Aptamers for Detection and Diagnostics (ADD) is a proposed mobile app acquiring optical data from conjugated quantum nanodots to identify molecules indicating presence of SARS-CoV-2 virus: Why public health and healthcare need smartphone sensors as a platform for early detection and prevention

Shoumen Palit Austin Datta1,2,3,*, Brittany Newell4, James Lamb5, Yifan Tang6, Patrick Schoettker7, Catherine Santucci8, Theresa Gräfin Pachta8, Sanjay Joshi10, Oana Geman11, Diana C. Vanegas12, Carmen Gomes13, Pramod P. Khargonekar14, Molood Barati15 and Eric Scott McLamore16

1 MIT Auto-ID Labs, Department of Mechanical Engineering, Massachusetts Institute of Technology, Room 35-206, 77 Massachusetts Avenue, Cambridge, MA 02139, USA (soumen@mit.edu)
2 MDPnP Interoperability and Cybersecurity Labs, Biomedical Engineering Program, Department of Anesthesiology, Massachusetts General Hospital, Harvard Medical School, 65 Landsdowne Street, Suite 232, Cambridge, MA 02139, USA (sdatta8@mgh.harvard.edu)
3 NSF Center for Robots and Sensors for Human Well-Being (RoSeHuB), Collaborative Robotics Lab, School of Engineering Technology, Purdue University, 193 Knoy Hall, West Lafayette, IN 47907, USA
4 Adaptive Additive Technologies Lab (AATL), School of Engineering Technology, 189 Knoy Hall, Purdue University, West Lafayette, IN 47907, USA
5 Saturn Cloud Inc and former IoT Data Scientist, Amazon Web Services (AWS), Amazon Inc, USA
6 Biosystems Engineering, Department of Environmental Engineering and Earth Sciences, Clemson University, Clemson, SC 29631, USA
7 Department of Anesthesiology, Lausanne University Hospital and University of Lausanne, Rue du Bugnon, CH-1011 Lausanne-CHUV, Lausanne, Switzerland
8 Barts and the London School of Medicine and Dentistry, Queen Mary University of London Malta Campus, Triq l-Arċiṣqof Pietru Pace, Victoria, Gozo, VCT 2520, Malta
9 Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Mile End Road, Bethnal Green, London E1 4NS, London, UK
10 Industry CTO Healthcare, Global CTO Office, Dell EMC, Hopkinton, MA 01748, USA
11 Department of Health and Human Development and Department of Computers, Electronics and Automation, Stefan cel Mare University of Suceava, Strada Universităţii 13, Suceava 720229, Romania
12 Biosystems Engineering, Department of Environmental Engineering and Earth Sciences, Clemson University, Clemson, SC 29631, USA
13 Department of Mechanical Engineering, Iowa State University, Ames, Iowa 50011, USA
14 Vice Chancellor for Research, University of California, Irvine and Department of Electrical Engineering and Computer Science, University of California, Irvine, California 92697, USA
15 School of Engineering, Computer and Mathematical Sciences Auckland University of Technology, Auckland 1010, New Zealand
16 Department of Agricultural Sciences and Department of Civil and Environmental Engineering, Clemson University, Clemson, SC 29634, USA

*Corresponding author – Dr Shoumen Datta shoumen@mit.edu or sdatta8@mgh.harvard.edu
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ADD for SARS-CoV-2</td>
<td>3</td>
</tr>
<tr>
<td>2. Aptamers for Detection and Diagnostics (ADD)</td>
<td>4</td>
</tr>
<tr>
<td>3. ADD Digital Data Design</td>
<td>5</td>
</tr>
<tr>
<td>4. Beyond ADD</td>
<td>8</td>
</tr>
<tr>
<td>5. Prevention follows Detection and Diagnostics</td>
<td>10</td>
</tr>
<tr>
<td>6. Alternative Arrays</td>
<td>12</td>
</tr>
<tr>
<td>7. Array of Targets</td>
<td>14</td>
</tr>
<tr>
<td>8. Discussion</td>
<td>21</td>
</tr>
<tr>
<td>9. Complexity of the biomedical scenario, socio-economic catastrophe and public health crisis</td>
<td>23</td>
</tr>
<tr>
<td>10. Post-Pandemic Health &amp; Healthcare: Broad Spectrum use of Sensors with Smartphones</td>
<td>28</td>
</tr>
<tr>
<td>11. Comments</td>
<td>33</td>
</tr>
<tr>
<td>APPENDIX I</td>
<td>35</td>
</tr>
<tr>
<td>Brief Description of the Components for SARS-CoV-2 ADD Decision System</td>
<td></td>
</tr>
<tr>
<td>APPENDIX II</td>
<td>37</td>
</tr>
<tr>
<td>THE PROPOSED “SENSIBLE” SYSTEM AND APTAMER-AS-A-DRUG (AAAD)</td>
<td></td>
</tr>
<tr>
<td>APPENDIX III</td>
<td>41</td>
</tr>
<tr>
<td>THE MOLECULAR BASIS OF DISEASE</td>
<td></td>
</tr>
<tr>
<td>APPENDIX IV</td>
<td>48</td>
</tr>
<tr>
<td>APTAMERS as ADJUVANTS and/or ALTERNATIVE to ANTIBODIES</td>
<td></td>
</tr>
<tr>
<td>APPENDIX V</td>
<td>52</td>
</tr>
<tr>
<td>ENGINEERING RNASE RESISTANT RNA APTAMERS</td>
<td></td>
</tr>
<tr>
<td>APPENDIX VI</td>
<td>52</td>
</tr>
<tr>
<td>SCIENCE, SOCIETY VS SENSITIVITY, FREQUENCY OF TESTING AND COST</td>
<td></td>
</tr>
<tr>
<td>APPENDIX VII</td>
<td>56</td>
</tr>
<tr>
<td>ROOM FOR IN SILICO CREATIVITY IN PREDICTING APTAMER DESIGN?</td>
<td></td>
</tr>
<tr>
<td>APPENDIX VIII</td>
<td>61</td>
</tr>
<tr>
<td>FIELD TESTING THE APTAMER SENSOR, ALONE, MAY NOT BE PRUDENT</td>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>64</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>66</td>
</tr>
<tr>
<td>Cartoon for Reference #371</td>
<td>115</td>
</tr>
</tbody>
</table>

2 ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
Aptamers for Detection and Diagnostics (ADD) is a proposed mobile app acquiring optical data from conjugated quantum nanodots to identify molecules indicating presence of SARS-CoV-2 virus: Why public health and healthcare need smartphone sensors as a platform for early detection and prevention

Datta et al

“CITCOM” precedes this proposal. PDF is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017

ABSTRACT

Engineering a biomedical device as a low-cost, non-invasive, detection, and diagnostic platform for surveillance of infections in humans, and animals. The system embraces the IoT “digital by design” metaphor by incorporating elements of connectivity, data sharing and (secure) information arbitrage. Using an array of aptamers to bind viral targets may help in detection, diagnostics, and potentially prevention in case of SARS-CoV-2. ADD may spawn a broader approach from ADD diagnostics to AAAD (aptamer-as-a-drug) therapeutics. CoVID-19 as an airborne endemic infectious disease (EID) may be analogous to a social IED (improvised explosive device) which may profoundly transform geopolitical economy. Desperate demand for alternatives to high energy consuming RNA vaccines may be partially met by lower cost DNA and RNA aptamers both for diagnostics and for therapeutics, globally.

1. ADD for SARS-CoV-2

The scale of mortality and morbidity due to SARS-CoV-2 evokes us to explore unconventional approaches to mitigate the risks presented by pandemics. Scientists may be less aware of the discovery of aptamers thirty years ago but the “fit” of aptamers with respect to the molecular biology of the current problem makes it worthwhile to propose new tools. Innovation may arise from the combination of chemistry and molecular biology with sensor engineering and opportunity for data dissemination to benefit public health by integrating the principle of internet of things (IoT as a design metaphor).

