. Kohlrausch, *Ann. Phys. Chem. (Pogg.)* **1876**, *159*, 233-275.
[22] F. Kohlrausch, *Götting. Nachr.* **1877**, *9*, 181-199.
[23] F. Kohlrausch, *Ann. Phys. Chem. (N. F., Wied.)* **1879**, *6*, 145-210.
[24] F. Kohlrausch, *Ann. Phys. Chem. (N. F., Wied.)* **1879**, *6*, 1-51.
[25] F. Kohlrausch, *Ann. Phys. Chem. (N. F., Wied.)* **1880**, *11*, 653-660.
[26] R. Lenz, *Ann. Phys. Chem. (Pogg.)* **1877**, *160*, 425-435.

[27] R. Lenz, *https://www.biodiversitylibrary.org/item/104844#page/1/mode/1up* **1877**, *23*, 250-279.

[28] R. Lenz, *Mém. Acad. Impér. Sci. St.-Petersbourg* **1878**, *26 [VII]*, 1-51.

[29] R. Lenz, *Mém. Acad. Impér. Sci. St.-Petersbourg* **1882**, *30 [7]*, 1-64.

[30] F. Kohlrausch, *Götting. Nachr.* **1876**, *10*, 213-224.

[31] F. Kohlrausch, *Ann. Phys. Chem. (Pogg.)* **1878**, *Erg. Bd. VIII*, 1-16.

[32] F. Kohlrausch, *Ann. Phys. Chem. (N. F., Wied.)* **1885**, *26*, 161-226.

[33] H. É. S.-C. Deville, *Compt. rend.* **1857**, *45*, 857

[34] E. Edlund, *Ann. Phys. Chem. (Pogg.)* **1875**, *156*, 251-278.

[35] E. Dorn, *Ann. Phys. Chem. (N. F., Wied.)* **1880**, *9*, 513-552.

[36] E. Dorn, *Ann. Phys. Chem. (N. F., Wied.)* **1880**, *10*, 46-76.

[37] E. Edlund, *Ann. Phys. Chem. (N. F., Wied.)* **1881**, *12*, 149-155.

[38] in *Report of the Fifty-Sixth meeting of the British Association for the Advancement of Science, held at Birmingham in September 1886*, **1886 (1887)**.

[39] K. F. Mohr, *Lehrbuch der chemisch-analytischen Titrirmethode. Nach eigenen Versuchen und systematisch dargestellt; für Chemiker, Ärzte und Pharmaceuten, Berg- und Hüttenmänner, Fabrikanten, Agronomen, Metallurgen, Münzbeamte etc.*, 1 ed., Friedrich Vieweg und Sohn, Braunschweig, **1855**.

[40] W. Hittorf, *Ann. Phys. Chem. (Pogg.)* **1859**, *106*, 513-586.

[41] M. Berthelot, *Chimie organique fondée sur la synthèse, Vol. 1*, Mallet-Bachelier, Paris, **1860**.

[42] M. Berthelot, *Chimie organique fondée sur la synthèse, Vol. 2*, Mallet-Bachelier, Paris, **1860**.

[43] M. Berthelot, *La synthèse chimique*, Germer Baillière, Paris, **1876**.

[44] M. Berthelot, *Essai de mécanique chimique fondée sur la thermochimie. Tome I. Calorimétrie, Vol. 1*, Dunod, Paris, **1879**.

[45] M. Berthelot, *Thermochimie. Données et Lois Numériques. Tome I. Les Lois Numériques, Vol. 1*, Gauthier-Villars et Fils, Paris, **1897**.

[46] M. Berthelot, *Thermochimie. Données et Lois Numériques. Tome II. Les Données Expérimentales, Vol. 2*, Gauthier-Villars et Fils, Paris, **1897**.

[47] M. Berthelot, *Science et Philosophie*, Calmann Lévy, Paris, **1886**.

[48] M. Berthelot, *Les Origines de l´Alchimie*, Georges Steinheil, Paris, **1885**.

[49] M. Berthelot, *Collection des Anciens Alchimistes Grecs, Vol. 1-3*, Georges Steinheil, Paris, **1887-1888**.

[50] M. Berthelot, *Histoire des sciences. La chimie au moyen âge, Vol. 1-3*, Imprimerie Nationale, Paris, **1893**.

[51] M. Berthelot, *Essai de mécanique chimique fondée sure la thermochimie. Tome II. Mécanique, Vol. 2*, Dunod, Paris, **1879**.

[52] O. J. Lodge, in *Report of the Fifty-Fifth meeting of the British Association for the Advancement of Science, Aberdeen, September 1885, Vol. Aberdeen, September 1885*, London, **1885**, pp. 723-772.

[53] T. v. Grotthuß, *Ann. Phys. (N. F., Gilb.)* **1819**, *1*, 54-70.

[54] A. W. Williamson, *Phil. Mag.* **1850**, *37*, 350-356.

[55] S. Arrhenius, Proc. Roy. Inst., 1904, **1903**.

[56] P. L. Dulong, *Ann. Phys. Chem. (Pogg.)* **1838**, *45*, 461-467.

[57] G. H. Hess, *Bull. scient. Acad. Impér. Sci. St.-Pétersbourg* **1840**, *8*, 257-272.

[58] G. H. Hess, *Compt. rend.* **1840**, *10*, 759–763.

[59] H. Hess, *Ann. Phys. Chem. (Pogg.)* **1841**, *52*, 97-114.

[60] H. Hess, *Ann. Phys. Chem. (Pogg.)* **1841**, *52*, 107-114.

[61] H. Hess, *Ann. Phys. Chem. (Pogg.)* **1841**, *53*, 535-547.

[62] H. Hess, *Ann. Phys. Chem. (Pogg.)* **1841**, *53*, 499-512.

[63] H. Hess, *Ann. Phys. Chem. (Pogg.)* **1842**, *56*, 593-604.

[64] H. Hess, *Ann. Phys. Chem. (Pogg.)* **1842**, *56*, 463-479.

[65] H. Hess, *Ann. Phys. Chem. (Pogg.)* **1842**, *57*, 569-584.

[66] Abria, *Ann. phys. chim.* **1844**, *12 [3]*, 167-176.

[67] P.-A. Favre, J.-T. Silbermann, *Compt. rend.* **1849**, *28*, 627-632.

[68] P.-A. Favre, J.-T. Silbermann, *Compt. rend.* **1847** *24* 1081-1090.

[69] T. Andrews, *Ann. Phys. Chem. (Pogg.)* **1845**, *66*, 31-57.

[70] T. Andrews, *Ann. Phys. Chem. (Pogg.)* **1848**, *75*, 244-255.

[71] T. Andrews, *Ann. Phys. Chem. (Pogg.)* **1848**, *75*, 27-50.

[72] T. Andrews, *J. prakt. Chem. (Erdm., March.)* **1850**, *2*, 468-493.

[73] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1853**, *88*, 349-363.

[74] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1853**, *90*, 261-288.

[75] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1854**, *92*, 34-57.

[76] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1854**, *91*, 83-104.
[77] M. Berthelot, *Ann. chim. phys.* **1869**, *18 [4]*, 5-108.
[78] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1869**, *138*, 65-102.
[79] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1869**, *138*, 201-213.
[80] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1869**, *138*, 497-514.
[81] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1870**, *140*, 88-114.
[82] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1870**, *140*, 497-540.
[83] J. Thomsen, *Ann. Phys. Chem. (Pogg.)* **1870**, *139*, 193-224.
[84] J. Thomsen, *Thermochemische Untersuchungen. I. Neutralisation und verwandte Phänomene, Vol. 1*, J. A. Barth, Leipzig, **1882**.
[85] J. Thomsen, *Thermochemische Untersuchungen. II. Metalloide, Vol. 2*, J. A. Barth, Leipzig, **1882**.
[86] J. Thomsen, *Thermochemische Untersuchungen: III. Wässrige Lösung und Hydratbildung. Metalle, Vol. 3*, J. A. Barth, Leipzig, **1883**.
[87] J. Thomsen, *Thermochemische Untersuchungen. IV. Organische Verbindungen, Vol. 4*, J. A. Barth, Leipzig, **1886**.
[88] in *Report of the Fifty-Sixth meeting of the British Association for the Advancement of Science, held at Birmingham in September 1886*, **1886 (1887)**.
[89] O. J. Lodge, *Report of the Fifty-Sixth Meeting of the British Association for the Advancement of Science, Birmingham, September 1886* **1886**, *56*, 389-412.
[90] W. Ostwald, *Lebenslinien. Eine Selbstbiographie, Vol. 1-3*, Klasing, Berlin, **1926/27**.
[91] W. Ostwald, *J. prakt. Chem.* **1879**, *19 [2]*, 468-484.
[92] W. Ostwald, *J. prakt. Chem.* **1880**, *22 [2]*, 251-260.
[93] W. Ostwald, *J. prakt. Chem. (N. F., Kolbe, Meyer)* **1881**, *23*, 517-536.
[94] W. Ostwald, *J. prakt. Chem. (N. F., Kolbe, Meyer)* **1881**, *24*, 486-497.
[95] W. Ostwald, *J. prakt. Chem. (N. F., Kolbe, Meyer)* **1881**, *23*, 209-227.
[96] W. Ostwald, *J. prakt. Chem. (N. F., Kolbe)* **1877**, *16*, 385-423.
[97] W. Ostwald, *J. prakt. Chem. (N. F., Kolbe)* **1878**, *18*, 328-371.
[98] W. Ostwald, *J. prakt. Chem. (N. F., Kolbe, Meyer)* **1883**, *27*, 1-39.
[99] W. Ostwald, *J. prakt. Chem. (N. F., Kolbe, Meyer)* **1883**, *28*, 449-495.
[100] M. Berthelot, L. Péan de St. Gilles, *Recherches sur les affinités. De la formation et de la décomposition des éthers*, Mallet-Bachelier, Paris, **1862**.
[101] M. Berthelot, L. Péan de St. Gilles, *Ann. chim. phys.* **1862**, *65 [3]*, 385-422.
[102] M. Berthelot, L. Péan de St. Gilles, *Ann. chim. phys..* **1862**, *66 [3]*, 5-110.
[103] M. Berthelot, L. Péan de St. Gilles, *Ann. chim. phys.* **1863**, *68 [3]*, 225-359
[104] W. Ostwald, *J. prakt. Chem. (N. F.; Kolbe, Meyer)* **1884**, *29*, 385-408.
[105] L. Wilhelmy, *Versuch einer mathematisch-physikalischen Wärme-Theorie*, Akademische Anstalt für Literatur und Kunst, Heidelberg, **1851**.
[106] L. Wilhelmy, *Ann. Phys. Chem. (Pogg.)* **1850**, *81*, 413-428.
[107] L. Wilhelmy, *Ann. Phys. Chem. (Pogg.)* **1850**, *81*, 499-526.
[108] J. Löwenthal, E. Lenssen, *J. prakt. Chem.* **1862**, *85*, 321-358.
[109] J. Löwenthal, E. Lenssen, *J. prakt. Chem.* **1862**, *85*, 401-416.
[110] W. Ostwald, *J. prakt. Chem. (N. F.; Kolbe, Meyer)* **1884**, *30*, 93-95
[111] G. Arrhenius, K. Caldwell, S. Wold, in *2008 Annual Meeting of the Royal Swedish Academy of Engineering Sciences*, The Royal Swedish Academy of Engineering Sciences, **2008**, pp. 6-44.
[112] S. Arrhenius, *Z. phys. Chem.* **1888**, *2*, 284-295.
[113] M. Traube, *Arch. Anatom. Physiol.* **1867**, 87-165.
[114] M. Traube, *Gesammelte Abhandlungen*, Mayer & Müller, Berlin, **1899**.
[115] W. Pfeffer, *Osmotische Untersuchungen. Studien zur Zellmechanik*, 1 ed., Wilhelm Engelmann, Leipzig, **1877**.
[116] J. H. van ´t Hoff, *Kongliga Svenska Vetenskaps-Akademiens Handlingar* **1884-1885**, *21*, 1-58.
[117] J. H. van ´t Hoff, *Archives Néerlandaises* **1886**, *20*, 29-302.
[118] J. H. van 't Hoff, *Z. phys. Chem.* **1887**, *1*, 481–508.
[119] M. Planck, *Z. phys. Chem.* **1887**, *1*, 577-582.
[120] M. Planck, *Z. phys. Chem.* **1890**, *6*, 187-189.
[121] L. Boltzmann, *Z. phys. Chem.* **1890**, *6*, 474-480.
[122] F. Kohlrausch, *Ann. Phys. Chem. (N. F., Wied.)* **1885**, *24*, 48-52.
[123] C.-A. Valson, *Compt. rend.* **1871**, *73*, 441-443.
[124] J. H. Gladstone, *Phil. Mag.* **1868**, *36 (4)*, 313-316.
[125] C.-A. Valson, *Compt. rend.* **1872**, *74*, 103-105.
[126] F.-M. Raoult, *Ann. chim. phys.* **1885**, *4 [6]*, 401-430.
[127] M. Guldberg, *Compt. rend.* **1870**, *70*, 1349-1352.
[128] H. de Vries, *Pringsheims Jahrbücher für wiss. Botanik* **1884**, *14*, 427-601.
[129] W. Ostwald, *Z. phys. Chem.* **1888**, *2*, 270–283.
[130] W. Ostwald, *Z. phys. Chem.* **1888**, *2*, 36-37.
[131] W. Ostwald, W. Nernst, *Z. phys. Chem.* **1889**, *3*, 120-130.

[132] S. Arrhenius, *Z. phys. Chem.* **1887**, *1*, 110-133.
[133] S. Arrhenius, *Z. phys. Chem.* **1887**, *1*, 285-298.
[134] S. Arrhenius, *Ann. Phys. Chem. (N. F., Wied.)* **1887**, *30*, 51–76.
[135] S. Arrhenius, *Ann. Phys. Chem. (Wied.)* **1887**, *32*, 545.
[136] S. Arrhenius, *Z. phys. Chem.* **1888**, *2*, 491-505.
[137] S. Arrhenius, *Z. phys. Chem.* **1889**, *4* 96-116.
[138] S. Arrhenius, *Z. phys. Chem.* **1889**, *4*, 226-248.
[139] S. Arrhenius, *Z. phys. Chem.* **1889**, *3*, 115-119.
[140] S. Arrhenius, *Phil. Mag.* **1889**, *28 [5]*, 30-38.
[141] K. J. Laidler, *Can. J. Chem.* **1997**, *75*, 1552-1565.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: Michálek J., Podešva J., Dušková-Smrčková M. (2022) True Story of Poly(2-Hydroxyethyl Methacrylate)-Based Contact Lenses: How Did It Really Happen. *Substantia* 6(2): 79-91. doi: 10.36253/Substantia-1591

Received: Mar 01, 2022

Revised: Jun 26, 2022

Just Accepted Online: Jun 27, 2022

Published: September 1, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

True Story of Poly(2-Hydroxyethyl Methacrylate)-Based Contact Lenses: How Did It Really Happen

Jiří Michálek, Jiří Podešva, Miroslava Dušková-Smrčková*

Institute of Macromolecular Chemistry, Czech Academy of Sciences, Heyrovsky Sq. 2, 162 06 Prague 6, Czech Republic
E-mail: jiri@imc.cas.cz, podesva.ji@seznam.cz, m.duskova@imc.cas.cz
*Corresponding author

Abstract. Soft hydrogel contact lenses represent the most famous and commercially successful application of poly(2-hydroxyethyl methacrylate). The scarcely crosslinked network of this hydrophilic polymer finds its use also in many other fields, be it in (bio)medicine or technology. Moreover, the polymer itself and its crosslinked forms, discovered more or less serendipitously in the early fifties by a group of Czech chemists, is extremely interesting due to its exceptional properties: it readily swells in water, is optically clear, soft, biologically compatible, sufficiently strong, stable, gas-permeable, cheap, and easy to produce. Looking for its as-yet undiscovered qualities and possible utilization still continues. The story of the invention of hydrogel contact lenses was referred to many times in various literary sources which, however, contain numerous errors and misinterpretations. In the present article, we put these records straight and present the correct chronology of the hydrogel contact lenses development including the dramatic patent litigation. A brief overview of the chemical nature, properties, and applications of the constitutive substance of the lenses, i.e., the hydrophilic methacrylate, is also given.

Keywords: hydrogels, contact lenses, intraocular lenses, poly(2-hydroxyethyl methacrylate), Otto Wichterle.

1. INTRODUCTION

Modern hydrogels are usually tailor-made for the given purpose and application, be they synthesized by radical-initiated or stepwise processes, performed in a standard way, or by 3D printing. Since the times of the invention of the first hydrophilic plastic “swellable Perspex”, prepared by O. Wichterle’s group in the 1960s using the radical polymerization of 2-hydroxyethyl methacrylate (HEMA),[1] much effort has been devoted to a detailed study of this polymer. This was due both to its use for pioneering hydrogel contact lenses (the so-called “swelling plastic”) and to its interesting properties. Poly(2-hydroxyethyl methacrylate) (PHEMA) is distinguished by a good swellability (primarily in hydrophilic and partially also in hydrophobic media) and by a very good compatibility with living tissues. Even after swelling in aqueous media it keeps its mechanical strength and flexibility and is stable in time. That is why this material has found so many applications. Besides the medicinal use in the fields of ophthalmology, implants, or systems for drug transport and releasing, there are less known but no less successful uses for sorbents with a large intrinsic surface or separation monoliths in chromatography.[2-4] Thus, PHEMA remains a subject of lively scientific interest, as indicated by the number of papers with this keyword, published every year. At the same time, it represents an important model polymer both for the scientific research of synthetic hydrogels and for biomedicinal applications, including testing experiments of tissue engineering.

This paper brings information on the history of the research and applications of this unique monomer and its polymers, with special regard to hydrophilic contact lenses. It is the authors' ambition to put some erroneous historical data straight. Moreover, we consider it useful to briefly outline also the classification and history of the whole phenomenon of contact lenses.

2. EXCITING HISTORY OF CONTACT LENSES IN GENERAL

What is the contact lens? The basic definition reads: *Contact lens is a small optical system placed directly on the cornea.* All the issues and problems related to the contact lenses follow therefrom.

Contact lenses can be categorized in various ways. However, according to M. F. Refojo,[5] the fundamental division is based on the nature of the material. Most simply, contact lenses could be distinguished into rigid ones and soft ones, the latter then into hydrophobic and hydrophilic. Further categorization, necessary in connection with the development of new materials for contact lenses, is given in more detail in the Appendix (Tab. I). In current sources, this division is, regrettably, often oversimplified.

The idea of contact lenses is very old, reaching back as far as the 16th century and Leonardo da Vinci concepts, and its implementation is closely connected with the development of material science. Various inventors tried to use a broad spectrum of materials for contact lenses.

For example, when poly(methyl methacrylate) (PMMA) was introduced into the market (1933) and its relatively good biocompatibility was discovered, a way was opened for new medicinal applications of this plastic. Thanks to its optical properties, PMMA found its main use in ophthalmology (as a material for contact lenses, later for intraocular lenses, spectacles, etc.). This was the beginning of the era of polymers or covalent polymer networks in contactology, a brief history of which is presented in a tabulated form in the Appendix (Tab. II).[6-9]

After PMMA had been tested and finally abandoned, the following development of contact lenses was carried out to improve the properties of the lenses, namely, their permeability for gases (primarily oxygen) and also for water-soluble substances and ions. Although both of these requirements were met excellently by hydrogels studied by Wichterle and Lím,[1] another branch of the research continued towards the silicone elastomers (1965) which offered a high permeability for gases and showed good softness but were hydrophobic. These properties were then responsible for problems met when removing these lenses from the eye, namely, mechanical damage to a testing person's cornea. As a consequence of this, contact lenses based purely on silicone hydrophobic elastomers are no more accessible in the common market.[10]

Still another route of the development resulted in rigid gas-permeable (RGP) materials (1974), usually copolymers of alkyl methacrylates and siloxane methacrylates (possibly also fluoroalkyl methacrylates) which guarantee a high permeability for oxygen[11] but are hydrophobic and do not allow the transport of water-soluble substances.

Diverse variants of high-swelling hydrogels for contact lenses have continuously been being developed which had, in dependence on the equilibrium water content, a higher permeability for both water-soluble substances and gases. In addition to the basic sparsely crosslinked PHEMA, other glycol methacrylates were used, such as diethylene glycol methacrylate, triethylene glycol methacrylate, dihydroxyalkyl methacrylates (e.g., glycerol methacrylate), acrylamide, and, for ionogenic materials, also methacrylic acid sodium salt. Besides the acrylic acid derivatives, also 1-vinyl-2-pyrrolidone and polyvinylalcohol found their use as materials for high-swelling hydrogel contact lenses.[12]

Thus, in the sixties and seventies, the development headed toward soft contact lenses based on PHEMA or similar hydrophilic methacrylates, as will be discussed below. Later, however, silicone hydrogel lenses of the first generation were developed and introduced (1998-1999, according to the territory) and became an important milestone. Based on the first experience, the second generation arrived in 2004 and soon after (2006) even the third one. Interestingly, the first relevant patent dates back to 1979.[13]

3. TRUE STORY OF SOFT PHEMA-BASED CONTACT LENSES

3.1. Origins of the idea

The story of the origin of PHEMA-based contact lenses from the primal idea to the invention itself and its putting into practice seems to be generally known. The discovery of the synthetic hydrogel based on sparsely crosslinked PHEMA and its successful application as a biomimetic material for soft contact lenses are often mentioned in introductory parts of scientific papers. Similarly, the pioneering article by Wichterle and Lím[1] on the unexpected hydrophilic behavior of certain plastics and future possibilities of their biological applications, as well as the corresponding patents (see, e.g.,[14]) are frequently cited, too. However, although the history of the development of PHEMA, its polymerization, and properties, as well as hydrogel lenses based on it, has been published many times in various literary sources, the interpretations very often digress from reality. Hence, the following chapter aims to bring a systematic survey of events that led to the worldwide known invention and to the subsequent global development of soft contact lenses. The text is based on reviewed sources, Otto Wichterle's book of memoirs,[15] and a personal experience of the first author, i.e., his collaboration with the famous inventor for fifteen years.

The primary impulse arose from a fortuitous meeting of Prof. Wichterle with Dr. Pur, the secretary of a certain committee for the application of plastics in medicine at the Czechoslovak Ministry of Health Care. By coincidence, in 1953, they traveled together by train and looked through an ophthalmological journal with an advertisement for a tantalum prosthesis to substitute the eyeball. As he later mentioned in his memoirs,[15] Wichterle had expressed an opinion that it would be more suitable to prepare such implants from biocompatible polymers and suggested an idea of three-dimensional sparsely crosslinked hydrophilic gels.

This idea attracted Wichterle's attention so much that he started to put it immediately into practice in the Department of Plastics at the then Czech Technical University in Prague, together with his younger colleagues, especially Drahoslav Lím. At that time, research on methacryloyl derivatives of oligoethylene glycol was already running with the aim to get new plastics for future biomedical applications. The first hydrogel prepared and identified by D. Lím was crosslinked triethylene monomethacrylate, as described in a paper by J. Kopeček.[16] Later, as mentioned in another paper by Kopeček *et al.*,[17] in 1953 D. Lím succeeded in synthesizing the first hydrogels by the copolymerization of HEMA with ethylene dimethacrylate. In the same year Wichterle, as the only inventor, submitted a patent application for an invention, in which he claimed the whole class of sparsely crosslinked hydrophilic polymers including a description of many potential uses including even contact lenses unless he (or whoever else) had prepared this material.[15, 18] Of course, this was a pure fantasy at that time but, as it turned out later, also a realistic prophecy. Later on, this application was withdrawn and substituted by another one[19], which finally led to a patent entitled "The way of preparation of hydrophilic gels".[20] In the meantime, however, patents were granted to translated versions of the applications with differing delays in various territories. For example, in Great Britain and the then Federal Republic of Germany, it was granted still to the earlier application from 1953, while in other countries already to the one from 1955. That is why various literary sources differ in dating the origin of hydrogel lenses.

Since 1956 the contact lenses have been being prepared in Wichterle's lab in Prague but their ridges were of poor quality so testing persons were able to tolerate them on their eyes only for a few minutes at most. In the meantime, however, part of the applied research was transferred under the supervision of the Ministry of Health (Dental Laboratory, Prague). Several good lenses could have eventually been selected from the production of this laboratory where they were being prepared in polystyrene molds (1957).

The tests on patients (performed in the 2nd Ophthalmology Clinic at the General University Hospital in Prague, Mr. Dreifus, M. D.) proved that the soft hydrophilic lenses, prepared on a lab-scale but using ground glass molds, can ensure a very good correction of vision and are excellently tolerated (1959).

We quote here from the paper cited above (entitled „Hydrophilic Gels for Biomedical Use"):[1] "Promising results have also been obtained in experiments in other cases, for example, in manufacturing contact lenses, arteries, etc." That is why some sources proclaim 1960 as the year of the origin of soft hydrophilic lenses. Till today, this publication has been cited almost 1100 times.

However, most authors consider 1961 to be a true year of the origin of the hydrogel lenses. At the end of December 1961, prof. Wichterle, using a Czech-made children's toy building set *Merkur* (similar to the well-known *Erector Kit*), assembled at his home a device for the spin casting of contact lenses and named it (with his typical sense of humor) the "lens-machine" (Fig. 1, left). The principle of the spin casting consists in that the starting liquid polymerization mixture, dropped into a mold with a precise inner shape, is rotated by fine-tuned

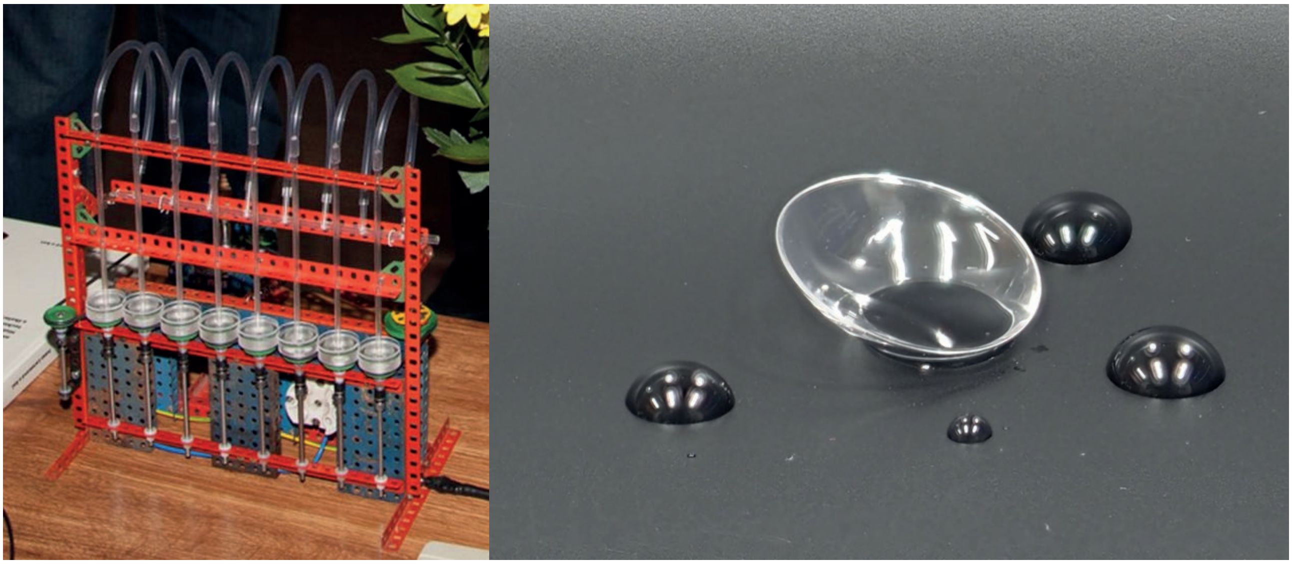


Figure 1. Replica of the building set *Merkur* (improvised lens-machine) for spin-casting (left), an example of a soft hydrophilic PHEMA-based contact lens (right).

number rpm. Due to a combination of the mold shape, the centrifugal force, and the surface tension, a proper lens shape is formed and, after the polymerization is finished, the solid contact lens acquires also the desired optical properties. With this improvised pilot-plant device, the first hydrogel contact lenses were produced (Fig. 1, right).

Later on, but still before the end of the same year, Wichterle patented a method to produce contact lenses.[21] In this way, the patents protecting the material for contact lenses were complemented by those describing the production method and the foretold use of synthetic hydrogels for contact lenses came into existence. A typical appearance of a contact lens is in Fig. 1.

A meeting with G. Nissel, a British producer of lathes and facilities for lathe-cutting of hard contact lenses, inspired Prof. Wichterle to submit another patent application of the invention to produce soft hydrogel lenses by turning from xerogel blocks, i.e., from prefabricated parts constituted by hydrogel in a dry state (Fig. 2), followed by fine polishing and swelling the lathed lenses.[22]

In 1964 Prof. Wichterle met his license partners-to-be from the National Patent Development Corporation (NPDC, USA). During the negotiations, he took out a lens from his eye, put it down to the ground, trampled it, then picked it up, removed the dirt from it first by fingers and then in his mouth, and finally put it back on his eye. This impressed his guests enormously. In 1965, the first license deal was signed between the then Czechoslovak Academy of Sciences and NPDC. Later on, in 1966, NPDC transferred the sub-license for soft contact lenses to Bausch & Lomb Co. which started to produce them in the USA, to prepare the distribution network and the marketing support, while waiting for approval of the production from the Food and Drug Administration (FDA). This was granted as late as 1972 but thanks to thorough preparation, Bausch & Lomb quickly penetrated the market and met a considerable demand for lenses.

3.2. Fascinating lawsuit on the patent priority

Already at the beginning of the seventies, infringements of Wichterle's patents by some producers appeared and even the Bausch & Lomb Co. took part in the litigations to save money for license fees. They used a tactic of denying the validity of Wichterle's patents with an argument of alleged pre-publication of some results and an absence of clinical tests. After NPDC had requested Wichterle's personal participation and testimony in American courts, the lawsuits began. To make the long story short, we set aside complications and obstacles laid by Czech communist authorities to block Wichterle's travel to the USA. Fortunately, he was allowed to testify in the end.

These legal disputes stretched till the beginning of the 80ies, although, thanks particularly to Wichterle's unambiguous replies to questions, became increasingly obvious that the validity of the patents will be confirmed.

By the end of 1976, despite this promising course, the Czech side acceded to an out-of-court settlement,

Figure 2. Special lathe for contact lens manufacturing (left) and the lathing of the contact lens from xeroblock.

and, for receiving an amount equal to the license income for one year, the Czechoslovak Academy of Sciences, controlled by the communist regime, stupidly opted out of the contractual liability for the participation in the patent lawsuits. In this way, the Czech side forfeited not only the license contracts but also the share of the proceeds of the lawsuit.

In 1980, a radical turnaround happened in the lawsuit which meant a full victory because all disputed issues were explained and Dr. Dreifus, who had been apparently manipulated by the infringers, was convicted of false testimony.

Still, it had taken two years of thrilling waiting before the final verdict was delivered (1982). In the meantime, still in 1981, NPDC made, probably as an expression of gratefulness to O. Wichterle for his contribution to the victory at the Court, a new license contract regarding the preparation of contact lenses by a photopolymerization initiated by UV radiation.[20-22] License fees from this contract have been coming to the Czech Republic till 2000.

3.3. Further development

Simultaneously with improving the quality of the contact lenses, also the means of maintenance of them had to be adapted to the newly developed materials. Thus, the physiological solution, used in the beginning, was substituted by multipurpose solutions containing, e.g., a disinfection or conservation component, a buffer system, detergents, wetting agents, and auxiliary substances, such as those with chelating effects. Similarly, the regime of wearing the lenses, as well as the planning replacement of them (rate), have been developing. In this way, the development resulted in disposable lenses.

In the nineties (1993) a one-time non-recurring contract was made with South Korean partners who took over a new lens-making machine ("lens machine") of the carousel type with an electronic-pneumatic control of functions and documentation for innovative technological processes including a new version of the software (Fig. 3).

Although the Koreans paid for a corresponding part of the charges, they never started to produce so the fees derived from the number of pieces produced were never received by the Czech side.