<table>
<thead>
<tr>
<th></th>
<th>Emergence</th>
<th>Cases</th>
<th>Fatality Rate</th>
<th>Transmissibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS</td>
<td>2003</td>
<td>8,098</td>
<td>11%</td>
<td>+</td>
</tr>
<tr>
<td>MERS</td>
<td>2011</td>
<td>2,519</td>
<td>34%</td>
<td>+</td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td>2019</td>
<td>~100 million</td>
<td>0.5-1% Est</td>
<td>+++</td>
</tr>
<tr>
<td>SARS-CoV-3?</td>
<td>??</td>
<td>??</td>
<td>??</td>
<td>??</td>
</tr>
<tr>
<td>SARS-CoV-4?</td>
<td>??</td>
<td>??</td>
<td>??</td>
<td>??</td>
</tr>
</tbody>
</table>

Table I: Is the eight year interval between SARS, MERS and COVID-19 just an unrelated coincidence?
2. Aptamers for Detection and Diagnostics (ADD)

Single stranded (ss) DNA aptamers (ssRNA\(^2\) are equally useful\(^4\) but susceptible to degradation by ribonuclease) bind with specificity to SARS-CoV-2\(^9\) proteins (Nucleocapsid, Spike, Nsp1). Aptamers are conjugated with carbon or cadmium quantum (QT) nano-dots. If there are viruses (1, 10, 100, 1000) in a sample (sputum, saliva) at a detectable level, then it triggers QT.DNA (QTD) conjugated complex to transmit optical property change (EIS or electrochemical impedance spectroscopy is another option for signal transduction). An optical signal transduction mechanism may offer low cost data acquisition, enabling billions of people to use ADD (detection tool) at home or anywhere (AHA). The end-user must have access to the "QTD" conjugate (distributed by health departments in hamlets, towns and cities). QTD (product) may be a slurry in a tube labeled as "CoV-2-DETECTION & DIAGNOSIS" (C2DD). It remains to be investigated if inclusion of endo-b-N-acetylglucosaminidases (ENGases\(^10\)) in the slurry may be necessary to expose the binding sites by partially removing the N-glycan coat if the viral Spike protein is the target (Figure 20). Imagine C2DD as a tube of lip balm or similar form factor. For supply chain and logistics, it will reduce operational cost of distribution if C2DD may be shipped as a tamper-proof sterile vial without the need for cold supply chain or special storage to extend shelf-life.

First, end-user uses her smartphone holo-lens "QTD" app (not limited to Microsoft HoloLens, the concept can be re-developed anywhere to reduce cost) to take an image of the C2DD vial/tube without sample (no virus). Priming (tuning) step is critical to establish a baseline for signal transduction and app-embedded data analytics engine to set the system to "without virus" ground state to obtain an optical "ground zero" (baseline will be different for EIS). Open question for instrumentation is the need for UV activation (for traditional nanodots) to record the shift (valence electron transfer). Can the app be configured to perform the activation and record the photoluminescence change? Using visible light to activate and coupling activation/quenching with the app needs innovative chemical/device engineering.

Second, the end-user spits (or adds a small volume saliva or sputum using a swab/spoon) in the test tube (vial). There is room for controversy in this step but it is the easiest non-invasive procedure.

Third, end-user uses her smartphone holo-lens "QTD" app to record optical change (as soon as possible after adding saliva/sputum). Perhaps similar to bar code or EPC or QR code scanning.

Fourth, end-user uses her smartphone holo-lens "QTD" app to record optical changes every 5 minutes for 30 min (from the time of adding the sample). There will be questions about ENGase activity, binding kinetics of the aptamer, signal to noise ratio ([filtering algorithms (Kalman\(^11\) filter), error correction], activation/quenching issues, damping of signals due to interference from host proteins, salinity and pH of mucus-mucin/saliva/sputum sample (any or all could jeopardize binding and signal).
3. ADD Digital Data Design

Baseline versus change over time will appear as a plot in the app (analytics, Figure 1, uses basic machine learning (ML) tools, for example, SVM or support vector machine). Fool-proof visualization by generating a "traffic signal" visual [green oval (NO virus detected); red oval (virus detected); yellow oval (inconclusive/ambiguous)]. Data gathered by the smartphone app (if enabled by user) to be transmitted to national centers of epidemiology (eg CDC in USA, ECDC in EU) and local hospitals (the choice will be user-dependent). Allowing collection of anonymized data may be one alternative (without recording IPv6/IPv4 addresses) but pros/cons to be considered for the greater good, public safety and privacy12.

This app is a "frontline" detection tool which may be used everyday or each week, At Home or Anywhere (AHA), by individual users. The "C2DD" vial has no therapeutic value. Positive results (red oval - virus detected) may have to be re-confirmed using lab tests (PCR, mAbs) in a clinic or hospital. C2DD PRODUCT and associated SERVICE "QTD" app if combined, are data-informed tools. It does not offer or guarantee further testing or treatment. Distribution and pricing of the hypothetical C2DD product and proposed pay-per-use (PAPPU13) service for QTD will be debated by corporations. Free distribution of C2DD and a micro-payment model (pay-per-use) for the "QTD" app is advocated.

Users may hide or selectively control data/information sharing as well as access to surveillance data (data from daily screening for infection by the infectious agent in question). Secure sharing of surveillance data by users (citizen science) is recommended to generate a robust and representative status of the community or infected demographics in the region in terms of molecular epidemiology.

In general, data from molecular epidemiology is critical for resource-constrained healthcare supply chains to optimize planning (humanitarian logistics), allocate human resources (medical professionals) and organize transportation of materials to areas where assistance is needed. Citizen science14 efforts are germane for the efficacy of healthcare systems in case of widespread infections (epidemics/pandemics). The tools which makes citizen science possible and effective may be viewed as global public goods. Similar systems for animal surveillance (farms, cattle, poultry, meat) are necessary to reduce infection in domestic animals (pets) and from crippling the food supply chain.

Components of the ADD system (QTD, C2DD) including mobile data collection, information arbitrage and public health applications are not limited to SARS-CoV-2 but is a platform approach which includes digital design elements illustrated15 in Figure 1. Citizen science supported public health may immensely benefit from detection of viruses, bacteria, fungi, prions or any infections agent as long as an aptamer (oligonucleotide based on the idea16 of an “anti-sense” approach17) may bind a small molecule or a macromolecule (peptides or proteins) with sufficient specificity, sensitivity and selectivity to generate credible data which may be distributed in real-time to inform and initiate subsequent steps. Scoring data from test sample, negative and positive control (for same person) will improve accuracy.
Figure 1: ADD system also includes data acquisition, analytics and data distribution which includes elements of the concept referred to as internet of things (IoT) which is a “digital by design” metaphor. Cartoon shows the potential path of raw data from the hypothetical binding between a sensor and a target molecule. Raw data from signal transduction due to binding activity is transmitted and acquired at the “edge” by the smartphone. The raw data is “processed” using tools either at the edge (embedded operations in the smartphone) or data may be uploaded to the cloud. Post-cloud computing analytics is returned to the edge device for display within an ADD application portal on the smartphone. The choice between edge versus cloud computing is a function of infrastructure (availability of wireless bandwidth, at the edge). The user may observe a difference in the time that it takes to process the data and display information (delayed visualization due to latency, function of bandwidth and speed).

Data scoring and processing is recommended due to variability of systems. ADD proposes the use of aptamers but other alternative arrays (see section 6) which may use the general approach (above) may “weigh” the information based on probability of false positive / false negative outcomes from tests (separate from false positives / false negatives in machine learning models). Assigning weighted risk to data and running other analytics can be performed on the mobile device (smartphone) or in the cloud, depending on access to and quality of telecommunications infrastructure. Cloud computing could add latency between upload and display of information or prediction, depending on availability, reliability and connectivity to the internet. Several regions of the world still lack sufficient access to the internet.
Scoring, processing and assigning risk within the analytical engine may benefit from machine learning (ML) tools to create a set of models or rules, to be described by and agreed on by experts. The system may scan and screen the image or data from the holo lens app (optical signal) to compare with these models or rules. Assigning an “image risk score” (IRS) may influence the presentation of the raw data where the “traffic signal” “red” may be provided with a sub-text containing a confidence score or include a qualitative comment (likely presence of virus) associated with a Likert-type indicator/scale.

In any procedure, enabling the IRS to influence the raw data must be stringently controlled. Models or rules must be agreed by global experts whose credibility may be above question. Hence, these models and rules must stay outside the realm of testing services or labs or groups that are involved with creating systems, for example, ADD. It is preferable if model scoring (assigning risk score to an upload) runs on a platform which is not influenced by the local operator or the mobile user. The smartphone uploading the data may use a “tool” that applies the IRS engine residing in a secure infrastructure in a public cloud (FDA, NIH, NSF, CDC, ECDC) using appreciable level of cybersecurity (eg. Microsoft Azure, Amazon AWS). The smartphone must have the permission and physical availability to wireless internet or mobile data network to remotely access analytical tools in the cloud, such as, the IRS engine.