Prof. Wichterle's decease in 1998 symbolically closed the era of the early development of PHEMA-based hydrogel contact lenses. In the same year, the first "silicone hydrogels", constituted partly of polysiloxane chains, were introduced into the market. The polysiloxane structure, hydrophobic by nature, is made sufficiently hydrophilic by the covalent attaching of methacryloylated segments and other hydro-

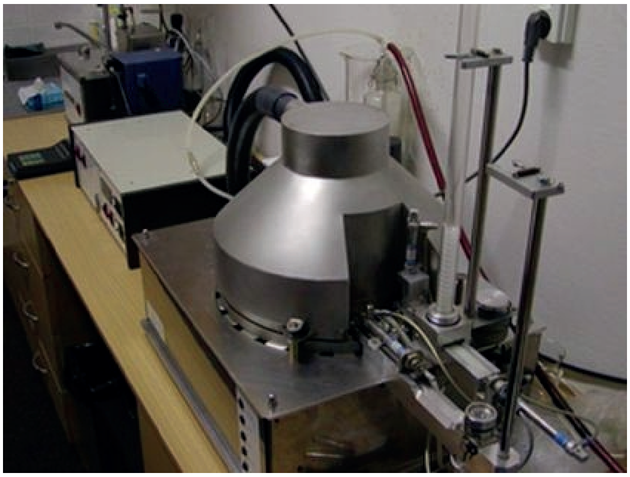


Figure 3. Lens machine for spin casting, the carousel type from the nineties.

philic vinylic polymers. [23] Silicone hydrogel contact lenses arrived at their 3rd generation and the "tricks" of attaining hydrophilicity differ from generation to generation. The type Dailies Total One, which was introduced on the market in 2012, represents a unique type of lens with a swelling gradient. However, hydrogels based on polymethacrylates or poly(vinyl alcohol) still constitute a substantial part of the world's production of contact lenses. Supposedly, for some clients, they will remain a suitable variant of the ocular refraction defect correction. Innovations still appear, for instance, the product called Hypergel from Bausch & Lomb, which is a bio-inspired hydrogel material containing 78% of water and showing an increased oxygen permeability (Dk = 42 barrer). This multicomponent polymer formulated on the basis of HEMA, *N*-vinylpyrrolidone, and 2-hydroxy-4-*tert*.butyl-cyclohexyl methacrylate, and crosslinked by ethylene dimethacrylate and allyl methacrylate, contains also a UV stabilizer based on benzotriazole and incorporated in the chain by a methacryloyl substituent. Undesirable drying of the lens surface made of a highly swelling material is prevented by a block copolymer formed by two outer blocks of poly(ethylene oxide) and a central block of poly(propylene oxide). The copolymer is terminated on both ends by two methacrylate groups, through which it is incorporated into the structure of the whole polymer network. Contact lenses made from it were introduced in the market under the trademark Biotrue ONEday in 2014.

4. HEMA AND ITS POLYMERS

4.1. History of HEMA and PHEMA

The first notices on HEMA and its polymers date back to the Thirties, namely in the US patent No. 2,129,722 entitled *Esters of Methacrylic Acid* and registered on September 13, 1938, for John C. Woodhouse as the inventor and DuPont de Nemours Co. as the applicant.[24] In several claims (1-4), esters of methacrylic acid and a series of aliphatic diols, triols or pentaerythritol, etc. are generally presented; among these alcohols, also ethylene glycol is mentioned. Claim 8 is devoted solely to polymeric monomethacrylate prepared by heating the monomeric ester to 60-100 °C in the presence of dibenzoyl peroxide. Although the monomer, the polymer, and their preparations were thus described, a real utilization of them came as late as during the systematic study of the hydrophilic structures performed by Wichterle and Lím.[1,15]

4.2. Nomenclature, structure, and properties of the HEMA monomer

The most frequently used, non-systematic but the deep-rooted name is 2-hydroxyethyl methacrylate (usually acronymed as HEMA), sometimes also glycol methacrylate. Names like glycol monomethacrylate, hydroxyethyl methacrylate, ethylene glycol methacrylate, or 2-(methacryloyloxy)ethanol are also used. According to IUPAC, the systematic name is 2-hydroxyethyl-2-methylprop-2-enoate. To preserve intelligibility and to comply with the scientific community's common usage, the name 2-hydroxyethyl methacrylate (HEMA) is used throughout the text; similarly, ethylene glycol will be used instead of the systematic 1,2-ethanediol.

The structure of the monomer is presented in Fig. 4 together with its basic physical properties. If not stated otherwise, the values correspond to standard conditions, i.e., 25 °C and 101.325 kPa.[25]

4.3. Preparation of the HEMA monomer

Of the procedures to produce HEMA, two have been used on a larger scale. The Czechoslovak patent was based on the reesterification of methyl methacrylate by glycol.[26] This process led to a product with a relatively high content of diester (ethylene dimethacrylate causing a crosslinking during the polymerization), the concentration of which had to be decreased by subsequent purification procedures. In addition to that, the prod-

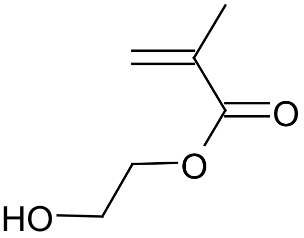


Figure 4. The schematic formula of 2hydroxyethyl methacrylate (properties: colorless liquid, density 1.07 g·cm^{-3}, melting point 99 °C, boiling point 213 °C, vapor pressure 0.08 hPa).

uct contained traces of diethylene glycol methacrylate and diethylene glycol dimethacrylate (the latter being a crosslinking agent, too) but was free of methacrylic acid.

Nowadays HEMA is commonly produced by a reaction of ethylene oxide with methacrylic acid. The resulting product contains a low level of the crosslinking agent and traces of methacrylic acid (see, e.g.,[27]).

4.4. Polymerization of HEMA

The double bond of 2-hydroxyethyl methacrylate reacts readily under normal pressure in bulk or in a solution, similarly to other methacrylates. The temperature range of the radical polymerization of HEMA has its upper limit at ca. 160 °C; at this and higher temperatures, depolymerization of the polymer chain takes place. Practically, the lower limit corresponds to the solidification (vitrification) temperature of the polymerizing system; however, it is possible to perform a redox-initiated polymerization under the condition of the so-called cryogelation, i.e., at sub-zero temperatures, e.g. around -20 °C and in presence of a diluent, when interesting macroporous structures are formed in the resulting gel thanks to freezing of the diluent (typically aqueous) off the system.[28] A living anionic polymerization of HEMA with a protected hydroxyl group has also been reported, [29,30] proceeding at much lower temperatures (40 to 80 °C) and yielding an isotactic polymer. In the latest decade, papers have been published reporting on the possibility to control the HEMA polymerization by the RAFT (reversible addition-fragmentation chain transfer)[31] or ATRP (atom transfer radical polymerization)[32] methods. It is the aim of these controlled radical polymerizations to get a polymer with the distribution of molar mass narrower than that obtained by standard (uncontrolled) free radical polymerization and to possibly attach certain functional groups onto the chain ends.

Interestingly, the sparsely crosslinked PHEMA (i.e., with the level of the crosslinker below ca. 1 mol.%) significantly swells in water attaining swelling equilibrium at approx. 36-38 wt.% of water at room temperature.[33] The swelling behavior of the PHEMA macromolecular network is very interesting and shows a certain "swelling anomaly": the equilibrium swelling degree does not depend much on the crosslink density which is also true for a linear PHEMA of a high degree of polymerization. PHEMA belongs to the UCST-LCST[1] system exerting swelling minimum at 55°C.[34]

4.5. Physical prerequisites for making the perfect contact lens

The PHEMA-based hydrogel suitable for lenses (PHEMA prepared with 38–40 wt.% of water and ca. 1 mol.% crosslinker) is characterized by some key properties such as the equilibrium content of water (approx. 38 wt.%), the oxygen permeability (8-12 x 10^{-11} barrer), and modulus of elasticity (typically 0.5-0.6 MPa).[8,28] However, these parameters strongly depend on the starting conditions and exact way of hydrogel preparation, especially on the concentration of the crosslinking agent and diluent (water) at polymerization. Here we focus solely on the microstructure and porosity. The PHEMA hydrogels can be prepared either as macroscopically homogeneous (optically transparent) or, inversely, as a heterogeneous substance, showing a loss of transparency and a formation of opalescence, thus indicating refraction of light on microscopic interphases due to the formation of pores. At this point, our report deserves a more detailed explanation of the PHEMA hydrogel optical clarity. In the early studies, when Wichterle and his coworkers observed the first crosslinked PHEMA gels, the pieces of water-swollen material were rather transparent and colorless. Their observations were truly serendipitous as the material resembled clear glass and provided an index of refractivity very close to that of the biological cornea, so the ideas about a gel-based soft contact lens could be explored ever since. But it soon became evident that not always the free radical crosslinking of the HEMA-based system leads to an optically clear material and that there are critical limits of composition beyond which the resulting material turns irreversibly hazy, or completely non-transparent – and thus not useful for an optical lens. These "clarity limits" for HEMA-based systems were subjected to thorough experimental studies in the Institute of Macromolecular Chemistry in Prague in the 1970s. It was found that when the content of water as a diluent in the polym-

---

[1] UCST – upper critical solution temperature, LCST – lower critical solution temperature

erizing system exceeds ca. 50 vol.%, an opaque or white, or even porous heterogeneous material is obtained. Indeed, the limits also correlated with the amount of crosslinker. The reasons for the existence of the limits were in the meantime explained by K. Dušek who put forward the analysis of the formation of thermodynamic phases leading to the porosity of the crosslinking system styrene-divinylbenzene investigated for ion exchange resins.[35] Deeper studies of PHEMA and its solution and gel properties continued in the seventies.[36] Dušek derived a generalized thermodynamic treatment for phase separation in a three-dimensional polymer system based on the analysis of the Flory-Huggins swelling equation and he coined the term microsyneresis (or syneresis). This term denotes a separation of phases in the so-called quasi binary system where the phase of the swollen gel separates from that of the diluent, the latter, however, possibly containing residua such as a soluble monomer or its oligomers. This separation is a consequence of the change of miscibility within the polymerizing system with conversion, so-called c-syneresis, and/or is induced by increasing crosslink density, so-called n-syneresis.[37] Whereas HEMA monomer is unlimitedly miscible with water (starting state), the growing chains only have limited solubility in the water-HEMA mixture and limited entropy of chain arrangements (crosslinked state). Microsyneresis in water-HEMA crosslinking system proceeds through the mechanism of the nucleation and growth which leads to a typical structure of mutually connected microscopic spheres providing a heterogeneous gel well visible in Fig. 5. These gels, when swollen to equilibrium volume in water, macroscopically appear white or opaque – far from the perfectly transparent appearance necessary for a contact lens. Interestingly enough, even standard hydrogel of composition used for contact lenses showed, already during polymerization, the formation of nanosized inhomogeneities, supposedly pores, of several typical dimensions between 1 and 10 nm.[28] Such inhomogeneities do not deteriorate the optical clarity of the final product but can enhance the transport of water, oxygen, and small ions. [38]

Microsyneresis provides an interesting and well-explored way nowadays leading to a formation of porous systems, predominantly with communicating pores having their size in the range of 10^0-10^1 μm. It is a system-specific thermodynamic phenomenon that can be predicted, is perfectly reproducible, and is inevitable within a certain compositional range.

As mentioned above, the HEMA monomer always contains a little amount of bis-methacrylic units (ethylene dimethacrylate, EDMA). During the polymerization, EDMA is gradually incorporated through its two vinyl groups into the polymer chains so that the branching and, at higher degrees of conversion, also crosslinking inevitably takes place.

During the development, various methods have been used to achieve the porosity of PHEMA:[40] besides the thermodynamic demixing, also introducing washable microparticles (porogen) into the gel matrix. In this way, interesting porous structures based on PHEMA have been prepared, including (nano)fibers.[41] Also composites of PHEMA, e.g. with bacterial cellulose, [42] or interpenetrating networks,[43] as well as materials with dual porosity[44] have been described.

5. PHEMA − APPLICATIONS OTHER THAN CONTACT LENSES

5.1. Medicinal applications

Since the seventies, within the group of younger Wichterle's colleagues, there existed a lively activity in the field of biological application of PHEMA materials other than ophthalmology.[45]

Due to its good compatibility with living tissue, PHEMA was predetermined for medicinal applications. During its decades-long history, this biocompatibility was proved beyond any doubt by its long-term use in this field. Some later studies then confirmed that not only the high-molar-mass polymer of HEMA but also its very short chains (oligomers) are well biocompatible.[46]

In fact, PHEMA has become a material of the first choice for biomedicinal applications, in particular for pilot experiments; subsequently, the material can be modified in many ways according to the needs of the particular application. Thanks to their transparency, homogeneous HEMA polymers found their first medicinal applications in ophthalmology. In addition to the already discussed soft hydrophilic contact lenses which aroused a global response, PHEMA has its history too as a material for intraocular lenses implanted into the eye during cataract surgery,[47] artificial vitreous body, [48] etc. Wichterle himself proposed many medicinal applications of PHEMA which were put in practice more or less successfully.

Of the other applications, known are implants for otorhinolaryngology,[49,50] plastic or general surgery,[51] gynecology,[52] urology,[53] and neurology,[54] as well as carriers for cell cultivation for dermal wounds healing, burns, or bedsores.[55,56] Polymers of HEMA are still used to prepare ointments/salves[57] and various gel preparations,[58] drug carriers,[59] tissue expanders,[60] synthetic emboli[61], or hemoperfusion detoxicating columns.[62] 3D microstructured carriers for cell cultivation, known as

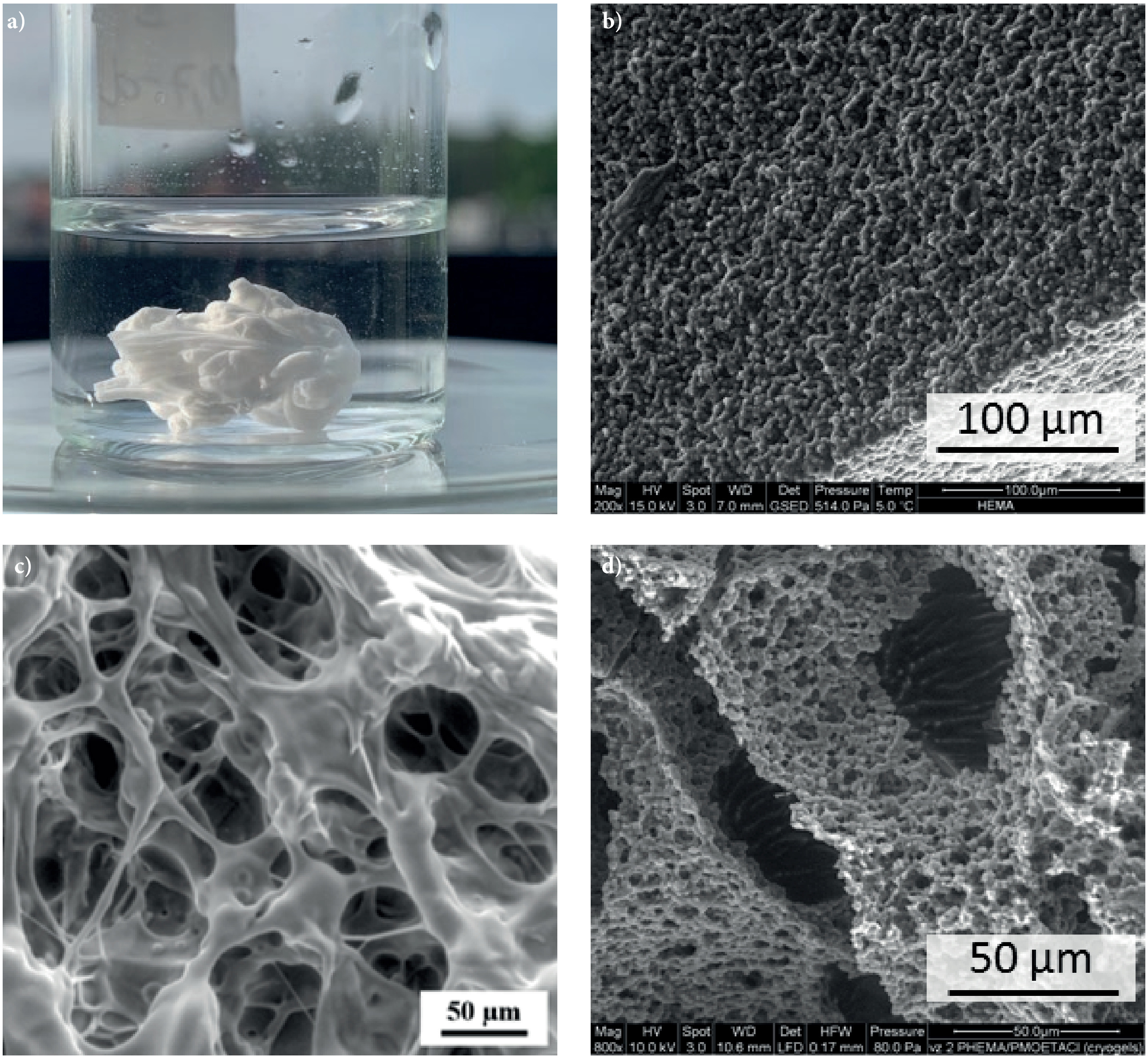


Figure 5. Porous hydrogels prepared from poly(2hydroxyethyl methacrylate) and visualized by the methods of scanning electron microscopy. (a) Macroscopic view; (b) PHEMA hydrogel showing after the microsyneresis a structure of connected spheres, (c) PHEMA hydrogel prepared from poly(HEMA-*stat*-MA) (fractionated NaCl was used as a porogen; after washing out the porogen, the gel was visualized by the AquaSEM method); (d) microscopic structure of a cryogel of HEMA showing the dual size of pores. Figs 5b and 5d were obtained by the so-called environmental SEM.[39]

scaffolds, have since recently been used. Thus, PHEMA has become a successful reference material also in the fields of cell therapy and tissue engineering.

Recently, with the development of additive manufacturing methods, HEMA finds its use as a photopolymerizing monomer in the resin compositions in stereolithographic 3D printing and 3D writing methods. It was used to constitute photopolymerizable ink for direct writing of 3D microarrays as scaffolds for neuronal cultures.[63]

5.2. Technical applications

To this category belong, e.g., (meth)acrylate coatings. PHEMA of technical grade is being used as a part of single-component dispersion coatings (together with

Figure 6. Historical cup restored using a preparation based on PHEMA. (a) an example of gluing glass on a base, (b) a detail of an attached substitutive bottom.

butyl acrylate or butyl methacrylate). As a comonomer, HEMA carries the functional reactive OH group into the polyol component of the two-component curable and highly resistant polyurethane coatings.[64] Another proven application, though not yet published, was the preparation of heterogeneous membranes with incorporated ion exchangers. The high adhesivity of PHEMA to other materials, as well as its transparency, enabled such technical applications as gluing of methacrylates or their layers. As an example, until now unpublished results of the tests (performed in 1982 and based on stress-strain curves) enabled one to assess the strength of the link formed by polymerization of 2-hydroxyethyl methacrylate in between two specimens, the latter being constituted by a common mineral glass, an organic glass, a polyamide, and steel of class 11. In all cases, very firm joints were obtained, resisting stress of about 2 MPa. The results, suitable especially for gluing glass, led to the testing of polymers based on PHEMA, to prepare permanent microscopic preparations, mechanically resistant layered glass or antifire layered glass, or to restore various historical glass objects (Fig. 6).

In an interesting application, water confined in certain hydrogels (semi-interpenetrating PHEMA/polyvinylpyrrolidone networks) was used to gently remove dirt from the surface of water-sensitive cultural artifacts.[38] Similarly, complex cleaning fluids confined in these hydrogels were used to remove aged varnishes.[65]

A highly diluted solution of PHEMA was tested by O. Wichterle as an "anti-spray" coating to prevent the creation of graffiti. Regrettably, to the best of our knowledge, this method has been neither patented nor published. Its advantage lies in that that the coating is cheap and can easily be removed by excess water.

6. CONCLUSIONS

It follows from the facts presented that the history of the origin, development, and applications of 2-hydroxyethyl methacrylate and its polymers is extremely interesting, varied, edifying, and sometimes even exciting. In this review, the development of the famous application of hydrogel based on poly(2-hydroxyethyl methacrylate) for contact lenses is presented. Inventors' effort was idea-driven rather than serendipitous: Otto Wichterle and his co-workers not only arrived at a technically useful product but also showed the general importance of hydrogels. The dispute over the validity of the corresponding patents became a subject of a thrilling lawsuit that ended with the victory of the inventors. The eventual success was possible thanks to inventors' endurance and ability to overcome the obstacles, both technical and political. The whole process from idea to final product took twenty years. When inspected in more detail, the present state of the art in the field suggests a possibility of further and deeper studies and even broad projects on the subject. In this way, some new properties, behavior, and applications of poly(2-hydroxyethyl methacrylate) hydrogels, so far unexplored, could be discovered.

ACKNOWLEDGMENT

The authors acknowledge the institutional support of the Institute of Macromolecular Chemistry of the Czech Academy of Sciences RVO: 61389013. The authors also thank Dr. Ivana Lorencová who kindly provided several photographs from the archives of the National Technical Museum in Prague.

REFERENCES

[1] O. Wichterle, D. Lím, *Nature* **1960**, *185*, 117-118. https://doi.org/10.1038/185117a0.

[2] C. Yu, M. Xu, F. Svec, J. M. J. Fréchet, *J. Polym. Sci. Part A, polym. Chem.* **2002**, *40*, 755-769. https://doi.org/10.1002/pola.10155.

[3] S. Kovačič, D. Stefanec, P. Krajnc, *Macromolecules* **2007**, *40*, 8056-8060. https://doi.org/10.1021/ma071380c.

[4] O. Kulygin, M. S. Silverstein, *Soft Matter* **2007**, *3*, 1525-1529. https://doi.org/10.1039/B711610A.

[5] M. F. Refojo in *Kirk-Othmer Encyclopedia of Chemical Technology (Contact Lenses)*. 3rd edition, vol. 16, Wiley, New York, **1979**, pp. 720-742.

[6] J. Lamb, A. G. Sabell in *Contact Lenses*, 5th Edition (Eds.: A. J. Phillips, L. Speedwell), Butterworth-Heineman, Elsevier, Oxford, UK, **2007**, pp. 1-20.

[7] T. J. Bowden, *Contact Lenses, the Story*, Bower House Publications, Gravesend, UK, **2009**, pp. 247-363.

[8] N. Efron in *Contact Lens Practice, Historical Perspective* (Ed.: N. Efron), Butterworth-Heinemann, Elsevier, Oxford, UK, **2010**, pp. 3-10.

[9] N. Efron in *Contact Lens Practice*, 3rd Edition (Ed.: N. Efron), Butterworth-Heineman, Elsevier, Oxford, UK, **2018**, pp. 3-9.

[10] R. B. Mandell in *Contact Lens Practice, Historical Development* (Ed.: R. B. Mandell), C. C. Thomas, Springfield, IL, **1988**, pp. 5-22.

[11] N. G. Gaylord (Polycon Lab Inc.), US 3808178, **1974**.

[12] J. Michálek, R. Hobzová, M. Přádný, M. Dušková in *Biomedical Appli cations of Hydrogels Handbook* (Eds.: R. Ottenbrite, K. Park, T. Okano), Springer, New York, **2010**, pp. 303-315.

[13] K. Tanaka, K. Takahashi, M. Kanada, S. Kanome, T. Nakajima (Menicon Co Ltd.), US 4139513, **1979**.

[14] O. Wichterle, D. Lím (O. Wichterle, D. Lím), US 3220960, **1965**.

[15] O. Wichterle, *Recollections*, Impreso, s.r.o., Žďár nad Sázavou, **1994**.

[16] J. Kopeček, *J. Polym. Sci., Part A, Polym. Chem.* **2009**, *47*, 5929-5946.

[17] J. Kopeček, J. Yang, *Polym. Int.* **2007**, *56*, 1078-1098, and the references No. 3 and 4, cited therein. https://doi.org/10.1002/pi.2253.

[18] O. Wichterle, Czechoslovak Patent Application **1953**-187-**53**.

[19] O. Wichterle, D. Lím, Czechoslovak Patent Application **1955**-1284.

[20] O. Wichterle, D. Lím, CS 91918, **1959**; GB 814009 (A), **1959**.

[21] O. Wichterle (Czechoslovak Academy of Sciences), CS 108895, **1963**; CA 710706 (A), **1963**.

[22] O. Wichterle (Czechoslovak Academy of Sciences), CA748158 (A), **1966**; CS 132141, **1969**.

[23] J. Michalek, R. Hobzova, M. Pradny, M. Duskova in Biomedical Applications of Hydrogels Handbook (Ed.: R. M. Ottenbrite, K. Park, T. Okano), Springer, New York, 2010, pp. 303-315.

[24] J. C. Woodhouse (Du Pont de Nemours), US 2129722, **1938**.

[25] https://en.wikipedia.org/wiki/(Hydroxyethyl)methacrylate, last accessed on 10/04/2018.

[26] M. Borovička, J. Vacík (M. Borovička, J. Vacík), CS 216150 (B1), **1982**.

[27] G. Chen (Wuxi Huaming Chemical Co. Ltd), CN 103304413A, **2013**.

[28] M. Dušková Smrčková, Z. Sadakbayeva, M. Steinhart, K. Dušek, *Macromol. Symp.* **2017**, *372*, 28-42. DOI 10.1002/masy.201700018.

[29] A. Hirao, H. Kato, K. Yamaguchi, S. Nakahama, *Macromolecules* **1986**, *19*, 1294-1299. DOI 10.1021/ma00159a002.

[30] T. Breiner, H. W. Schmidt, A. H. E. Müller, *E-Polymers* **2002,** art. No. *022.*

[31] H. Macková, Z. Plichta, H. Hlídková, O. Sedláček, R. Konefal, Z. Sadakbayeva, M. Dušková Smrčková, D. Horák, Š. Kubinová, *ACS Appl. Mater. Interfaces* **2017**, *9*, 10544-10553. DOI 10.1021/acsami.7b01513.

[32] C. Hou, S. Lin, F. Liu, J. Hu, G. Zhang, G. Liu, Y. Tu, H. Zouab, H. Luo, *New J. Chem.* **2014**, *38*, 2538-2547. DOI 10.1039/c3nj01398g.

[33] K. Dušek, B. Sedláček, *Coll. Czech. Chem. Commun.* **1969**, *34*, 136-157. https://doi.org/10.1135/cccc19690136.

[34] M. Dušková-Smrčková, K. Dušek, *ACS Macro Lett.* **2019**, *8*, 272–278. DOI: 10.1021/acsmacrolett.8b00987.

[35] K. Dušek, *J. Polym. Sci. B: Polym. Letters*, **1965**, *3*, 209-211. https://doi.org/10.1002/pol.1965.110030311.

[36] K. Dušek, B. Sedláček, *Collect. Czech. Chem. Commun.* **1971**, *36*, 1569-1577. DOI 10.1135/cccc19711569.

[37] K. Dušek, *Polym. Bull.* **2007**, *58*, 321-338. DOI 10.1007/s00289-006-0671-7.

[38] J. A. L. Domingues, N. Bonelli, R. Giorgi, E. Fratini, F. Gorel, P. Baglioni, *Langmuir* **2013**, *29*, *8*, 2716-2755. https://doi.org/10.1021/la3048664.

[39] J. Kaňka, E. Chylíková Krumbholcová, D. Chmelíková, J. Pánek, B. Podhorská, S. Krejčíková, T. Vampola, N. R. Banafshehvaragh, M. Vetrík, Z. Kaberova, M. Dušková Smrčková, *Chem. Listy* **2019**, *113*, 726-731.

[40] M. Přádný, M. Šlouf, L. Martinová, J. Michálek, *E-Polymers* **2010**, *043*, Art. No. 043.

[41] M. Přádný, L. Martinová, J. Michálek, T. Fenclová, E. Krumbholcová, *Cent. Eur. J. Chem.* **2007**, *5*, 779-792. DOI 10.2478/s11532-007-0021-0.

[42] R. Hobzova, J. Hrib, J. Sirc, E. Karpushkin, J. Michalek, O. Janouskova, P. Gatenholm, *J. Nanomater*. **2018**, Art. No. 5217095. DOI: 10.1155/2018/5217095.

[43] Z. Sadakbayeva, M. Dušková Smrčková, A. Šturcová, J. Pfleger, K. Dušek, *Eur. Polym. J*. **2018**, *101*, 304-313. DOI 10.1016/j.eurpolymj.2018.02.035.

[44] M. Přádný, M. Dušková Smrčková, K. Dušek, O. Janoušková, Z. Sadakbayeva, M. Šlouf, J. Michálek, *J. Polym. Res*. **2014**, *21*, Art. No. 579. DOI 10.1007/s10965-014-0579-0.

[45] L. Šprincl, J. Kopeček, D. Lím, *Calc. Tis. Res.* **1973**, *13*, 63-72 (1973). DOI 10.1007/BF02015397.

[46] E. Brynda, M. Štol, V. Chytrý, I. Cífková, *J. Biomed. Mater. Res*. **1985**, *19*, 1169-1179. DOI 10.1002/jbm.820190923.

[47] C. Anderson, D. D. Koch, G. Green, A. Patel, S. Van Noy in *Foldable Intraocular Lenses* (Eds.: R. G. Martin, J. P. Gills, D. R. Sanders), Slack Inc., Thorofare, NJ, pp. 161-177, **1993**.

[48] J. Vacík, J. Kopeček, D. Lím, O. Wichterle (Czechoslovak Academy of Sciences), CS 146642, **1972**; CA 913259 (A), **1972**.

[49] M. Zignani, S. F. Bernatchez, T. Le Minh, C. Tabatabay, J. M. Anderson, R. Gurny, *J. Biomed. Mater. Res.* **1998**, *39*, 277-285.

[50] Z. Voldřich, Z. Tománek, J. Vacík, J. Kopeček, *J. Biomed. Mater. Res.* **1975**, *9*, 675-685.

[51] V. Olivier, N. Faucheux, P. Hardouin, *Drug Discovery Today* **2004**, *9*, 803-811. DOI 10.1016/S1359-6446(04)03222-2.

[52] T. E. J. Ind, J. C. Shelton, J. H. Shepherd, *Br. J. Obstet. Gynaecol.* **2001**, *108*, 1013-1016. DOI 10.1016/S0306-5456(01)00251-0.

[53] L. Šefc, M. Přádný, J. Vacík, J. Michálek, C. Povýšil, I. Vítková, M. Halaška, V. Šimon, *Biomaterials* **2002**, *23*, 3711-3715. DOI 10.1016/S0142-9612(02)00104-7.

[54] P. Lesný, J. De Croos, M. Přádný, J. Vacík, J. Michálek, S. Woerly, E. Syková, *J. Chem. Neuroanat*. **2002**, *23*, 243-247. DOI 10.1016/S0891-0618(02)00011-X.

[55] K. Smetana, B. Dvořánková, M. Jelínková, J. Michálek, J. Vacík, *J. Mater. Sci. Mater. Med*. **1997**, *8*, 587-590.

[56] B. Dvořánková, Z. Holíková, J. Vacík, R. Königová, Z. Kapounková, J. Michálek, M. Přádný, K. Smetana, *Int. J. Dermatol*. **2003**, *42*, 219-223. DOI 10.1046/j.1365-4362.2003.01792.x.

[57] J. S. Ahn, H. K. Choi, M. K. Chun, J. M. Ryu, J. H. Jung, Y. U. Kim, C. S. Cho, *Biomaterials* **2002**, *23*, 1411-416. DOI 10.1016/S0142-9612(01)00261-7.

[58] J. Labský, J. Vacík, P. Hošek (Institute of Macromolecular Chemistry, Czechoslovak Academy of Sciences), CZ 293419, 2004; US 6610284 (B1), 2003.

[59] J. Sirc, J. Hrib, M. Vetrik, R. Hobzova, A. Zak, B. Stankova, O. Slanar, R. Hromadka, V. Sandrikova, J. Michalek, *Physiol. Res*. **2015**, *64*, S51-S60. DOI 10.33549/physiolres.933142.

[60] J. Hrib, J. Sirc, P. Lesny, R. Hobzova, M. Duskova Smrc kova, J. Michalek, R. Smucler, *J. Mater. Sci. Mater. Med*. **2017**, *28*, Art. No. 12.

[61] J. H. Peregrin, R. Janoušek, D. Kautznerová, M. Oliverius, E. Sticová, M. Přádný, J. Michálek, *Physiol. Res.* **2015**, *64*, 841-848. https://doi.org/10.33549/physiolres.932992.

[62] J. Kulíšek, J. Kálal, M. Tlusťáková, J. Horák, J. Vavrouch, CS 233015, 1984.

[63] J. N. Hanson Shepherd, S. T. Parker, R. F. Shepherd, J. A. Lewis, *Adv. Funct. Mater.*, DOI: 10.1002/adfm.201001746.