One alternative is to install (and update) the data scoring models/rules (IRS engine) in the ODS (operational data store, see Figure 1). The ’message broker’ receives uploads and sends them to the ODS, which serves “hot” data to the app. ODS database is tuned for rapid reads, and serves requests made by the mobile app (only recent uploads and metadata about those uploads, including “risk scores”).

Risk scores are generated from models which are trained from historical data relevant to the test in question (using aptamer or antibody or other molecules, for example, hACE2). There must be access to sufficient statistical data from each type of test to create a credible risk score. If the model is based on bad data (garbage in), the risk score and IRS engine will spew bad information (garbage out). The model’s responsibility to assign “risk score” impacts the “traffic signal” and could alter the outcome. Model training requires vast quantities of historical data, curated and pooled across multiple users who used the test and verified their outcome. If the binding was positive it must be corroborated by PCR or another test with even higher specificity to confirm the result from the binding test using ADD tool.

Model-building is an iterative exercise that requires lab data from testing to be evaluated by credible scientists before data scientists can use it (curate?) to train ML models, which are error prone. In model scoring, a model (in the IRS engine) is called to act on the uploaded (input) data. This analysis generates a prediction, displayed on the smartphone as information or recommendation for the user. The outcome the user views depends on the design choices made in ML model training. It is absolutely central that model scoring requires “features” (characteristics germane to model-analysis). Creating features is the task of a team of specialists (scientists collaborating with data experts). Harvesting

---

7 ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
feature vectors and data relevant to the feature is the task of feature extractors. It may be provided by humans or we may use automated selection/extraction to generate features from raw data.

4. Beyond ADD

Scientific and engineering challenges to design ADD must embrace trans-disciplinary activities. But, no new physics is necessary. ADD may be available to billions, as a low-cost mobile AHA (at-home) product linked to IoT-type service app. The user experience is related to the service, not the product. The convergence of hardware and software with science and engineering as well as analytics and machine learning to meaningfully ascend the DIKW pyramid (data, information, knowledge, wisdom) is key to creating any detection platform where other tools and devices may upload data using open data APIs and standards-compliant data interoperability (DDS) tools to aggregate or explore cumulative analytics, integrated with other systems, for example, geographic information systems or GIS.

In the broader spectrum, ADD is an embryonic element of a potential global health surveillance platform (GSP) which may be pivotal as an early warning signal for humans and animal farms. Lessons from tsunami detection are sorely missing from public health policy discussions. Implementation of GSPs are neither a part of any local public health strategy nor on the agenda of precision population health management organizations (CDC, ECDC, WHO).

An important element of the global health surveillance platform (GSP) may include data from non-invasive profiling, referred to as “pay-per-pee” healthcare, which may be instrumental in molecular profiling for longitudinal studies on health and wellness. GSPs may try not to dwell on genomics (DNA) and expression (RNA) in imprecision medicine but include proteomics because gene expression is insufficient unless the functions are implemented by proteins. Aptamers in proteomic profiling (GWAS, metabolomics) and other applications including ADD may benefit from synergistic integration to help predict status of health (collected papers provide select applications of aptamer).

Genomics is a “snapshot” (static structure of the infrastructure) and transcriptomics (RNA, GTEx) is an indicator of expression, which is data, but data may not (always) contain information. Proteins bind in a myriad of ways and translates data to usable information to maintain standard dynamic operating procedures (physiology, homeostasis, metabolomics).

Proteomics is a “time series” but its analysis over time may be interrupted due to feasibility and logistics of implementing programs like pay-per-pee healthcare, not to mention the complexity involved in extracting sense, often cryptic, from thousands of protein profiles, over time. Static protein profiles using NMR and mass spec tools only capture snapshots. Can proteomics make sense of a cytokine storm as markers of counter-anti-inflammatory response even before the infectious agent is detected? Perhaps it is utopian to expect proteomic profiling as a daily practice in healthcare and home-health.
Figure 2: (Top) Pay-per-pee healthcare may provide time series data for precision medicine. (Left) “Collection of saliva samples by patients themselves negates the need for direct interaction between health care workers and patients. This interaction is a source of major testing bottlenecks and presents a risk of nosocomial infection. Collection of saliva samples by patients themselves also alleviates demands for supplies of swabs and personal protective equipment. Given the growing need for testing, our findings provide support for the potential of saliva specimens in the diagnosis of SARS-CoV-2 infection.”
5. Prevention follows Detection and Diagnostics

If viewed as non-classical antibodies then the role of aptamers vastly exceeds that of detection. It spills over into prevention, perhaps as an alt-vaccine, albeit non-immunogenic. Identifying aptamers that can detect viral proteins in saliva implies that the aptamers may also bind the same protein (albeit with altered kinetics) if administered topically (nasal spray, throat spray, soft-mist inhaler). Protecting the naso-pharyngeal area by saturating it with aptamers which binds (irreversibly?) to proteins from respiratory viruses (SARS) may be a preventative measure. Asymptomatic, pausi-symptomatic and COVID-19 patients clearly expressing symptoms associated with SARS-CoV-2 may continue application of the aptamer cocktail to reduce the spread of infection by disabling nascent virions. Aptamers preventing the spike protein (S1 RBD) of SARS virion from attaching to the ACE-2 viral receptor protein of uninfected cells may slow down the infection and development of COVID-19.

It follows that aptamers can also bind to any or all viral proteins not only in the extracellular space but also inside the cell. Delivering a portfolio of functional aptamers inside the cytosol must face the challenges posed by bio-availability and toxicity due to the potential for perturbing functions of essential cellular proteins. Creating aptamers as alt-vaccines for any infecting organism (virus, bacteria, fungi, prion) which uses a protein in its lifecycle may be an (~30 year) old idea. Will the use of aptamers gain greater prominence in global public health practices, as a low-cost global public goods tool to contain the current and future epidemics and/or pandemics, worldwide, in humans and animals?

Single stranded RNA or ssDNA aptamers are not linear “tapes” but 3-dimensional shapes as illustrated by the discovery of tRNA by Paul Zamecnik, Mary Louise Stephenson and colleagues at MGH, HMS. Publication of the discovery of tRNA by Zamecnik in 1958 catalyzed an array of milestones including the discovery of mRNA by Brenner and Gros as well as the lac operon model of feedback inhibition by Jacob and Monod, all three published in 1961. The role of proteins in regulation emerged as central to physiology and metabolism. In transcription, translation and replication the binding between proteins and nucleic acids acted as a “switch” (mechanism of action). The notion of aptamers germinated in 1990 but it drew on knowledge from binding between oligonucleotides and proteins. Aptamers may be 20-60 oligonucleotides or more. Binding specificity of an enriched pool may be orders of magnitude different (Kd) between a nearest neighbor or an analog. Sequential steps are necessary from a starting sample (for example, 9×1014 ssDNA oligonucleotides) to arrive at an enriched pool of aptamers (19 ssDNA aptamers). The process has evolved in complexity and unique structures may be involved in conferring specificity. In many applications of aptamers the debate also involves issues pertaining to trust and doubts due to the constant demand for increasing accuracy and precision with respect to sensitivity, selectivity and specificity, in detection and diagnostics.
Current and future\textsuperscript{74} application\textsuperscript{75} of aptamers include chemistry\textsuperscript{76}, chemotherapy\textsuperscript{77}, food\textsuperscript{78} safety, diagnostics\textsuperscript{79}, antibodies\textsuperscript{80}, alt-vaccines\textsuperscript{81}, imaging\textsuperscript{82} and different\textsuperscript{83} types\textsuperscript{84} of biosensors\textsuperscript{85}. ADD as a detection tool for SARS-CoV-2 proposes aptamer-based sensors (aptasensors) to detect SARS-CoV-2 proteins. When an aptamer binds with the target, the signal (data) will be transduced and captured by a mobile device. Analytical tools will process data and display information on smartphones (Fig 1). Data dissemination will follow according to user preferences, to inform public health authorities or hospitals.