[64] J. Huybrechts, A. Vaes, K. Dusek, M. Duskova-Smrckova, R. J. Barsotti, *Surf. Coat. Int., Part B, Coatings Trans.* **2006**, *89*, 275-283. DOI 10.1007/BF02765579.

[65] M. Baglioni, J. A. L. Domingues, E. Carretti, E. Fratini, D. Chelazzi, R. Giorgi, P. Baglioni, *ACS Appl. Mater. Interfaces* **2018**, *10*, *22*, 19162-19172. https://doi.org/10.1021/acsami.8b01841.

APPENDIX

Table I. Categorization of contact lenses from the material point of view. How it changed in time

Original division	hard			
	soft		hydrophobic hydrophilic	
Recent division	hard	impermeable	glass, PMMA	
		gas permeable	rigid gas permeable (RGP)	
	soft	hydrophobic	silicone elastomers	
		hydrophilic	standard hydrogels	PHEMA
			high swelling hydrogels	
			hybrid hydrogels	silicone hydrogels
Present day division	rigid gas permeable (RGP) hydrogels silicone hydrogels			

Table II. Important dates in global contactology (from the viewpoint of polymer materials and manufacturing methods)

1933	Rohm and Hass Co. introduced transparent polymethyl methacrylate (PMMA) into the market.
1936	William Feinbloom described a scleral lens composed of a central clear part (glass) and an opaque edge (PMMA). Soon after that, rigid lenses have been produced by turning solely from PMMA.
1948	By mistake in turning, Kevin Tuohy prepared a very small size lens of PMMA and found that it was better tolerated than that of the original size. Afterward, he patented hard corneal lenses of PMMA.
1953	D. Lím successfully prepared the first hydrogel following the idea of Otto Wichterle; application of the first O. Wichterle's patent.
1956	The first hydrogel contact lens was prepared in Wichterle's Prague laboratory.
1959	Tests on volunteers showed good correction of *visus* and excellent tolerance of hydrogel contact lenses.
1960	Wichterle and Lím published an article in *Nature*, entitled "Hydrophilic Gels for Biomedical Use" where they described PHEMA gels.
1961	Priority of spin casting method of hydrogel contact lens fabrication (Wichterle)
1963	Priority of lathe cutting method of lens fabrication from xerogel blocks (Wichterle)
1965	Hydrophobic soft contact lenses made of silicone elastomers
1972	Hydrophilic (hydrogel) soft contact lenses were introduced to the global market.
1974	RGP – rigid gas permeable lenses
1988	Lenses with regular replacement (cast molding technology began to prevail)
1994	Disposable lenses (regular replacement after one day)
1998	Silicone hydrogels, 1st generation
2004	Silicone hydrogels, 2nd generation
2006	Silicone hydrogels, 3rd generation (till present day)
2014	New highly swollen hydrogel contact lenses (Biotrue ONEday) were introduced on the global market. Their material (Hypergel™) contains in equilibrium 78% of water.

Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

Citation: Franza A., Pratesi G. (2022) Just a Grand Duke who Loves Chemistry. Peter Leopold of Habsburg-Lorraine (1747–1792) and his Chemical Cabinet at the Imperial and Royal Museum of Physics and Natural History. *Substantia* 6(2): 93-106. doi: 10.36253/Substantia-1600

Received: Mar 12, 2022

Revised: Jun 26, 2022

Just Accepted Online: Jun 27, 2022

Published: September 1, 2022

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

ORCID
AF: 0000-0003-3146-6957
GP: 0000-0001-6329-901X

Just a Grand Duke who Loves Chemistry. Peter Leopold of Habsburg-Lorraine (1747–1792) and his Chemical Cabinet at the Imperial and Royal Museum of Physics and Natural History

Annarita Franza, Giovanni Pratesi*

Department of Earth Sciences, University of Firenze, Via G. La Pira 4, 50121 Firenze,
E-mail: annarita.franza@unifi.it; giovanni.pratesi@unifi.it
*Corresponding author.

Abstract. This article dealt with the history of the chemical cabinet established by the Grand Duke of Tuscany, Peter Leopold of Habsburg-Lorraine (1747–1792), at the Imperial and Royal Museum of Physics and Natural History in Firenze during his regency. To achieve this goal, it investigated untapped archival sources (e.g., administrative and commercial documents, minutes, correspondences, inventories) concerning the museum management from its foundation in 1775 to the departure of the Grand Duke for Vienna to be crowned as Holy Roman Emperor Leopold II in 1790. The article analyzed the chemical cabinet's manuscript catalog, whose entire transcription is presented in the Supplementary Information Files. The work then examined the connections between the activities performed at the chemical laboratory and Peter Leopold's interest in experimental chemistry. Concerning this research question, the scientific relationship he held with the naturalist Giovanni Valentino Mattia Fabbroni (1752–1822) – Vice-director and then Director of the Imperial and Royal Museum of Physics and Natural History – who helped the Grand Duke navigate all aspects of his interests in chemistry and natural sciences, was also discussed.

Keywords: Peter Leopold, history of chemistry, Giovanni Fabbroni, catalog, museum.

INTRODUCTION

Writing about a notable historical figure such as Peter Leopold of Habsburg-Lorraine (1747–1792) is not an easy task. So, it is not by chance that Soll stated that "there is no single way to characterize Peter Leopold. His approach to intellectual life, philosophy, and governing were pragmatic, eclectic, and a mix of his various influences".[1] In fact, it is difficult to account for the life[2] of a man who was the ninth son of Francis I (1708–1755) and Maria Theresa of Austria (1717–1780), brother of Joseph II (1741–1790) and Marie Antoinette (1755–1763),[3] Grand Duke of Tuscany from 1755 to 1790, and the next-to-last Holy Roman Emperor (Leopold II) from 1790 to 1792, year in which he suddenly died from pleurisy at the age of 44 years.[4] But that is not all. Peter Leopold was a Jansenist and an enlightened despot,[5] whose concrete projects – as stated by Maran, Castellini, and Bisman–[6] reformed the administrative, managerial, organizational, judicial,[7] and economic aspects of social and cultural life in the Grand Duchy of Tuscany. In this regard, Sarti found that these amendments "transformed Tuscany into a model state",[8] while Maran et al. underlined how the analysis of the municipalities' reform contributed to a better understanding of today's decentralization by governments in the context of the new public management.[9] However, a complete survey of Peter Leopold's reforms of Tuscany is out of scope here, and the reader is referred to the literature for further details.[10]

The principal aim of this study is to detail and analyze the history of the chemical cabinet Peter Leopold established at the Imperial and Royal Museum of Physics and Natural History through the investigation of the scientific relationship the Grand Duke held with the naturalist and Museum's Vice Director Giovanni Valentino Mattia Fabbroni (1752–1822)[11] together with the examination of the laboratory's catalog.[12] The latter unpublished source is preserved at the Museo Galileo's library, whose archival fund on the history of the Imperial and Royal Museum of Physics and Natural History[13] holds various documents that were investigated to achieve the goals of the present paper.

The existing literature on the Imperial and Royal Museum of Physics and Natural History is extensive and focuses mainly on its collections and history.[14] The Museum, established by Peter Leopold's *motu proprio* on 22 February 1775, was not only a place to gather and preserve the naturalistic and scientific collections inherited by the Medici family[15] and the Accademia del Cimento[16] but also acted as a research center to promote useful knowledge in the service of the public good.[17] In this regard, Contardi underlined how the museum collections were open to all and organized to encourage a visitor's self-instruction through the display of explanatory labels.[18] However, the collections did not comprise only the specimens and instruments belonging to the Medici family and the Accademia del Cimento. They were constantly enriched by new acquisitions[19] promoted by Peter Leopold, together with the instruments, preparations, and objects[20] realized at the various museum workshops.[21] Among these, there was the chemistry cabinet that was established to represent, according to both Felice Fontana (1730–1805) and Giovanni Fabbroni, one of the most advanced research centers at an international level. It encompassed, in addition to the standard equipment like flasks and bell-jars, diverse pneumatic pumps[22] together with a workbench –realized by the woodcutter Francesco Schmidt (dates uncertain)– [23] belonging to Peter Leopold. It was openable with a slate working surface for experiments. One of the three cavities in the workbench is linked to a bellow operated by pedals useful for calcination and combustion operations. Various shelves allowed to store glassware, tiny bottles, and chemical compounds (Fig. 1).

Figure 1. Leopold's chemistry cabinet. Museo Galileo, Room X.

In this regard, Scorrano et al. analyzed 38 samples using X-ray fluorescence, X-ray diffraction, gas chromatography-mass spectrometry, IR spectroscopy, thermal analysis, and chemical tests. According to the authors, most of the compounds were employed in textile manufacturing, while the remaining represented both chemicals helpful in improving wine production and substances of apothecary interest.[24] Many historians have connected the investigation on the workbench to the work of Huber Franz Hoefer (1728–1795) at the court apothecary, as well as to the research the Grand Duke patronized on water[25] and hot springs[26]. While recent studies have linked Peter Leopold's interest in chemistry to mineralogy, mining science, and mineral collecting.[27] However, so far, there has been little discussion about the involvement of Peter Leopold in the experiments being conducted at the chemistry cabinet and the organization of the chemical laboratory. In the following pages, the analysis of untapped archival documentation will generate fresh insights into these subjects.

MATERIALS AND METHODS

As mentioned in the Introduction Section, this study investigated the historical documentation that described both Peter Leopold's interest in chemistry and the chemical cabinet he established at the Imperial and Royal Museum of Physics and Natural History. The documentation covers a period from the museum's foundation in 1775 until the end of Peter Leopold's regency as Grand Duke of Tuscany in 1790. Exceptions were documents about the chemical cabinet covering a period up to 1807. The materials were analyzed according to the standard archival research methods illustrated, for example, in Ventresca and Mohr.[28] The benefit of this approach is providing access to data that would not otherwise be known. Furthermore, the presentation of the archival documents according to a schematic organization facilitates their use as a resource for scholars interested in the history of chemistry in 18th-century Italy.

Bonding in Chemistry: Peter Leopold and Giovanni Fabbroni

One of the first research questions that come to mind is: did Peter Leopold attend the Imperial and Royal Museum of Physics and Natural History? Or the museum's establishment was, as stated by Solomon,[29] just a result of the European Enlightenment ideas? If, on the one hand, it is beyond dispute that the museum foundation was a consequence of Peter Leopold's reformism, i.e., the promotion of science as an instrument of public utility,[30,] on the other hand, it represented a place that the Grand Duke used to attend in person as shown by a *billet* his Intimate Secretary, Giovanni Tommaso Mannucci (1750–1814), sent to Fabbroni on 3 September 1789 forewarning his visit for the following day.[31]

At the Imperial and Royal Museum of Physics and Natural History, Peter Leopold established a close scientific relationship with Giovanni Fabbroni. The latter also assisted him in handling the contacts with various foreign chemists and apothecaries, such as Antoine Baumé (1728–1804).[32] Through the mediation of Francesco Favi (1749–1823)[33] –who was the Secretary of the Tuscan Legation in Paris deputed to purchase and ship diverse kinds of scientific products to Firenze via Marseille, Livorno, and Pisa–[34] Fabbroni got in contact with Baumé and received a copy of the catalog of his preparations in July 1787. The *Genres du Catalogue de M. de Baumé* encompassed 40 items, among which there were cobalt crystals useful for both nitrous and vitriolic acids, flowers of benzoin (benzoic acid), fusible salts of urine (i.e., the complex of salts present in human urine, which may indicate ammonium sodium phosphate or sodium ammonium chloride) and diverse essential oils (e.g., chamomile and cardamon).[35] Since then, Fabbroni bought various boxes of Baumé's preparations until 1788.[36] One of the last products Fabbroni purchased from Baumé at Peter Leopold's request was a supply of *pâte de guimauve*, i.e., a demulcent lozenge prepared from the root of *Althea officinalis*.[37] On Tuesday 19 May 1788, Mannucci advised that "His Royal Highness ordered Fabbroni to write to Monsieur Baumè to send him, as soon as possible, 12 pounds of *pâte de guimauve* within a small cup, so that it would not be broken". The order had to be of particular urgency since Mannucci sent the same note twice on Friday 22 May 1788.[38] At the end of June, Fabbroni noticed that the box was just arrived in Firenze, hoping that it would be delivered at the Royal Chamber soon.[39] Unlike the previous shipments, the *pâte de guimauve* was transported from Paris to Firenze via Genova, and its delivery was in the care of the courier Francesco Maria Vignolo (dates uncertain).[40]

On the same note, Fabbroni gave positive news about sending 17 pounds and 2 ounces of Venus crystals (copper acetate) that had been bought from Bertrand Pelletier's (1761–1767) laboratory.[41] The products sold by Pelletier had already aroused the Grand Duke's interest in the past. As an example, on 6 April 1787, during one of his stays in Pisa, Peter Leopold charged his Intimate Secretary Ranieri Fulger (dates uncertain) writing Fabbroni to inquire about the essences and essential oils sold by the Parisian apothecary.[42]

Peter Leopold was then very interested in the work of another French chemist, Jean Antoine Claude Chaptal (1756–1832).[43] In an undated document, Mannucci referred to Fabbroni that the Grand Duke had heard the news about a "Monsieur Chaptal who has a big store in Montpellier selling all kinds of chemicals, acids, and drugs of excellent quality at a meager price". Mannucci then said that "His Royal Highness will appreciate Fabbroni writing to some of his correspondents to know whether this is true or not. In case of a positive response, His Royal Highness will be pleased to receive a note on all the products he sells with their respective prices".[44] Fabbroni wrote to Chaptal[45] and the latter sent back a descriptive letter[46] on both the compounds he produced and the procedures he followed for realizing the crystallization of vitriol oil. In this regard, it has to be said that Chaptal sent Fabbroni letters like this more than once. For example, he illustrated the products he sold in Montpellier and his research methods in a missive dating 24 December 1786. This document shows on the top right of its first page a Fabbroni's brief remark in which he noted that the response was sent on 15 January 1787. He also noticed to have rewritten to the French chemist on 22 February 1787 because he was waiting for a box of chemicals that was not delivered yet. Later, Chaptal would have informed Fabbroni about his chemical products and his upcoming scientific publications. This is the case of a letter, partially gone missing, in which he noticed the forthcoming release of a treatise in two volumes on modern chemistry (i.e., *Elements de Chemie*, 1790).[47]

On 10 November 1786, Fabbroni sent a request for chemicals to Chaptal following a detailed list Peter Leopold had sent to him some days earlier.[48] The content of the list is reported in Table 1. It is interesting to note that, on 27 November 1786, the Grand Duke required that the products, once in Firenze, be delivered to him in Pisa. They had to be shipped via river using Monti's boat that sailed every Friday with the groceries for the Royal Pantry.[49]

Table 1. Peter Leopold's order of chemicals purchased from Jean Antoine Claude Chaptal. 4 November 1786.

Substances	Weight
Vitriol oil	10 pounds
Alkali Volatile Fluor	6 pounds
Alkali Volatile Concrete	6 pounds
Alkali Volatile Caustic	6 pounds
Salt of Tartar	8 pounds
Volatile Salt of Hart's Horn	2 pounds
Cinchona Essential Oil	1 pounds
Butyric Antimony	1 pounds
Volatile Salt of Hart's Horn	3 pounds
Oil of Tartar per Deliquium	4 pounds
Diaphoretic Antimony	1 pounds
Flowers of Benzoin	1 pounds
Emetic Tartar	1 pounds
Infernal Stone	3 ounces
Marine Ether	1 pounds
Acetic Ether	1 pounds
Sugar-cane Acid	1 pounds
Regulus of Cobalt	10 ounces

A second list, containing the request for other three chemicals (e.g., six ounces of phosphor, six ounces of pyrophorus,[50] and one pounds of tartar acid) to buy from Pelletier's laboratory, was attached in a letter Peter Leopold sent to Fabbroni from Pisa on 17 January 1787.[51]

The delivery of the products coming from Chaptal's store in Montpellier was late, so the company Durand Martin et C.ie wrote to Fabbroni to apologize for any inconvenience it might have caused on 3 March 1787.[52] When the shipment finally arrived in Pisa on 26 March, Peter Leopold personally opened the three boxes and checked their content. After separating what he would have kept for himself from what would have been sent to the chemistry cabinet, the Grand Duke noted that three chemicals were missing (e.g., the essential oil of cinchona, the flowers of benzoin, and the marine ether), so he asked Fabbroni to contact Chaptal to send them as soon as possible. He was also informed that an unusual quantity of vitriol oil had been delivered to the Imperial and Royal Museum of Physics and Natural History and asked to set aside a bottle he would have used for his experiments when back in Firenze. In this regard, it has to be noted that Peter Leopold carried out chemical experiences outside the Royal Cabinet.[53] For instance, on 18 December 1786, he gave instructions to the chancellor Gaetano Stefani (dates uncertain) to contact Fabbroni asking for a supply of lute[54] to be sent to Pisa. The compound, mixed with boiled oil and clay, had to be like the one previously prepared by Fabbroni for the experiments he conducted in the chemistry cabinet. Peter Leopold's request was quite urgent. Fabbroni had to prepare and deliver the lute in a day to the grand-ducal Intimate Secretary, so it could be placed in a suitcase that was ready to be shipped to Pisa.[55]

Coming back to the order of chemicals delivered by the French apothecary Chaptal in March 1787, Peter Leopold gave away the butyric antimony to the chemistry cabinet. He advised Fabbroni that the products' jars were extremely fragile and had to be handled with highly caring.[56] One of the last orders Fabbroni commissioned to Chaptal dated from mid-January to mid-March 1788.[57]

As mentioned before, Fabbroni was not only an agent for purchasing chemicals but also a mentor who guided the Grand Duke through the path of chemistry. For example, while he was experiencing cinchona, Peter Leopold asked Fabbroni if its extract could be reverted to gel in a brass pot, in a copper and silver evaporator, or if it was necessary to store the compound in an earthenware evaporator.[58] Fabbroni was also Peter Leopold's mentor for mineral chemistry and its application to mining science. In this regard, after visiting the Habsburg iron mines in 1779, Peter Leopold realized that the Tuscan iron deposits were not adequately exploited. Therefore, he asked Thaddeus Rauscher (dates uncertain), a mining expert in Carinthia, to come to the Grand Dutchy of Tuscany to evaluate the ore processing.

Fabbroni discussed the results of Rauscher's investigations in two reports in 1780. In the first document, the events concerning these inspections were summarized. Fabbroni stated that two experts from Carinthia arrived in September to check the quality of the iron ores and any abuses committed by the local mining administrations. However, one of the two men left the country promising to return after a few days, while the other remained in Firenze without continuing the work. New mining experts arrived from Carinthia, and after analyzing some ores extracted in Maremma, they came to different conclusions about the iron processing systems in use. The first one stated that the defects in ironworking originated from the small amount of coal used in processing the raw materials. The second expert thought that they were caused by an insufficient quantity of iron ores being processed.[59] So ends the first document. In the following *Report on the iron mines in the Grand Duchy of Tuscany and the sampling made by Taddeo Glauscher of Carinthia in 1780*,[60] Fabbroni disagreed with the opinions of the foreign mining experts, saying that "they were not aware of the progress in iron processing made outside their homeland". As an example, he stated that the three furnaces employed in the smelting were necessary because of the samples' size, while the iron produced in Livorno, whose excessive malleability could be easily corrected, could also be used to manufacture nails and plowshares. However, at the end of his second report, Fabbroni agreed with Rauscher about the necessity to reorganize the wood supply for the furnaces in Val di Cecina. Moreover, the excellent quality of the coal deposits in those areas was worthwhile for further analysis.

Peter Leopold was no stranger to this dispute. He was pleased with the results shown in Fabbroni's reports to the point that he supported the research of new ore deposits in Val di Cecina in the years to come. Starting from January 1789, Fabbroni investigated the quality of the soil near the coal mines, their extent, and the possibility of finding other deposits nearby. The Grand Duke also ordered that an essay on coal mining and its processing methods be written to improve and promote this activity in the Grand Duchy of Tuscany. One year later, Fabbroni informed Alessandro Pontani (dates uncertain), the Secretary of the State Council, to have finished the book the Grand Duke requested. The volume was called *Dell'antracite o carbon di cava detto volgarmente carbon fossile* (On anthracite or quarry coal, commonly called hard coal) and was printed in 250 copies by Gaetano Cambiagi (1725–1795), who was a typographer and the owner of the grand-ducal printing house.[61] Besides the investigations he carried out in Val di Cecina, Fabbroni was also deputed to the characterization of presumed coal samples. These analyses were performed in the chemistry cabinet. For example, on 26 December 1788, he examined 20 coal samples brought to the museum by Francesco Henrion (dates uncertain), an archivist at the Archives of the Tithes of the Grand Duchy of Tuscany.[62] On Peter Leopold's request, Fabbroni also examined ores coming from Montecerboli to ascertain the presence of borax[63] and copper specimens to verify the possibility of opening a new mine near Arcidosso, a project carried out some years later under the grand-ducal regency of Ferdinand III (1769–1824).[64]

In the following sub-section, the organization of the chemistry cabinet is described through the examination of its catalog –comprising instruments, books, and compounds, some of which belonged to and were made by Peter Leopold. Information is also given on the destiny of the cabinet after the Grand Duke left Firenze in the summer of 1790 to be crowned as Holy Roman Emperor Leopold II.

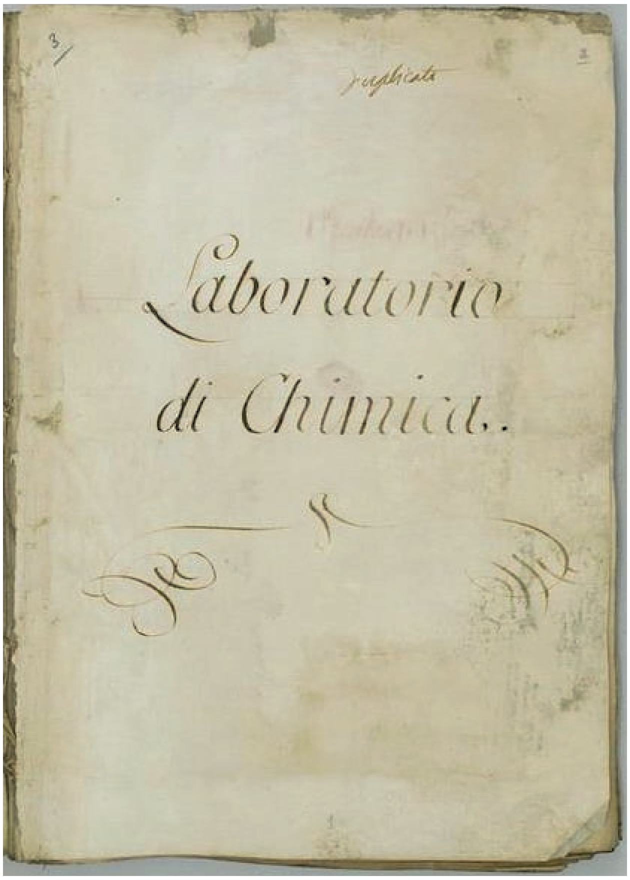
A Grand Duke at Work: Peter Leopold's Chemistry Cabinet

The organization of the chemistry cabinet is illustrated in a 100-pages manuscript entitled *Laboratorio di Chimica* (Chemistry Laboratory), which is preserved at the Museo Galileo's Library.[65]

The volume covers a period from 1780 to 1807. It is divided into four sections, the first of which is represented by the *Indice delle droghe e preparazioni chimiche lasciate al Reale Museo da S.M.I. e collocate nella prima nuova stanza del Laboratorio chimico* (Index of the drugs and chemical preparations left to the Royal Museum by H.R.H. and placed in the first new room of the Chemical Laboratory). On the first page was reported, in red ink, the location (Room II) where the preparations were kept, and each compound was preceded by the symbols "·" or "x", which probably indicated its presence or lack ascertained during an inventory. The chemicals were then grouped according to the shelves where they were stored, as shown in Table 2. The Results and Discussions section will analyze this part of the chemistry cabinet's catalog.

The following catalog's section concerned the chemicals obtained by processing the "three Kingdoms of Natures" (i.e., animals, plants, and minerals), which were housed in the second of the three rooms designed to accommodate the chemical cabinet within the Imperial and Royal Museum of Physics and Natural History.[66] In this regard, it has to be noted that most of these compounds were in poor conservative conditions, since they were evaporated or altered, and needed to be restored. The preparations' list has been transcribed in Supplementary Information Files I–IV, while some information of potential interest resulting from the transcription activity is reported below.

In a closet on the left of the room were substances of animal and/or human origin (e.g., cow's milk serum,



Laboratorio di Chimica.

Figure 3. Frontispiece chemistry cabinet's catalog. Biblioteca del Museo Galileo, Firenze.

Table 2. Numbers of shelves and preparations as listed in *Index of the drugs and chemical preparations left to the Royal Museum by H.R.H. and placed in the first new room of the Chemical Laboratory.*

Shelf	Drugs and Preparations Nos.
I	81
II	45
III	79
IV	63
VIII	92
IX	89
X	36
V	40
VI	145
VII	85
IX	54
XII	51
XIII	75
Above the shelves	15

gelatinous part of dried human blood, urine salts) for a total of 129 preparations. Some of these compounds' descriptions are followed by the letters "MB" in red ink, while the red-ink string "O-KI-AO-" is placed before the "donkey glue as prepared by Chinese people" (*Colla d'Asino come preparate dai Cinesi*). Subsequently, there were the plant chemicals comprising 295 preparations. Among them, seven samples were preserved without their containers.[67] It is then interesting to note that diverse samples of honey were listed as preparations of vegetable origin, and one of these varieties (e.g., honey without phlegm, *Miele sflemmato* in the original text) was marked with the red-ink letters "MB". It is also noteworthy to highlight the presence of two kinds of milk sugars (e.g., impure sugar milk, *Zucchero di Latte impuro*; white sugar milk, *Zucchero di Latte bianco*) as well as the numbers "II" and "III" to indicate two different samples of rectified ether. Numbers "2", "2.3", and "3" were used to show respectively a sample of turpentine essential oil (*Olio essenziale di Terebentina*), and two samples of turpentine oil (*Olio di Terebentina*).

The only compound that was cataloged as a poison was a sample of Ticunas, i.e., an American poisonous substance whose effects were studied by Felice Fontana in an essay published in 1780.[68] On the right side of the room, a closed kept the compounds that were realized by processing minerals for a total amount of 436 samples. Two preparations (e.g., Earth-based Sea salt, *Sale marino a base terrosa*; Hombergy's sedative salt, *Sale sedativo d'Hombergio*) were inventoried without containers (vase missing, *manca il vaso*). The red-ink numbers "I", "II", "III", and "IV" indicated four samples of blue enamel (*Azzurro di smalto*), while the string "100.6" and "8.00" was reported on a cobalt sample (*Cobalto reputato fattizio con 100.6. e 8.00. di Vienna*). The number "2" showed that a sample of Berlin-blue was prepared using Sheele's method (*Azzurro di Berlino privato della parte colorita col metodo 2: di Sheele*). The letters "A.B." were then present in the description of a crystallized fusible alloy (*Lega fusibile dell'A. B. cristallizzata*). The cabinet also preserved 104 chemistry and natural sciences books constituting the third catalog's section.[69] Many of these volumes were written in German and French, while only two editions in Italian were present.[70] The inventory did not provide information about the books' publishing house or their year of publication with the exception of *Taschenbuch für Scheidekünstler und Apotheker auf das Jahr* (1789) and Göttling's *Almanach* (1780). The books' titles were usually shortened as well as the names of their authors. The number of tomes in a single work followed the title. At the end of the section, a brief note remarked that all the compounds and preparations are usually used in experiments and other investigations. For this reason, they were not included in the general catalog. The transcription[71] of the entire inventory is provided in the Supplementary Information Files.

The fourth and last catalog section described 56 scientific instruments (e.g., balances, mortars, boilers). The transcription of this section is also presented in the Supplementary Information Files. It is interesting to note that the title of this section referred to the objects stored in all three cabinet rooms (*Utensili e arnesi che si conservano nelle tre stanze del Laboratorio*). Nevertheless, ten instruments kept in a fourth room, i.e., the cabinet's obscure hallway (*androne oscuro*), were also inventoried. It is then worth of mention that the descriptions presented in this section are more detailed than those given in its previous catalog parts, providing accurate data about the instruments' physical and external characteristics, the materials they were made from, and whether accessories were present or not. By analyzing this descriptive model, it was discovered that two anvils showed the emblem of the Medici family.[72]

After Peter Leopold returned to Vienna to be crowned Holy Roman Emperor Leopold II, he donated the workbench and all the objects preserved in the chemistry cabinet to the Imperial and Royal Museum of Physics and Natural History. The donation happened on 27 July 1790 when all the preparations belonging to the Grand Duke were transferred to the purpose-built museum room together with "various tools, housewares, earthenware, and crystal vases". Fabbroni's economic evaluation estimated the cabinet's value at 7217 Tuscan lire. In detail, the chemical preparations and the instruments were evaluated at 6235 Tuscan lire; the housewares, the furniture, and the tools at 952 Tuscan lire; the books at 30 Tuscan lire.[73]

RESULTS AND DISCUSSIONS

This study aimed to assess the history of the chemical cabinet at the Florentine Imperial and Royal Museum of Physics and Natural History during the Grand Duke Peter Leopold's regency. To achieve this goal, the catalog of the chemical cabinet was analyzed, and the entire manuscript text was transcribed in the Supplementary Information Files. The investigation showed that 1534 items were kept in the cabinet between 1780 and 1807. Among them, there were 1374 chemicals. What is interesting about the data is that 950 compounds belonged to Peter Leopold and therefore, they were inventoried and described altogether in the first part of the catalog. A closer inspection of their list transcribed in the Supplementary Information File I revealed as some of these preparations might have been prepared by the Grand Duke himself. For instance, a sample of *Lac Martin* – i.e., a mastic-based varnish used in the restoration of paintings – was described as "made by His Royal Highness [on] 29 December 1780".[74] One of the most intriguing aspects, besides the sample evidence of Peter Leopold's genuine interest in chemistry, is that it was still preserved in the early 20th century when Ugo Schiff (1834–1915) noted its presence among the compounds kept in the chemical cabinet. Schiff also outlined that Jakob Philipp Hackert (1737–1807), a painter working for Ferdinand I of the Two Sicilies (1751–1825) in Napoli, had probably illustrated the method for preparing the *Lac Martin* to Peter Leopold during the meeting he had with the Grand Duke at the end of 1778.[75] In this regard, it has to be noted that the sample made by Peter Leopold was stored on Shelf no. VIII, which was entirely devoted to the preservation of dyes and varnishes for painting, coating wood, gilded brass, and other minerals. Among the 92 preparations, there were two samples of *Lac Martin*: the one prepared by Peter Leopold and a second one briefly described as "Bechi's *Lac Martin*".[76] Unfortunately, the catalog provided neither further information on Peter Leopold's preparation nor on the Bechi's one.

Nevertheless, the latter could be possibly identified with Antonio Bechi (dates uncertain), a member of the Florentine Confraternity of the Misericordia and an impresario who established a theater in Via di Porta Rossa in 1760.[77] In the current state of research; it is neither possible to affirm nor deny that the Bechi who prepared the second Lac Martin sample was Antonio Bechi. And if this was the case, neither had he ever met Peter Leopold to explain the varnish preparation. However, this is a hypothesis valuable to be explored in further studies.

Another striking observation emerging from comparing the chemicals' inventory belonging to Peter Leopold is that various preparations were comprised in the orders the Grand Duke received by the diverse French apothecaries in the 1780s. As expected, not all the compounds bought through the years have been cataloged (e.g., the *pâte guimauve*) because they were used to perform experiments or make other preparations, such as the radical vinegar of Venus, which was based on Venus crystals. However, Table 1 showed that most of the chemicals ordered to Chaptal on 4 November 1786 (excepting the essential oil of cinchona) were cataloged in pure form (i.e., flowers of Benzoin, emetic tartar, infernal stone, and sugar-cane acid) or in combination with other compounds. This is the case of the marine ether, which was used to make four different preparations (i.e., manganese marine ether with vitriolic oil, marine salt, and wine spirit; marine ether with Libavio's liqueur; nitrous and vitriolic marine ether).