Optimism for aptamers as detection tools\textsuperscript{86} extend to SARS-CoV-2 due to the detection of SARS-CoV (etiologic agent of 2008 SARS epidemic) C-terminal of N (nucleocapsid) protein at a concentration as low as 2 picograms/mL using a RNA\textsuperscript{87} aptamer in a nanoarray. Tests using saliva\textsuperscript{88} may be unsuitable for RNA\textsuperscript{89} aptamers due to presence of ribonuclease\textsuperscript{90} (RNase). DNA aptamers previously shown to bind to the N protein of SARS-CoV (K\textsubscript{d} 4.93±0.3nM\textsuperscript{91}) also\textsuperscript{92} binds to the N protein of SARS-CoV-2. The N protein\textsuperscript{90} of SARS-CoV-2 shares 91% sequence homology with the N protein\textsuperscript{94} of SARS-CoV but is less similar (16% - 38%) with N protein from the other 5 known human coronaviruses. Thus, detection\textsuperscript{95} of N protein in saliva using an aptamer-based ADD aptasensor is possible. Aptamer-based technologies\textsuperscript{96} directed toward SARS-CoV-2 Spike protein are gaining\textsuperscript{97} momentum\textsuperscript{98}. Blocking\textsuperscript{99} the S protein from attaching to hACE-2 may perturb viral entry and prevent\textsuperscript{100} the spread of infection.

Aptamers created against the S1 RBD\textsuperscript{101} may block binding to hACE-2 (internally) or serve as a detection tool (external ADD aptasensor) to test saliva/sputum for SARS-CoV-2. Other\textsuperscript{102} SARS-CoV-2 targets\textsuperscript{103} including Nsp1\textsuperscript{104} may be less accessible in saliva because they are synthesized after viral entry. But, during the burst cycle, when new virions are released, viral proteins inside the host cell may be exposed. The targets are not limited to external viral proteins (spike, nucleocapsid, envelope proteins; Figure 4).

Signal transduction and data acquisition follows detection. In addition to EIS (electrochemical impedance spectroscopy\textsuperscript{105}) signals, optical signals are preferred because data acquisition using cameras and apps in smartphones are feasible in locations where resources may be limiting. Protein\textsuperscript{106} detection\textsuperscript{107} by conjugating aptamers with quantum dots\textsuperscript{108} is a tried\textsuperscript{109} and true\textsuperscript{110} process\textsuperscript{111} which may be the optical signal (data) for this system. Changes in optical characteristics due to binding may be captured by cameras on mobile phones or HoloLens\textsuperscript{112} app in smartphones may scan the saliva sample (think barcode or QR\textsuperscript{113} code scan). Cameras (sensors) associated with the holo-lens (Kinect\textsuperscript{114}) can scan the “field” and collects data to create a digital geometry\textsuperscript{115} (digital model, 3D image). For ADD, HoloLens tools required for holographic functions\textsuperscript{116} may be unnecessary, for example, accelerometer (speed of movement), gyroscope (tilt, orientation) and magnetometer (compass). Optical data captured from saliva containing testing vials will be analyzed (machine learning tools; see Figure 1) followed by visualization of information on the mobile device and (secure) information arbitrage, if authorized.

\textsuperscript{11} ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
6. Alternative Arrays

Figure 3: **Multiplexed** Detection Tool for SARS-CoV-2. Upper panel presents potential recognition and detection chemistries. The data (fusion, middle panel) will be analyzed and assay results displayed (bottom). The data and information will be displayed on a mobile device (see cartoon in Figure 1). Three distinct binding targets for SARS-CoV-2 spike protein RBD are presented in sections A, B (ADD) and C.

In (A) RBD-antibody (SARS-CoV-2 strain specificity) is functionalized with EDC-NHS chemistry to metal (gold, Au) nanoparticles (or may be attached/adsorbed on laser inscribed graphene, LIG).

In (B) single-stranded DNA aptamers with thiol linker is adsorbed to metallized LIG (ADD aptasensor).

In (C) histidine-tagged human ACE2 is adsorbed to metallized LIG. (D) Binding elicits signal (EIS, impedance spectroscopy) which is transduced to a mobile device. Data acquisition is followed by “hot” data upload to embedded tiny database (tinyDB) in the device (ODS in Figure 1). Analytics may be executed on the device (embedded logic, machine learning tools) or uploaded to cloud server. The data fusion (model scoring) step may be necessary to make sense of the data, *in combination*, to provide not only raw data (results from A, B and C) but information, extracted from data and processed according to a simple SNAPS paradigm to convey the *meaning of the outcome*, to inform the non-expert end-user.
Interpretation of data may be necessary due to the caveats of target binding and recognition. The specificity of the antibody used in the tool may not bind or bind with lower affinity ($K_d$) with viral target protein (Spike protein) due to mutations in the epitope which generated the immunoglobulin (IgG). Lack of binding or lower affinity of binding can interfere with signal generation and failure to log signal over noise. Thus, individuals carrying SARS-CoV-2 may fail to test positive (false negative) if the viral variant possesses mutations preventing the antibody (A) to bind with the mutated Spike protein. Other factors (temperature, pH, salinity) may also interfere with signal (see “model scoring” in Appendix).

In (B), binding with the aptamer is highly specific but it depends on precisely which oligonucleotide (sequence of the ssDNA from an enriched pool) binds to which part of the Spike protein. For ADD, one aptamer may bind to the RBD (receptor binding domain) of the SARS-CoV-2 Spike protein. The length of the RBD (primary sequence) used in screening and enriching for the aptamer(s) may influence the shape (structure) of the RBD during selection phase. The complementarity of the shape of the RBD and the secondary/tertiary structure of the ssDNA complex is key to the binding specificity and affinity. If the test sample contains the whole Spike protein (includes RBD) as well as fragments (peptides with different lengths of amino acid sequence) which may or may not contain the RBD then the binding to the aptamer may fluctuate (widely) because the primary sequence of the protein may influence the secondary and tertiary structural outcome. The latter may change the configuration of the RBD in a given fragment and prevent binding to the aptamer, generating a false negative. If a sample contains other proteins and peptides, it is possible that the 3D configuration of an arbitrary protein or protein fragment could mimic or compete, albeit partially, with the RBD, and elicit a signal by binding with the aptamer, even if the binding is ephemeral due to reduced affinity (false positive result).

Binding of the Spike protein RBD to the immobilized hACE2 protein target (C) is probably the weakest link in this tripartite approach. Presence of mutations, dynamic or modified configuration and the effect of the environment (temperature, pH, salinity) may perturb binding and corrupt the signal.

Error correction and data curation may be necessary to prevent data corruption (false negative, false positive, limit of detection) to improve the information and recommendation for end-users. If the confidence in the raw data from each element is high, then the data may be responsibly combined (after data scoring, image risk score) to display the information with an assigned degree of confidence which may be more than the sum of the parts (positive, negative, false positive, false negative). The strategy from data acquisition and display vs information and recommendation must reduce risk, optimize level of precision and accuracy to maximize the value of the information for the user and/or the community. Of greater concern is the accumulation of errors, which when aggregated (time series data from ADD used as a surveillance tool), may generate spurious results with respect to the status of the population.
7. Array of Targets

The ADD approach for detection of infectious agents is based on targets identified from the biology and/or lifecycle of the organism and its interaction with the host (humans, animals). The RBD (receptor binding domain) of the Spike protein from SARS-CoV-2 and the human ACE2 cellular receptor (in bats, rats, pangolins and related animals in the phylogenetic tree; reviewed in reference 9) are under intense scrutiny. But, exploring the biology of SARS-CoV-2 reveals other equally potent targets. Developing drugs, antibodies and aptamers may benefit from a brief review of the viral biology. For SARS-CoV-2 detection alone, there are at least two other external proteins which may serve as targets for binding to aptamers, the M protein and the E protein in addition to S protein (Figure 4).

Figure 4: ssRNA genome of SARS-CoV-2 is longer compared to other RNA viruses (HIV, Influenza, Zika, Ebola; see Figure 16). It is encased in a nucleocapsid core (N protein120) and resides inside the virus. External surface of the virus is studded with S (spike), M (membrane) and E (envelope) proteins.

The receptor binding domain (RBD) of the Spike protein appears to make the first contact with the human cellular receptor ACE2 (angiotensin converting enzyme 2). Disrupting this event is the Holy Grail for preventing the virus from entering the cell. The mechanism by which Spike protein facilitates viral entry is not merely due to the recognition (between RBD and ACE2) but a cascade of events that begins after successful binding. The events that follow result in fusion of the viral envelope with the cell membrane, thereby allowing the viral genetic material (+ssRNA) to be delivered inside the cell in order to create progeny viruses. Fusion is mediated by the fusion machinery and fusion peptide sub-segments of Spike S2 protein which includes a step resembling a “jack-in-the-box” toy121. These segments of the Spike protein are better conserved and occupy a distinctly different part of the Spike protein (Figure 5).