It is then worth of mention that Peter Leopold's interest in mineral chemistry outlined in the previous Section was confirmed by the catalog's analysis of his preparations preserved in the chemical cabinet at the Imperial and Royal Museum of Physics and Natural History. For example, diverse compounds were utilized in mineral processing, such as vitriolic oil, which was used in the distillation of manganese and the extraction of lead in association with cobalt and zinc. In this regard, it is interesting to note the presence of many preparations constituted by minerals such as ruby-Sulphur (realgar)[78] and various copper-, antimony-, and manganese-based compounds. About this topic, it is worth mentioning that on Shelf No. V was kept 30 cobalt preparations, such as a cobalt specimen from Vienna melted with nitrous acid and prepared with tartar oil, and four compounds made with zaffre (*zaffra*), which indicated impure cobalt arsenate.[79] Among them, one was used for realizing a sample of sympathetic green ink mixed with royal water (*aqua regia*), i.e., a 3:1 mixture of hydrochloric acid and nitric acid. In this regard, it is important to note that 13 different sympathetic cobalt-based inks were preserved, including five green and one rose samples. It can be assumed that Peter Leopold was genuinely interested in sympathetic inks and their chemical principles, a fondness he shared with the cultural milieu of his time. On this, Macrakis highlighted that sympathetic inks gained the interest of the intellectual community at the end of the 18th century. In particular, the progress in knowing the properties of cobalt (e.g., the changing of colors from rose to green and blue when heated) in the early 1700s and the research of chemists such as Jean Hellot (1685–1766) and Pierre Joseph Macquer (1718–1784) allowed the use of these writing tools to become much more advanced and fashionable throughout Europe.[80]

The remaining parts of the catalog gathering the chemical preparations belonging to Peter Leopold present various compounds that could be utilized in precious metal refining as a mixture of Salpeter and ammonia salt to refine gold. These preparations may help us understand the research on mineralogy, chemistry, and mining activities the Grand Duke performed with Giovanni Fabbroni's assistance.

The present results are significant in at least two major respects. Broadly speaking, the compounds belonging to Peter Leopold that were listed in the first part of the chemical cabinet's catalog reflected the institutional reforms he undertook to promote the development of chemical knowledge through the establishment of new research centers as the chemical cabinet at the Royal Museum of Physics and Natural History. The latter soon became a center of great importance from a European perspective, as shown by the French apothecaries with whom Fabbroni was in contact. The cabinet was then equipped with the most up-to-date instruments and tools to perform experiments in pneumatic chemistry, as revealed by the analysis of its catalog. Furthermore, the investigations carried out at the chemistry cabinet, such as those to improve iron processing in Maremma or the research on the characterization of valuable minerals, were aimed, as stated by Mokyr, at expanding the set of useful knowledge and applying natural sciences to solve technological problems and bring about economic growth.[81] To this theoretical framework has to be referred to Fabbroni's publication on the coal deposits, which represented a means to promote mining sciences and activities within the Grand Duchy of Tuscany. In this regard, it is noteworthy that the typologies of the preparations listed in the chemistry cabinet's catalog showed how chemistry was an applied science in 18th-century Italy, closely linked on the one hand with medicine and pharmacopeia and on the other hand with mineralogy and mining processes. On this subject, it is interesting to outline that Peter Leopold asked Thaddeus Rauscher to improve the working methods used in Tuscan iron deposits and ordered Fabbroni to observe the mining and mineral processing techniques utilized in the Habsburg mines. This action followed similar resolutions taken by the rulers of other Italian states that supported the educational travels to the mines of Central and Northern Europe (e.g., to the renewed Chemnitz mining school) for their naturalists, as happened to Marco Carburi (1731–1808) and Matteo Tondi (1762–1835).[82] So, it was not by chance that Fabbroni, after the Restoration, left his position as Director of the Imperial and Royal Museum of Physics and Natural History to be responsible for the mint and the mines of the Grand Dutchy of Tuscany. It should then be noted that the measures Peter Leopold implemented to improve chemical knowledge, such as the establishment of a chemistry chair at the University of Siena in 1771, continued the actions carried on by his father, Francis I, who instituted the first Tuscan chair of chemistry at the University of Pisa in 1757.[83]

Subsequently, the combination of the findings presented in this study raises the possibility that Peter Leopold's interest in chemistry, although influenced by the cultural context in which he lived, could be genuine. For example, he carried out experiments on the diverse phases of matter and performed personal investigations outside the chemistry cabinet, as shown by the orders of chemical products he asked to be delivered in Pisa with urgency at the end of 1786.

About the remaining parts that constituted the catalog of the chemical cabinet, it was considered not to explore them further than the materials presented in the previous section since these inventories are linked neither to Peter Leopold nor to the activities he carried out at the chemical cabinet, covering the volume a period up to 1807. Regarding the inventory of the books kept in the chemistry cabinet, it is interesting to outline that on 14 May 1789, Francesco Favi informed Fabbroni of the publication of Lavoisier's *Traité élémentaire de Chimie* and of the first volume of the *Annales de Chimie*. Favi then advertised to have shipped the books by courier to Fontana's address and asked whether he was to send the ongoing annals or not.[84] According to inventory, the museum acquired for the library of the chemical cabinet both the volumes of Lavoisier's *Traité élémentaire de Chimie* and the first ten issues of the *Annales de Chimie*.

CONCLUSIONS

The present study aimed to examine the history of the chemical cabinet at the Imperial and Royal Museum of Physics and Natural History up to 1790, when Grand Duke Peter Leopold, who founded the museum in 1775, returned to Vienna as Holy Roman Emperor Leopold II. The second goal of this work was to analyze Peter Leopold's interest in chemistry to investigate whether he had a sincere fondness for this branch of science. Thanks to the examination of the chemical cabinet's catalog, whose transcription is presented in the Supplementary Information Files, and the data comparison with other archival sources relative to the Imperial and Royal Museum of Physics and Natural History, this study has found that the establishment of the chemical cabinet fell under the policies for promoting scientific culture Peter Leopold implemented in the Grand Dutchy of Tuscany under his regency. However, one of the more significant findings to emerge from this work is that the establishment of the cabinet and its management also suited the Grand Duke's interests in natural and experimental sciences, disciplines that experienced remarkable advancements in the second half of the 18th century. Peter Leopold's fascination with analytical chemistry is evidenced by the collection of preparations he owned within the chemical cabinet –some of which he made– that were inventoried in the first part of the cabinet's catalog. Their analysis, together with the reconstruction of the activities he carried out with the help of Giovanni Fabbroni, showed the scientific and experimental interest in the chemistry of an Enlightened ruler, who was fully aware of the economic, social, and cultural benefits that would result from the development of chemical knowledge for the territories under his governance.

ACKNOWLEDGMENTS

The authors of this work thank all the Museo Galileo's library staff for their kindness and help in accessing the archival documentation. The authors then acknowledge the financial support of Fondi di Ricerca Scientifica d'Ateneo 2022 – Università degli Studi di Firenze.

REFERENCES

[1] J. Soll in *Florence after the Medici*, (Eds.: C. Tazzara, P. Findlen, J. Soll), Routledge, New York, **2020**, p. 331.

[2] For an overview of the biography literature on Peter Leopold see: M. Rastrelli, *Memorie per servire alla vita di Leopoldo II. Imperatore de' Romani già Gran-Duca di Toscana*, Firenze, **1792**; F. Becattini, *Vita pubblica e privata di Pietro Leopoldo d'Austria granduca di Toscana poi imperatore Leopoldo II*, all'insegna del Mangia, Siena, **1797**; A. Wandruszka, *Pietro Leopoldo: un grande riformatore*, Vallecchi, Firenze, **1968**; G. Capponi, Introduzione alla Istoria civile dei papi; Storia di Pietro Leopoldo, Le Lettere, Firenze, **1976**; H. Peham, *Pietro Leopoldo Granduca di* Toscana, Bonechi, Firenze, **1990**; R. Pasta, *Dizionario Biografico degli Italiani* **2015**, 83, 563–573.

[3] An investigation of the relationship between Peter Leopold, Joseph II and Marie Antoinette is included in A. Ritter von Arneth, *Marie Antoinette: Joseph II Und Leopold Ii: Ihr Briefwechsel: Joseph II Und Leopold II: Ihr Briefwechsel*, K.F. Köhler, Leipzig, **1866**.

[4] A. Wandruszka, *Physis* **1962**, 4(2), 116–124.

[5] L. Bellatalla, *Pietro Leopoldo di Toscana granduca-educatore: Teoria e pratica di un despota illuminato*, Pacini Fazzi Editore, Lucca, **1984**.

[6] L. Maran, M. Castellini, J. Bisman, *Manag. Organ. Hist.*, DOI: 10.1080/17449359.2013.826907.

[7] As an example of Peter Leopold's judicial reforms, it is worth mentioning that the Grand Duchy of Tuscany was the first country to abolish torture and the death penalty on 30 November 1786. More details on this topic can be found in: L. Berlinguer, F. Colao, *La leopoldina nel diritto e nella giustizia in Toscana*, Giuffré, Milano, **1989**; D. Edigati, *Prima della Leopol-*

dina: la giustizia criminale toscana tra prassi e riforme legislative nel XVIII secolo, Jovene, Napoli, **2011**; C.E. Tavilla in *Il diritto come forza, la forza del diritto. Le fonti in azione nel diritto europeo tra medioevo ed età contemporanea* (Eds.: R. Braccia, A. Carrera, A.A. Cassi, E. Fusar Poli, A. Marchisello, G. Rossi, C.E. Tavilla, A. Sciumè), Giappichelli, Torino, **2012**, pp. 151–185.

[8] R. Sarti, *Italy: A Reference Guide from the Renaissance to the Present*, Facts on File, New York, **2004**, p. 38.

[9] L. Maran, W. Funnell, M. Castellini, *A.A.A.J.,* DOI: 10.1108/AAAJ-10-2017-3180.

[10] For a more extensive survey of Peter Leopold's reforms see: C. Mangio, *La polizia toscana: organizzazione e criterio d'intervento 1765*–1808, Giuffré, Milano, **1988**; A. Tacchi, *Medioevo e Rinascimento*, **1992,** 6, 361–373; G. La Rosa, *Il sigillo delle riforme. La costituzione di Pietro Leopoldo di Toscana,* Vita e Pensiero, Firenze, **1997**; M. Rosa, *Settecento religioso. Politica della ragione e religione del cuore,* Marsilio, Venezia, **1999**; V. Baldacci, *Le riforme di Pietro Leopoldo e la nascita della Toscana moderna,* Regione Toscana Press, Firenze, **2000**; A. Contini in *La corte di Toscana dai Medici ai Lorena* (Eds.: A. Bellinazzi, A. Contini) Ministero per i Beni e le Attività Culturali, Direzione Generale per gli Archivi, Roma, **2002**, pp. 129–220; V. Perlato, *Studi Urbinati, A - Scienze Giuridiche, Politiche ed Economiche,* DOI: 10.14276/1825-1676.1070; R. Messbarger in *Florence after the Medici*, (Eds.: C. Tazzara, P. Findlen, J. Soll), Routledge, New York, **2020**, pp. 116–154.

[11] Giovanni Valentino Mattia Fabbroni was one of the most renewed intellectuals in the Tuscan cultural circles of the late 18th century. In their review of the history of analytical chemistry in Italy, Burns et al. stated that Fabbroni's research interests encompassed natural sciences, mineralogical and agricultural chemistry, electrochemistry, and political economy. Fabbroni was also involved in managing the Imperial and Royal Museum of Physics and Natural History, where he worked since he was sixteen years old as an assistant to the director Felice Fontana (1730–1805). In 1780 Fabbroni was appointed as Museum's Vice Director. He became Museum Director in 1825. See: T. Burns, G. Piccardi, L. Sabbatini, *Microchim Acta*, DOI: 10.1007/s00604-007-0769-0. For a more exhaustive Fabbroni's scientific biography see: R. Pasta, *Scienza, Politica e Rivoluzione. L'opera di Giovanni Fabbroni (1752–1822) intellettuale e funzionario al servizio dei Lorena*, L.S. Olschki, Firenze, **1989**; S. Contardi in *Linnaeus in Italy: the spread of a revolution in science* (Eds: M. Beretta, A. Tosi), Science history publications, Sagamore Beach, **2007**, pp. 113–128.

[12] Archivio Museo Galileo (AMG), Fabbroni 10, cc. 1–99.

[13] The Archive of the Royal Museum of Physics and Natural History preserves more than 5700 records about the administration, organization, and management of the Museum from 1775 to 1872. This archival fund represents one of the most important sources of information to investigate the fascinating and complex history of this institution.

[14] On the history of the collections housed at Imperial and Royal Museum of Physics and Natural History see: G. Barsanti, G. Chelazzi (Eds.), *Il Museo di Storia Naturale dell'Università degli Studi di Firenze. Le Collezioni della Specola*, Firenze University Press, Firenze, **2009**; M. Raffaelli (Ed.), *Il Museo di Storia Naturale dell'Università di Firenze – Le collezioni botaniche*, Firenze University Press, Firenze, **2009**; S. Monechi, L. Rook (Eds.), *Il Museo di Storia Naturale dell'Università degli Studi di Firenze. Le collezioni geologiche e paleontologiche*, Firenze University Press, Firenze, **2010**; G. Pratesi (Ed.), *Il Museo di Storia Naturale dell'Università degli Studi di Firenze. Le collezioni mineralogiche e litologiche*, Firenze University Press, Firenze, **2012**; J. Moggi Cecchi, R. Stanyon (Eds.), *Il Museo di Storia Naturale dell'Università degli Studi di Firenze. Le collezioni antropologiche ed etnologiche*, Firenze University Press, Firenze, **2014**; M. Borgheresi, F. Di Benedetto, A. Caneschi, G. Pratesi, M. Romanelli, L. Sorace, *Phys. Chem. Miner*., DOI: 10.1007/s00269-007-0175-5; G. Pratesi, A. Franza, E. Lascialfari, L. Fantoni, F. Malesani, A. Hirata, *Geoheritage*, DOI: 10.1007/s12371-021-00624-1.

[15] On 18 October 1771, Felice Fontana (1730–1805), who was the first director of the Imperial and Royal Museum of Physics and Natural History, received the inventory concerning the scientific instruments and the naturalistic specimens kept in the Royal Gallery (today Uffizi Gallery) that had belonged to the Medici family. AMG, ARMU 001, aff. 1, cc. 331–380. On the scientific Medicean collecting see: F. Camerota, M. Miniati, I Medici e le scienze: strumenti e macchine nelle collezioni granducali, Giunti, Firenze, **2008**. For an investigation of the Medici collections that are still preserved at the Florentine Natural History Museum see: A. Re, D. Angelici, A. Lo Giudice, J. Corsi, S. Allegretti, A.F. Biondi, G. Gariani, S. Calusi, N. Gelli, L. Giuntini, M. Massi, F. Taccetti, L. La Torre, V. Rigato, G. Pratesi, *Nucl. Instrum. Methods Phys. Res. B: Beam Interact. Mater. At*., DOI: 10.1016/j.nimb.2014.11.060.

[16] The Accademia del Cimento, founded by Leopold I de Medici (1617–1675) and Ferdinand II de Medici (1610–1670) in 1657, was the first scientific society in Europe. It remained active until 1667 and was devoted to studying the natural philosophy's principles in the light of the experimental method. During the Accademia's meetings, usually held at Pitti Palace, the participants –among which there were Francesco Redi (1626–1697) and Giovanni Alfonso Borelli (1608–1679)– performed experiments on thermometry, barometry, and pneumatics using the purpose-built instruments that were stored in the grand-ducal residence, until Peter Leopold transferred them to the Imperial and Royal Museum of Physics and Natural History at Torrigiani Palace. See: P. Findlen, Academies, networks, and projects: the Accademia del Cimento and its legacy, *Galilaeana*, 7, **2010**, 277–298.

[17] In the spirit of Enlightenment, useful knowledge covered many disciplines (from today's hard sciences to humanities) that could improve a country's education, progress, and civilization. See: T. Morel, G. Parolini, C. Pastorino (Eds.), The making of useful knowledge, Max Planck Institut für Wissenschaftsgeschichte, Berlin, **2016**. In this scientific, social, and cultural *milieu,* the establishment of the Imperial and Royal Museum of Physics and Natural History can be included among the policies – such as the reforms of hospitals and libraries – Peter Leopold implemented for the public good. See: M.M. Goggioli, *La Biblioteca Magliabechiana. Libri, uomini, idee per la prima biblioteca pubblica a Firenze,* L.S. Olschki, Firenze, **2000**; E. Chapron, Il patrimonio ricomposto. Biblioteche e soppressioni ecclesiastiche in Toscana da Pietro Leopoldo a Napoleone, *Archivio Storico Italiano,* **2009**, 167, 299–346; E. Diana, M. Geddes de Filicaia, *Regolamento dei regi spedali di Santa Maria Nuova e di Bonifazio*, Polistampa, Firenze, **2010**; S. Barchielli, L'Istituto vaccinogeno all'Ospedale di Santa Maria degli Innocenti di Firenze nel XVIII secolo, *Nuncius*, **1998**, 13(1), 247–263.

[18] S. Contardi, *La casa di Salomone a Firenze: l'Imperiale e reale museo di fisica e storia naturale, 1775–1801*, L.S. Olschki, Firenze, **2002**.

[19] For instance, an overview of the museum acquisitions coming from the Habsburg dominions is reported in: M. Benvenuti, V. Moggi Cecchi, L. Fantoni, R. Manca in *Collectio Mineralium. The catalog of Holy Roman Emperor Leopold's II mineralogical collection* (Eds: A. Franza, J. Mattes, G. Pratesi), Firenze University Press, Firenze, **2022**.

[20] A classic example of the objects realized at the museum laboratories is the anatomical models produced at the La Specola wax workshop. See: A. Maerker, Model experts: wax anatomies and Enlightenment in Florence and Vienna, 1775-1815, Manchester University Press, Manchester, **2011**.

[21] Among the museum workshops, Contardi highlighted how the physics cabinet represented one of the most renewed research centers on electrical phenomena at that time and was attended by various scholars such as Angelo Querini (1721–1796), Antonio Vallisnieri junior (1707–1777), and Carlo Barletti (1735–1800). The latter studied the machines that were built there. Furthermore, the physics cabinet also realized mechanical and pneumatic equipment and various optical instruments. See S. Contardi in *Il Museo di Storia Naturale dell'Università degli Studi di Firenze. Le collezioni della Specola* (Eds.: G. Barsanti, G. Chelazzi), Firenze University Press, Firenze, pp. 18–25.

[22] The surviving pneumatic pumps (ca. 13 instruments) are now preserved at the Museo Galileo (Inventory nos. 358–359, 423, 831, 1530–1537, 3777). Among these, it is worth mentioning a pear-shaped glass ampoule that was used to simulate the aurora borealis phenomenon (Inv. no. 423), two vitreous globes for experiments with a bladder in a vacuum (Inv. nos. 358–359), and an air pump that resembles the device described by Jean Antoine Nollet (1770–1770) in his *Leçons de physique expérimentale* (1743–1748) (Inv. no. 1534). On the latter instrument, see: P. Brenni in *The art of teaching physics: the eighteenth century demonstration apparatus of Jean Antoine Nollet* (Eds.: L. Pyenson, J.F. Gauvin), Septentrion, Sillery, **2002**, pp. 11–27.

[23] O. Gori, *Mitteilungen des Kunsthistorischen Institutes in Florenz*, **2002**, 46(2/3), 518–532.

[24] See: G. Scorrano, N. Nicolini, I.M. Masoner, *J. Chem. Ed.*, DOI: 10.1021/ed079p47. Peter Leopold's workbench is now displayed in Room X of the Museo Galileo's permanent exhibition.

[25] Hubert Franz Hoefer was born in Colonia in 1728. There was no news about him until 1765 when he arrived in Firenze along with Peter Leopold's retinue. He remained in Firenze for 25 years and was appointed director of the court's pharmacy. In 1766 he commissioned the *Tabula Affinitatum*, i.e., a table of the chemical affinities between different substances based on Étienne-François Geoffroy's *Table des differents Rapports observés entre differentes substances* (1718), for the apothecary's shop. The *Tabula Affinitatum* is now preserved at the Museo Galileo (Inv. no. 1899). From the early 1780s, Hoefer analyzed the Tuscan springs of Rapolano, San Quirico d'Orcia,

and Rio nell'Elba. Still, his most renewed investigation was the water analysis of Lagone Cerchiaio in Monterotondo Marittimo, during which he discovered the natural boric acid (1778). When Peter Leopold was crowned as Holy Roman Emperor in 1790, Hoefer returned to Vienna, where he died as a court chamberlain five years later. See: G. Piccardi, *La Farmacia granducale di Firenze*, L.S. Olschki, Firenze, **2018**; G. Piccardi, *Nuncius*, **2004**, 19(2), 545–568.

[26] After the renovation of Bagni di Pisa and Bagni di Lucca, Peter Leopold oversaw the establishment of a new thermal center near Montecatini. Following the works of David Becher (1725–1791), who was a physician and a balneologist also known as the "Hippocrates of Karlovy Vary", the Grand Duke ordered chemical analyses on both cold- and hot-water springs to formulate individual hydrothermal treatments. Similar investigations were performed in the Siena area by Giuseppe Baldassarri (1705–1785). See: V. Becagli in *Una politica per le terme: Montecatini e la Val di Nievole nelle riforme di Pietro Leopoldo. Atti del Convegno di studi: Montecatini Terme, 25-26-27 ottobre 1984*, Periccioli, Siena, **1985**, pp. 174–210; G.L. Corinto, *Geotema*, **2019**, 60, 44– 52; A. Guarducci in *Chiare, fresche e dolci acque. Le sorgenti nell'esperienza odeporica e nella storia del territorio* (Ed.: C. Masetti), Cisge, Roma, **2020**, pp. 355–370.

[27] A. Franza, R. Fabozzi, L. Vezzosi, L. Fantoni, G. Pratesi, *Earth Sci. Hist*., DOI: 10.17704/1944-6178-38.2.173; A. Franza, J. Mattes, G. Pratesi, *Collectio mineralium. The Catalog of Holy Roman Emperor Leopold's II Mineralogical Collection*, Firenze University Press, Firenze, **2022**.

[28] M.J. Ventresca, J.W. Mohr in *The Blackwell Companion to Organizations* (Ed.: J.A.C. Baum), Blackwell, Maiden, **2002**, pp. 805–828.

[29] E. Solomon, *Reference/Representation*, 105–108, **2011**.

[30] S. Contardi in *Il Museo di Storia Naturale dell'Università degli Studi di Firenze. Le collezioni mineralogiche e litologiche* (Ed. G. Pratesi), Firenze University Press, Firenze, p. 17.

[31] "Sua Altezza Reale fa sapere all'Aiuto del Direttore del Real Museo Gio: Fabbroni che domane verrà al d. Museo alle ore 2 pomeridiane scendendo alla porta della strada. 3 Settembre 1789." See: AMG, Fabbroni 04, c. 3.

[32] According to Simon, although he did not enjoy the same historical recognition as other pharmacists due to his opposition to Antoine Lavoisier's oxygen theory, Baumé was a qualified chemist. He wrote several works on the crystallization of salts, the phenomena of freezing and fermentation, and the properties of boric acid, Sulphur, and opium. Baumé's research also dealt with the practical use of chemistry. For example, he studied the characteristics of clays and other building materials and created a system for fabric dyeing and silk bleaching. Baumé then proposed a new method for the purification of saltpeter and established the first ammonium salt factory in France. He also perfectioned the empirical hydrometer that today is named after him. On Antoine Baumé see: J. Simon, *Osiris,* DOI: 10.1086/678108; C. Barnard, A. Fones, *Platin. Met. Rev*., DOI: 10.1595/147106712X650811. With regards to his studies on fermentation producers, on 9 November 1787, Baumé complimented Fabbroni for the essay about wine fermentation written by his brother Adamo. AMG, Fabbroni 04, c. 34. The work *Dell'arte di fare il vino* (1787) dealt with producing and conservating wines to be exported. On Adamo Fabbroni (1748–1816) see: R. Pasta in *Dizionario Biografico degli Italiani*, Istituto della Enciclopedia Italiana, Roma, **1993**, vol. 43, pp. 669–673.

[33] Francesco Raimondo Favi was a well-known diplomat in the Parisian political, economic, and cultural *milieu*. Thanks to his intermediation, new scientific publications as well as diverse mechanical drawings, botanical specimens, and scientific instruments reached Firenze in the last decades of the 1700s. See: Z. Ciuffoletti, *Parigi-Firenze 1789-1794. Dispacci del residente toscano nella capitale francese al governo granducale*, L.S. Olschki, Firenze, **1994**.

[34] Francesco Favi, Joseph Latour (dates uncertain), and David Durand (dates uncertain) oversaw the shipments on the French front, while Rocco Bacigalupo (dates uncertain) was responsible for the deliveries from Livorno to Pisa and Firenze.

[35] AMG, Fabbroni 04, cc. 69–72. Another list of chemicals Fabbroni asked Baumé for is given in AMG, Fabbroni 04, c. 57.

[36] AMG, Fabbroni 04, cc. 30–32, c. 52.

[37] R.D. Hoblyn, *Dictionary of terms used in medicine*, printed for Sherwood, Gilbert, & Piper, London, **1844**, p. 228.

[38] AMG, Fabbroni 04, c. 16.

[39] AMG, Fabbroni 04, c. 11.

[40] AMG, Fabbroni 04, c. 12.

[41] Bertrand Pelletier was an apothecary conducting diverse investigations into mineral chemistry at his Parisian laboratory in Rue Jacob 48. Following Jean Baptiste Romé de L'Isle's (1736–1790) studies, he realized salt crystals that were soluble at very slow evaporation. In 1785, he confirmed Carl Scheele's (1742–1786) discovery that chlorine could be pro-

duced from hydrochloric acid and manganese. See: W.A. Smeaton, *Platinum Metals Rev.*, **1997**, 41, 86–88.

[42] AMG, Fabbroni 04, c. 77.

[43] Jean Antoine Chaptal, Count of Chanteloup, was a chemist and a statesman. His studies dealt with the industrial manufacture of soda ash and sodium nitrate. With Louis Nicolas Vauquelin (1763–1829), he determined the alum composition, promoting its industrial production by synthetic means. See: J. Hoff, *Technology and culture*, **1998**, 39, 671–698.

[44] AMG, Fabbroni 04, c. 43. This document is related to a brief note, showing no signature or date, partially written in French, reporting general information about Chaptal and his well-stocked store in Montpellier. A closing remark in Italian says, "I would write for the prices and the kinds of products he sells". See: AMG, Fabbroni 04, c. 48.

[45] AMG, Fabbroni 04, c. 50.

[46] AMG, Fabbroni 04, cc. 101 and 108.

[47] AMG, Fabbroni 04, cc. 80–81.

[48] AMG, Fabbroni 04, c. 2. The document heading reported that if Fabbroni did not need to file the order note, this would be sent back to His Royal Highness.

[49] AMG, Fabbroni 04, c. 104.

[50] The *Pyrophorus Hombergii* was a flammable compound realized from human fecal matter created by the Dutch chemistry and physician Wilhelm Homberg (1652–1715) in the early 18th century. See: L. Principe, *The Transmutations of Chymistry. Wilhelm Homberg and the Académie Royale Des Sciences*, University of Chicago Press, Chicago, **2020**.

[51] AMG, Fabbroni 04, cc. 89–90.

[52] AMG, Fabbroni 04, c. 86.

[53] AMG, Fabbroni 04, c. 84.

[54] The lute was usually used in chemical distillation processes to seal a vessel and prevent the dispersion of the stem while protecting its surface from heat. See: S.J. Linden, The alchemy reader: from Hermes Trismegistus to Isaac Newton, Cambridge University Press, Cambridge, **2003**.

[55] AMG, Fabbroni 04, c. 126.

[56] AMG, Fabbroni 04, cc. 96 and 98.

[57] AMG, Affari 002, c. 355.

[58] AMG, Fabbroni 04, c. 30.

[59] AMG, Fabbroni 02, cc. 9–12.

[60] AMG, Fabbroni 02, cc. 2–8. The original title was *Relazione sopra le miniere di ferro nel Granducato di Toscana e saggi sopra le medesime fatte da Taddeo Glauscher di Carintia nel 1780*. Taddeo Glauscher can be identified with the mining expert Taddeus Rauscher, since the correct spelling of his name is mentioned in the opening of Fabbroni's report.

[61] AMG, Fabbroni 03, cc. 268–271.

[62] AMG, Fabbroni 03, c. 253.

[63] AMG, Fabbroni 03, c. 251.

[64] AMG, Fabbroni 03, c. 275 and 404.

[65] AMG, Fabbroni 10, cc. 1–99.

[66] The full title of the second catalog's section is *Serie di preparazioni chimiche risultanti dall'analisi dei tre regni della natura e che si conservano in piccoli saggi nella seconda stanza del Laboratorio* (Series of chemical preparations resulting from the analysis of the three kingdoms of nature and stored in small samples in the second room of the Laboratory).

[67] Acid and light cacao butter oil (*Acido e Olio leggiero del Burro di Caccao*), precipitated copal of the Spirit of Wine with the effusion of water (*Coppale precipitata dello Spirito di Vino con l'effusione dell'acqua*), cherry-laurel spirit (*Spirito di Lauro-ceraso*), essential oil and resinous part of laurel berries separated from its fixed oil by means of wine spirit (*Olio essenziale e parte resinosa delle Bacche di Alloro separati dal suo Olio fisso col mezzo dello Spirito di Vino*), aromatic part of wine dissolved in wine spirit (*Parte odorante del Vino disciolta nello Spirito di Vino*), litmus tincture (*Tintura di Tornasole*), litmus starch or mold laquer (*Fecula del Tornasole ò Lacca muffa*).

[68] In his essay on Ticunas, Fontana stated that he had the opportunity to study this vegetable-origin poison at the beginning of his stay in London (approximately in the summer of 1778). The English version of the essay's title reported the compound as belonging to the Grand Duke of Tuscany. Since no other mention of this fact was written neither in Fontana's text nor in the chemistry cabinet's catalog, it is impossible to include this sample within the preparations Peter Leopold owned. See: F. Fontana, *Philosophical Transactions of the Royal Society of London*, **1780**, 70:163–220, ix–xlv.

[69] The third catalog's section was entitled *Libri d'arte che si conservano nel Laboratorio* (Arts books kept in the Laboratory).

[70] They were Scopoli's *Materie spettanti alla chimica* and the Italian edition of Macquer's *Dizionario di chimica*.

[71] The transcription reports as precisely as possible how the books have been inventoried; therefore, possible spelling errors have not been corrected unless they made the text unintelligible.

[72] Here is the original text: *un Tasso ed una Bicornia sul medesimo Coppo lustri e intagliati a bulino con vari fregi e coll'Arme de Medici*.

[73] AMG, ARMU Affari 004, c. 307.

[74] "Lac Martin fatta da S.A.R. 29 dicembre 1780" in the original text.

[75] U. Schiff, M. Betti, *Archeion*, **1928**, 9:290–324.

[76. "Lac Martin del Bechi" in the text.

[77] See: *Tomo Settimo delle Gazzette Toscane uscite settimana per settimana nell'anno 1772,* Appresso Anton Giuseppe Pagani Stampatore e Librajo delle Scalere di Badia, Firenze, **1772**, No. 12, unnumbered page; P. Landini, *Istoria della venerabile Compagnia di Santa Maria della Misericordia della città di Firenze con i capitoli, riforme, e catalogo di tutti i capi di guardia dal suo primo principio*, nella stamperia di Pietro Allegrini alla Croce Rossa, Firenze, **1786**, p. CXVII;

[78] T. Wright, *The Universal Pronouncing Dictionary, and General Expositor of the English Language*, the London and New York Printing and Publishing Company, London and New York, vol. 5, p. 267.

[79] F. Flügel, J.G. Flügel, *A practical Dictionary of the English and German Languages: Deutsch- Englisch,* Julius Richter, Leipzig, **1861**, p. 1128.

[80] On the history of sympathetic inks, their usage in the late 18th century, and chemical research about their composition, see: J. Wisniak, *Revista CENIC Ciencias Químicas*, **2009**, 40(2): 111–121; C. Lehman, *Ambix*, DOI: 10.1179/174582310X12629173849881; K. Macrakis, *Prisoners, Lovers, and Spies: The Story of Invisible Ink from Herodotus to al-Qaeda*, Yale University Press, New Haven, **2014**.

[81] J. Mokyr, *The Journal of Economic History*, **2005**, 65(2):285–351.

[82] R. Vergani, *Quaderni Storici,* **1989**, 24: 123–141; C. Guerra, *Lavoisier e Partenope: contributo ad una storia della chimica del regno di Napoli*, Società Napoletana di Storia Patria, Napoli, **2017**.

[83] G. Fochi, *Annali di Storia delle Università Italiane,* **2010**, 14: 207–216.

[84] AMG, ARMU Affari 002, aff. 94, c. 374.

Finito di stampare da
Logo s.r.l. – Borgoricco (PD) – Italia

September 2022 No walls. Just bridges

Substantia

An International Journal of the History of Chemistry

Vol. 6 – n. 2

Table of contents