14 • ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
Figure 5: Sequence conservation\textsuperscript{122} of sarbecovirus S glycoproteins plotted on the SARS-CoV-2 Spike protein structure [viewed from the side (A) and top (B)]. The receptor binding domain (RBD S1) is separate from the region of the Spike S2 protein necessary to initiate viral entry. The latter is better conserved (fusion machinery, fusion peptide) and perhaps better targets for ADD aptasensors.

The better conserved segment of the Spike protein may offer valuable epitopes\textsuperscript{123} and potential binding sites for aptamers (unless glycan moieties interfere). In addition to the RBD (which appears to be more variable), the conserved portions of the S2 subunit responsible for fusion (fusion machinery, fusion peptide) are likely targets for aptamer binding. It remains to be seen if reagents (monoclonal antibodies, aptamers) aimed at the fusion specific domain of the S protein can disrupt viral entry and serve as tools for detection as well as prevention.

Interfering\textsuperscript{124} with the human cellular proteins ACE2 and TMPRSS2 (which are viral targets) to prevent viral binding may not be prudent. Reagents directed against proteases, usually non-specific, may perturb physiological functions essential for homeostasis. The events which follow after the viral Spike protein docks with the human ACE2 protein are illustrated (Fig 6 copied from Scientific American\textsuperscript{125}).
Figure 6: Cascade of events leading to viral entry into host lung cell identifies the “jack-in-the-box” mechanism as a pivotal tool used by the fusion machinery of the Spike protein to deliver the viral RNA inside the host cell. Selectively disabling the fusion machinery of the Spike protein is an attractive target for aptamers and other reagents. If available, the latter may not only detect and diagnose but prevent infection, even if virus particles may have already reached the human apical surface area. Superior region of the lungs are more vulnerable to infection due to higher number of hACE2 receptors. The number of hACE2 decreases from superior to inferior. Lower part of the lungs have less ACE2 and TMPRSS2 proteins, corroborated by the observation that these genes are expressed at a higher level in upper nasal epithelial tissue compared with bronchial and small airway epithelial brushings.
The +ssRNA of SARS-CoV-2 (positive strand serves as mRNA) generates at least 27 or more viral proteins by creatively manipulating the host translational machinery. Theoretically, any or all viral proteins could serve as targets for anti-viral strategies. Virus-encoded proteases are distinct from cellular proteases and may serve as good targets. The viral protease 3-chymotrypsin-like protease or 3CLpro aka M pro is encoded by Nsp5 and appears to cleave (see “scissors” in Figure 7) essential viral proteins from “polyproteins” generated from translation of open reading frame (ORF) 1a and 1b (Fig 7). Papain-like protease PLpro (ORF 1a, Nsp3), cleaves proteinaceous post-translational (ref 131) modifications on host proteins to evade host anti-viral immune responses. Nsp1 suppresses host translation by cleaving cell mRNAs and competes with mRNAs for binding to human 40S ribosomal mRNA channel (as well as 43S, 80S subunits). Type 1 interferon response is modulated by Nsp1, Nsp 6 and Nsp13, which interferes indirectly with IFN-1 by suppressing the phosphorylation and/or nuclear translocation of other cellular molecules involved in catalyzing the IFN-1 response.

Figure 7: (Top) SARS-CoV-2 genome encodes nonstructural proteins (nsp), structural and accessory proteins. Nsp5 are encoded by ORF1a & ORF1b generating pp1a (nsp 1-11) or pp1ab (nsp 12-16). The structural and accessory proteins are synthesized by translation of their respective sub-genomic mRNAs. (Bottom) Translational repression (Kamitani et al) and binding to 40S ribosome (Thoms et al) by Nsp1.
Figure 8: The positive sense (+ss) RNA genome is translated by the host translation machinery to make polyproteins that are co-translationally cleaved by proteases (PLpro/Nsp3 and 3CLpro/Nsp5) encoded in the polyprotein to generate components of RdRp or RNA dependent RNA polymerase (Hartenian and Nandakumar et al). The RdRp complex uses the genome as a template to generate negative sense subgenome and genome length RNAs, which are in turn used as templates for synthesis of positive sense full length progeny genomes and subgenomic mRNAs. Each and/or any protein factor in this complex may be a target for anti-viral reagents, for example, aptamers, antibodies, small molecules and inhibitors.

The conundrum and complexity presented by an abundance of anti-viral targets, a variety of strategies and potentially many cell types susceptible to infection, adds to the pharmaceutical dilemma where the problems of bio-availability, cross-reactivity and toxicity may force a solution to extinction. Viral proteins are distinct but structural homologies and overlapping functional issues are non-trivial.

Figure 9: Identification of ACE2 receptors on many other cells (humans). The SARS-CoV-2 virus is not only a respiratory virus or results only in pneumonia. It is causing systemic diseases presenting a vast array of symptoms and acuity.
The medical chaos due to our lack of understanding of the biological minutiae of SARS-CoV-2 is not completely without a silver lining, albeit bleak. The ray of “hope” emanates from ExoN (Figure 8) the protein produced from Nsp14 segment of ORF1b (see Figure 7). It appears that SARS-CoV with inactivated ExoN is growth impaired and mutates at a much higher level (>20-fold\(^{149}\) higher, see right panel in Figure 10). SARS-CoV with one of the longest genomes (see Figure 16) among common RNA viruses (HIV, Influenza, Rhino, Ebola) abhors errors\(^{150}\) in replication (not corrected in other common RNA viruses with low fidelity RNA replication). High fidelity replication has enabled SARS-CoV to maximize its genome size (see Figure 16) using RNA-dependent proof reading system, repair and error correction implemented by Nsp14-ExoN (there are Nsp14 homologs in other viruses). Lack of error correction in humans\(^{151}\) may result in disease, dysfunction and death, even due to point mutations.

Figure 10: Mutated or inactivated Nsp14-ExoN results in >20-fold increase (Eckerle et al \(2010\)) in genomic errors (B, right panel). ExoN in RdRp of SARS-CoV-2 enables error correction (left panel).

Error correction in SARS-CoV-2 may have implications for optimizing target selection for anti-viral strategies. The choice of the receptor binding domain in subunit 1 (S1 RBD) of the SARS-CoV-2 Spike protein, therefore, may be incomplete as a target (Figure 3). It appears that the fusion machinery and the fusion peptide (FP) region of the Spike protein (subunit 2) is better conserved and will continue to remain better conserved due to the error correction mechanism (see Figure 10). Hence, sub-segments within subunit S2 of S protein may be better targets. The obvious caveat in this discussion is whether the chosen sub-segments in S2 may be sufficiently exposed or available to bind with the anti-viral molecules.
Figure 11: S protein (trimer) consists of 2 structurally noncovalently linked domains, S1, contains RBD (receptor binding domain) and S2 contains the fusion machinery and the **fusion peptide**\(^{152}\) (FP). Site of proteolytic cleavage → vertical arrow. S2 contains 2 HR (heptad repeat)\(^{153}\) regions HR1 (898–1005) and HR2 (1145–1184) connected by 22-amino acid linker (LVPRGSGGSGGGLVELFQGP). Hydrophobic residues (a and d positions in heptad repeat regions) are conserved. SS (N-terminal signal sequence), TM (transmembrane domain, C terminus), FP (fusion peptide, **bottom**\(^{154}\)), IBV (infectious bronchitis virus), FIPV (feline infectious peritonitis virus), MHV (murine hepatitis virus - murine coronavirus).
8. Discussion

Could we detect SARS-CoV-2 in saliva, prevent membrane fusion and block viral entry with the same aptamer or another type of molecule based on the better conserved S2 of the SARS-CoV-2 Spike protein? Could we detect SARS-CoV-2 in saliva of asymptomatic individuals without COVID-19? ADD may use better targets for its aptasensors beyond RBD S1 and hACE2. ADD can be accomplished, as suggested by the evidence from creation of DeMEA (but it uses high cost microfluidics).

Even if ADD is successfully engineered to be a low-cost biomedical device for non-invasive detection, dissemination of ADD and other systemic surveillance tools will still depend on community-specific economics of technology to facilitate diffusion and adoption. Bringing data and information together to make sense and extract foresight (uncertain of the value of hindsight) will be a challenge which new initiatives must address. Diffusion of the tool to vulnerable communities will be restricted unless the end-to-end system is cost-effective at a level where it is sustainable for repeated use, preferably daily, as a surveillance tool for humans, pets and farm animals.

Data when transformed into usable information may deliver value for the greater good, for the greatest number. ADD is one small surveillance tool but it isn’t enough. Healthcare cannot be a knee-jerk reaction to epidemics and pandemics. Continuous monitoring (even for high risk individuals) may remain a mirage in view of the disproportionate socio-economic imbalance. While we must ADD up to address the crisis at hand, we must also utilize this disaster as an opportunity to deploy profiling as a healthcare staple. Other tools, for example, wastewater analysis may offer transparency and guide public health strategies regarding elements the community must address, in advance, to prevent meltdown of health services. When an emergency presents itself we must not disintegrate into quagmire.

Precision medicine and precision public health may benefit if we probe the broader question of physiological status as expressed by proteins but further complicated by our microbiomes. Isolated snapshots of data may be rate-limiting for communities under economic constraints. But, convergence of data from ADD along with multiple levels of profiling (DNA, RNA, protein, RDW) as well as environmental and wastewater data, if included, may augment the value of information, which could be catalytic for medicine, in general, if aggregated and shared between open platforms.

Analytical skills necessary to deconstruct the data and reconstruct its meaning, relevant to the individual and/or the community, may pose a rather insurmountable barrier in terms of tools and/or human resources. The ill-informed inclination is to hastily pursue a “quick and dirty” version (perhaps shoddy, yet masquerading as good enough) without a long term view or a vision that embraces a sense of service, science for the good of society and access to global public goods for all. It goes without saying that one shoe does not fit all. It is obvious that ADD is not enough to better prepare for the future.
Figure 12: Serum from a significant percentage of patients (one third) recovering from COVID-19 have low viral neutralizing activity. Depending on the acuity of the infection, patients may or may not follow standard immune profile (top). Low variation (Fig 10) in SARS-CoV-2 Spike protein is good news but mutations, D614G (middle panel and bottom) may still complicate the immune response and expected anamnestic response to reinfection or use of classical approaches to vaccination.

If immunity from traditional vaccines are uncertain, can we supplement with alt-vaccines (which are non-immunogenic, for example, aptamers), to better prepare for low-cost and rapid response to public health during future epidemics / pandemics?
9. Complexity of the biomedical scenario, socio-economic catastrophe and the public health crisis

Since 1980’s the HIV epidemic has infected ~76 million people\textsuperscript{183} (~1% of the global population) and almost half are dead (~33 million AIDS related deaths, disease caused by HIV) and currently the other half is still living or struggling with the disease. Yet, the thrust for HIV vaccine pales compared to the warp speed vaccine development collaboration\textsuperscript{184} against SARS-CoV-2, which erupted in 2020 as the COVID-19 pandemic. Is it because SARS-CoV-2 is irreverent and indiscriminate in infecting humans?

Figure 13: The timeline of SARS-CoV-2 Vaccine Development\textsuperscript{184} (mRNA-1273 vaccine\textsuperscript{185}) to control COVID-19 (codeveloped by NIAID, NIH and Moderna, Cambridge, MA). The mRNA encodes the SARS-CoV-2 full-length spike glycoprotein trimer, S-2P (stabilized\textsuperscript{186} with two Proline\textsuperscript{187} substitutions at the top of the central helix in S2 subunit). mRNA is encapsulated in lipid nanoparticles (0.5 mg per mL) and diluted with normal saline to achieve the final target vaccine concentrations\textsuperscript{188}.

Figure 14: 7-day rolling average of new COVID-19 cases\textsuperscript{189} from January through September 21, 2020.
Coronaviruses have long co-existed with humans and animals. Error correction (Figure 10) has made the genome of the coronavirus one of the largest among viruses (Figure 16). What does it mean? Compared to diseases due to relatively unknown viruses, and despite the flu pandemic ~100 years ago, the coronavirus, in less than six months, has changed, perhaps permanently, global thinking, trends and technology. Tobacco Mosaic Virus (TMV) was discovered around 1890-1892 but after more than 100 years of virus discovery, we have just now acknowledged the threat to global health from viruses. Understanding the molecular basis of virulence is the single most important questions in basic biology which must be investigated by the best and brightest, if we ever expect to mitigate the risk from viruses.

Figure 15: A family of (corona) viruses with pandemic potential (courtesy of S. M. Gygli, NIAID)

Figure 16: Coronavirus pandemic wasn’t really expected according to at least one global expert. The coronavirus has the largest RNA genome. Is it just a coincidence or is there any bio-medical correlation?
While we still remain clueless about what constitutes virulence, the genome size does not offer any solace or solution. New evidence about Neuropilin-1 as a host factor which facilitates cell entry further thickens the plot. The deep quagmire about R0 and k for COVID-19 defies epidemiological models but prefers apocryphal Pareto principles which suggests that 80% of new infections may be caused by only 20% (or less - 10%) - see Figure 17 and reference 216 of the infected individuals. Are these individuals a high risk group due to inborn errors of immunity caused by mutations of genes involved in regulation of type I and type III interferons (IFN)? How many more genetic factors may underlie the differentiation between super-spreaders vs sub-spreaders for SARS-CoV-2? If the latter is true then how valuable is generalizing infection dynamics from communities as a prediction tool for overall public health, advance planning and use as early warning for cautionary preparation?

In future, genomic analysis may enlighten us if there are polymorphisms which may partially account for this differentiation. It may be worth digressing to note that some individuals may be more susceptible to leprosy, caused by Mycobacterium leprae. Genes associated with leprosy include HLA (human leukocyte antigen) proteins. Analysis of eleven HLA genes in 1155 Vietnamese individuals revealed 4 leprosy-associated independent amino acid variants [HLA-DRβ1 positions 57 (D) and 13 (F), HLA-B position 63 (E) and HLA-A position 19 (K)] which comprised 2 pairs of linked genes, with one set conferring susceptibility [HLA-DRβ1 and HLA-A] and one being protective.

The demographics of infection by SARS-CoV-2 may be due to genetic determinants and individual outcomes may be determined by our genes as well as epigenetic factors which may be mapped to biomarkers. At this point it is unclear whether the etiologic agent of this 2019 coronavirus pandemic should be referred to as SARS-CoV-2 where SARS imply severe acute respiratory syndrome.

Figure 17: What is COVID-19? Respiratory illness? Blood clotting disorder? Cardiovascular disease? Autoimmune disease? Opportunistic "killer" for (~10%) patients with severe COVID-19 pneumonia and high titers of autoantibodies against different types (type I IFN-α2 and IFN-ω) of interferons?
Figure 18: Hypothetical déjà vu graphene sensor (RIGHT) detects SARS-CoV-2 antigens. Can smartphone detection adapt to other sensors (LEFT) to detect SARS-CoV-2 in exhaled breath by mouth? Smartphone breathalyzer for malaria and marijuana (tetrahydrocannabinol) is close at hand. Can it serve as a global surveillance tool to bridge the chasm of inequity?

Figure 19: Detection of one copy RNA per µL (microL) from SARS-CoV-2 with mobile phone camera. Cas13a (C2c2) is complexed with a CRISPR RNA (crRNA) containing a programmable spacer sequence (red tube) to form a nuclease-inactive ribonucleoprotein complex (RNP). When the RNP binds to a complementary target RNA, it activates HEPN (higher eukaryotes and prokaryotes nucleotide-binding domain) motifs of Cas13a that then indiscriminately cleaves surrounding ssRNAs. Target RNA binding and subsequent Cas13 cleavage activity can therefore be detected with a fluorophore-quencher pair linked by an ssRNA, which will fluoresce after cleavage by active Cas13. Ott et al used the SARS-CoV-2 nucleocapsid (N) gene as the template (detection target) to create an array of crRNA spacer (red tube).
The socio-economic fall-out from the stochastic spread of infection and non-deterministic trends affecting certain countries, select groups (race, ethnicity) and underserved clusters, may be an example of “writing on the wall” we are slow to acknowledge. The cost of testing 100,000 individuals in the US approximate $6 million. If 30 million tests are performed weekly it would require an additional $75 billion and adding the cost of contact tracing might bring the total to approach $100 billion232.

The “writing” says that the successful NIAID-Moderna mRNA-1273 vaccine or any other safe and effective vaccine against SARS-CoV-2, when it may become available in 2021 or earlier, may still be out of reach for billions of people. CRISPR233-based tests may be promising234 in the future (see Fig 19). BinaxNOW $5 test235 is at hand but may not be feasible for daily use in communities under economic constraints. The case of Hepatitis-C236 is an example how even after nearly 50 years, anti-viral drugs are not within the buying power of billions of people. Success of vaccine is not equal to access to vaccine.

Death, destruction and the decay of civilization237 may continue and may continue to amplify in certain regions of the world, long after the pandemic. If the current pandemic is substantially contained by the end of 2021, then the aggregated loss from mortality, morbidity, mental health conditions, and direct economic losses in the US alone is conservatively estimated at $16 trillion238. The US economy is about a quarter of the global economy239, hence, extrapolation suggests that losses due to this pandemic may be an estimated $64 trillion, globally (about 80% of the global GDP240).

This mundane proposal is an elusive quest for an alternative path, albeit temporary and vastly incomplete, perhaps through the use of aptamers (or other variations based on oligonucleotides241) to partially bridge the chasm of inequity242 and cushion the blow from the mortality and morbidity, yet to be witnessed. Healthcare is a pillar (FEWSHE - food, energy, water, sanitation, healthcare, education) of life and living but it is prudent to avoid indulging in any illusion or delusion because neither aptamers nor vaccines or CRISPR tools, irrespective of their respective efficacies, are a panacea for the restoration of civilization, even if this pandemic subsides in a few years. The quintessential ingredients for public health and global rejuvenation are scientific credibility, color-blind magnanimity and ethical leadership.

Figure 20: Similar to HIV, SARS-CoV-2 Spike protein uses a N-glycan shield243 to thwart the host immune response (coating of N-glycans in cobalt blue, right). Mutations244, inborn errors of the immune system and other factors may make SARS-CoV-2 endemic245. Thus, it behooves us to explore other risk mitigation strategies. Anti-sense oligonucleotides, mini protein inhibitors246 and aptamers (this proposal) are alternatives.
10. Post-Pandemic Public Health and Healthcare: Broad Spectrum use of Sensors with Smartphones

Transaction cost\(^{247}\) of *humans-in-the-loop* detection or surveillance\(^{248}\) is often astronomical\(^{249}\) and the burden of cost\(^{250}\) for public health and healthcare systems is prohibitive even for affluent nations. For resource constrained communities, the mortality and morbidity due to lack of access to primary care must be reduced. Can we effectively combine the ecosystem of sensors, smartphones and data informed decision analytics to deliver usable information at the point of care or point of need, in near real-time?

In general, part of the solution may be found in remote sensing and imaging tools (oil and gas pipelines\(^{251}\), leaves\(^{252}\), tree\(^{253}\) canopy\(^{254,255}\), radiation\(^{256}\)). SEE or “sense everything everywhere” (paint-based computation\(^{257}\), sensors in fabrics\(^{258}\)) was a ‘touchy-feely’ mantra at the turn of the millennium buoyed by the principle of ubiquitous\(^{259}\) computing\(^{260}\) but stumbled in practice due to the cost\(^{261}\) of computation\(^{262}\). Vinton Cerf’s “I P on Everything”\(^{263}\) (I *pee* on everything) was the witty clarion call for embedding the IPv6\(^{264}\) standard\(^{265}\) in all things\(^{266}\) to enable “bit dribbling” between “digital” objects. These ideas were preceded by “tangible bits”\(^{267}\) from Hiroshi Ishi\(^{268}\) and the “atoms to bits” paradigm\(^{269}\) of “Internet 0” from Raffi Krikorian\(^{270}\) and Neil Gershenfeld\(^{271}\) followed by the origins\(^{272}\) of internet of things\(^{273}\) by Sanjay Sarma\(^{274,275}\) and others\(^{276}\). The borborygmi of radio frequency identification (RFID) and standardization of the electronic product code (EPC\(^{277}\)) shifted the thinking from stationary goods and products with static bar codes to dynamic digital objects which can be uniquely identified in any process or supply chain and tracked and traced digitally between any number of transactions, globally.

Project Oxygen\(^{278}\) offered extraordinary insight into the art of the possible\(^{279}\) and represented a consilience and confluence\(^{280}\) of ideas but it was cost-prohibitive for real world applications, circa 2000. With decreasing cost of computation\(^{281}\), memory\(^{282}\), data storage\(^{283}\) and transmission\(^{284}\), these streams, which were occasionally bubbling since Isaac Asimov’s\(^{285}\) *Sally*\(^{286}\) in 1953, turned into a raging river bursting its banks. The convergence of these tools with initial thoughts about the networked physical world\(^{287}\) were far more than the sum of the parts. It exploded to become the inescapable tsunami of IoT\(^{288}\) which has infected every domain. The anastomosis\(^{289}\) of IoT with cyberphysical systems\(^{290}\) (CPS\(^{291}\)) has penetrated almost every field from asteroids to zeolites and engulfed them within the new\(^{292}\) laissez-faire world of DIKW\(^{293}\) hierarchy. The mobile smartphone represents the grand conduit for the aggregated dissemination of distributed facets emanating from the DIKW pyramid. The mobile platform appears to be the global choice to access and implement all and any service which is possible, via the smartphone, in some form or the other, where the ubiquitous device serves as the platform for information\(^{294}\) arbitrage.

ADD is a recognition element and a tiny part of this landscape. ADD enables the sensor, data is captured, analyzed, communicated and visualized on a smartphone. It may detect infectious molecules relevant to SARS-CoV-2 in saliva and nasal swabs for early detection to prevent the spread of the virus. This principle is applicable to *any* infectious agent as well as any physiological molecule of interest (see Figure 21). The potential of developing a “breathalyzer” (identifying molecules in breath, see Figure 18) may make it easier to detect any molecule or molecules which are either volatile or lighter than air. ADD may be the “*killer-app*” that IoT\(^{295}\) was searching since its conception\(^{296}\) circa 1999.

---

\(^{28}\) ADD is in the MIT Library [https://dspace.mit.edu/handle/1721.1/128017](https://dspace.mit.edu/handle/1721.1/128017) and ChemRxiv [https://doi.org/10.26434/chemrxiv.13102877](https://doi.org/10.26434/chemrxiv.13102877)
Based on the idea of swappable, modular, flash drives, sensors-on-a-chip in the form factor of flashdrives are not hypothetical but frontrunners as potential tools for dealing with infectious diseases. Cameras, accelerometers (speed, movement), gyroscopes (tilt, orientation), magnetometers (compass), lidars (range, depth sensing from reflected laser signal), GPS and other “sensing” tools are increasingly “standard” with smartphones. These “detectors” makes it possible to use multiple mediums and phases for detection of signals from molecules, changes in dipole moments (electro-magnetic field) and perhaps even perturbation ambient electromagnetic waves (transmission and capture of reflected radio waves).

Figure 21: Billions of users in underserved geographies may access limited health services by using ubiquitous tools that does not require installation of new infrastructure and re-uses “mobile lifestyle” devices to partially bridge the scarcity of resources. Smartphones may be catalytic for delivery of service, remote monitoring and health surveillance, not restricted to infectious diseases but as physiological probes for health and homeostasis or detecting onset of disequilibrium (BNP, Brain Natriuretic Peptide).

Using information arbitrage to better contain the pandemic is the thrust of ADD. Expanding this principle as a routine for public health and healthcare, in general, is not a leap of imagination but natural progression. It bears reiteration that data informed decision analytics (DIDAS) must embrace sensor data plus smartphone (SDS) applications not as “pilot” projects but science in the service of society to catalyze the SENSIBLE system (SENSors and Information arbitrage via mobile system). The marriage of DIDAS with SDS in the affluent world is a matter of social acceptance of SENSIBLE but the penchant for profit-first and lack of leadership are holding us back. In the rest of the world the barrier to diffusion of life-saving tools are, albeit with exceptions, greed, pursuit of unethical profitability, cost or paucity of infrastructure (engineering, energy, telecommunications) and rampant inequity in social cohesion.

ADD is in the MIT Library [https://dspace.mit.edu/handle/1721.1/128017](https://dspace.mit.edu/handle/1721.1/128017) and ChemRxiv [https://doi.org/10.26434/chemrxiv.13102877](https://doi.org/10.26434/chemrxiv.13102877)
One milestone for smartphone-based health surveillance may be the non-invasive glucose monitoring system which the healthcare system failed to aggressively adopt despite significant advances. The chest-thumping about diabetes pandemic continues in parallel with avoidance of available solutions. Cholesterol monitoring using smartphones may be a preventative measure for adults at increased risk for a slew cardiovascular diseases, a few of which may not show symptoms.

Decades after the discovery of atrial natriuretic peptide (ANP, 1981), brain or B-type natriuretic peptide (BNP, 1987) and C-type natriuretic peptide (CNP, 1990), we still do not have SENSIBLE monitoring for BNP even though BNP sensors including an aptamer-based sensor for BNP-32 and cardiac Troponin I are available. These and other biomarkers (CRP, TNFα) are indicators of cardiovascular dysfunction including congestive heart failure (CHF) and state of the patient after myocardial infarction, in addition to other conditions. BNP and other markers are key to risk stratification, diagnosis, prognosis, disease monitoring, titration of therapy, and identification of therapeutic targets for cardiovascular disease. Brain Natriuretic Peptide concentrations >400 pg/mL and N-terminal (NT) pro-BNP >400-900 pg/mL (age related) are prognosticators of congestive heart failure. Analysis of 48,629 patients of acute decompensated heart failure found linear correlation between BNP levels and in-hospital mortality. Failure of BNP to decline during hospitalization predicts death or rapid re-hospitalization. However, BNP levels of 250 picograms per mL (pg/mL) or less during discharge predicts potential for survival. Accelerating availability of sensors and transforming innovations to SENSIBLE systems for prevention of cardiovascular disease should neither suffer from paralysis due to analysis nor asphyxiated by the rancour over margin of profitability.
Figure 23: Smartphone cameras, LEDs, LiDARs and a FEAST of signal transduction tools (optical, Raman spectroscopy, electrochemical impedance spectroscopy, surface plasmon resonance) are now available as data carriers. Cartoon (top) shows smartphone-based blood pressure and non-invasive blood glucose monitoring. The SENSIBLE system may be used to estimate blood cholesterol level, hemoglobin and uric acid as indicators of health, albeit imperfect. Data from smartphone based optic disc exam, photoplethysmography, electrocardiograph for arrhythmias (see bottom panels), general ECG, heart rate, respiratory rate (reflexion of radio waves), pulse oximetry and other vitals, collectively, may create precision physiology portfolios (open data source interoperability).
It is likely that *hundreds* of papers are published *weekly* on sensors, many of which may be adapted/adopted as a part of the SENSIBLE family. Biosensing using mobile devices at the point of use is a staple, for example, in the food industry (spoilage, contamination, security), soil, water, agriculture, manufacturing, chemical industry, transportation, energy, etc. But, lack of open data and restrictions on data interoperability makes the transformation of data to information quite difficult. Scientists are eager to drill deeper to develop yet another sophisticated sensor but real-world user are in quest of answers at the point of use, and does not have the luxury to deal with numbers, compilers and programmers.

This discussion about the cacophony of available sensors is an embarrassment of riches from decades of research and development scattered as parts, in silos or locked by patents. The form and functional orchestration and integration necessary for sensors to contribute to precision physiology requires cross-pollination of ideas. Multi-disciplinary teams are necessary to create end-to-end working SENSIBLE systems which can be synchronized, if authorized, as a part of the public health information system. If that data is shared in real time, it may reduce mortality, morbidity, cost to society, decrease the burden on emergency medical professionals, and actually aid in preventing dysfunction. If this data is anonimized to serve molecular epidemiology, it may help precision public health and channel benefits to the community by revealing the environmental conditions or instances which need additional attention.

A plethora of brilliant experts with deep knowledge can fill any university hall but few have the breadth of ideas which, if synthesized, synergized and integrated, may helps to address or even solve a real problem in the public domain where non-experts are the end users. Solutions based approach must combine depth with breadth to deliver the fruits of science to society as global public goods. The latter may be missing in the academic context where chronic search for scholastic erudition is the norm. The concept of essential products and services as global public goods may not be appetizing in the corporate context due to their perpetual penchant to promote profit and profiteering, first. The cleavage between purpose and profit needs a new bridge and a new breed of thinkers and leaders with altruistic traits.

The laser-focus of biomedical professionals on saving the lives of those affected by COVID-19 and the public health community on preventing the spread of infection by SARS-CoV-2 is the only path, at present, to lift us out of the quagmire of the raging pandemic. Yet it may be crucial to use this disaster as a global opportunity to strengthen public private partnerships (academic-industry-government) for the ubiquitous deployment of global tools for early detection and prevention, not only for pandemics, but for public health and healthcare, in general. It is an enormous task and requires global leadership.

Ubiquity of smartphones is the available SENSIBLE platform to create at least one bridge over the chasm separating the *haves* from *have nots*. Inextricably linked economies of the under-developed, developing and developed nations makes it imperative that the leadership for global public health must be agnostic of prejudice. Trans-disciplinary cooperation and collaboration between corporations must rise above conventional economics, narcissism, egocentricity and personal wealth creation. We need an overwhelming force for good, for a greater purpose, for the greatest number (of people).
11. Comments

The complexity of the physiological dysfunctions associated with the SARS-CoV-2 pandemic calls for multiple routes for detection, prevention and therapy. ADD is a tool for detection and perhaps one of many other alternative treatment (nanobodies\textsuperscript{342}) that does not depend on the immune system. Immune response may be heterogeneous\textsuperscript{343} based on individual genetic constitution (inborn errors) as well as natural decay\textsuperscript{345} of antibodies and antigenic evolution\textsuperscript{346} despite error correction in coronaviruses (see Figure 10, antigenic drift\textsuperscript{347} may not be exclusively due to genomic mutations during replication).

The global diffusion of smartphones as ubiquitous devices is an opportunity for digital\textsuperscript{348} public health to accelerate the use of smartphones for detection of any infectious agent. ADD is one tiny sub-component of the proposed SENSIBLE system. Realization of the SENSIBLE system is a difficult task. The deployment of the SENSIBLE system, with ADD and other sensor components, is an essential task, unless the veneer of health equity\textsuperscript{349} is only that, that is talk, or in other words lippenbekenntnis\textsuperscript{350}.

The caveat in this line of thinking is the over-emphasis on the sensor and the SENSIBLE system as if the value proposition is undeniable. The latter is true for the affluent economies of the world but detection without follow-up is an exercise in futility. The latter is common in communities under economic constraints where tools to detect (SENSIBLE system) are impotent because there are very few resources to attend to the public health or healthcare need identified by the SENSIBLE system. Just because the user can detect the presence of mercury in drinking water does not mean that the user has another alternative source of drinking water in under-served or dystopian communities. Is it more or less psychologically debilitating to drink water or consume food if the user is cognizant that the water is contaminated with mercury\textsuperscript{351} or the food is laced with bacteria\textsuperscript{352} beyond the level of food safety?

Incongruity between the pursuit of science, implementation of the fruits of science and science as a measurable service to society is a conundrum beyond the horizon of tools and technology, for example, the SENSIBLE system. Entrepreneurial innovation can create SENSIBLE but implementing SENSIBLE requires leadership imbued with a sense of the future, especially for low-income countries with ultra-low per capita disposable income. In the absence of charity, SENSIBLE for public health and healthcare must carry with it a pay-per-use price tag which may be a micro-payment or even a nano-payment but still it must be a non-zero payment for the system to be sustainable and survive long enough to deliver value.

Ephemeral gimmicks demonstrating SENSIBLE in geographies with GDP which may be less than an average household income in Europe is a deliberate act to deceive, dressed up in a marketing\textsuperscript{353} garb by the glib, the smug and the smarmy. Enabling a SENSIBLE future and making it sustainable for most segments of the community, in greatest numbers, is a mission for a visionary leader who radiates the aura that kindness is a strength, not a weakness, humility is a virtue, not a lack of knowledge, that agreeing to disagree is a mark of civility\textsuperscript{354} and dignity, not a character flaw, that fear is not a tool for maiming diversity, speech or peace, that progress of civilization is development of the freedom\textsuperscript{355} to act on remediable injustices\textsuperscript{356} and lift many boats, not a few yachts. The best man for that job is a woman.
APPENDIX
APPENDIX I - Brief Description of the Components\textsuperscript{357} for SARS-CoV-2 ADD Decision System (Fig 1)

**Message Broker**
When users upload images (the data after scanning with the HoloLens app or equivalent mobile tool), the mobile application (on their phones) writes messages with the image content and other metadata to a message-broker, which may be cloud-based message queuing (MQ\textsuperscript{358}) protocol (open source software). The message broker allows devices to quickly offload data and confirm “sent” to a user (if cloud based), thereby decoupling the user experience from the data store (even if it uses a temporary tinyDB on the device, if the network is unavailable to access the cloud in real-time at the point of use). Messages can be queued in topics and the system may enable autoscaling (as usage of the application increases, more users can be provisioned, process user uploads and get them stored). The uploads (data) are also sent by the message broker to the feature extractor and long-term storage database (may use the batch upload option when device is proximal to a high bandwidth gateway which can offer access to cloud services).

**Operational Data Store**
The message broker transfers uploads to ODS (Operational Data Store\textsuperscript{359}), which may be a cloud-based managed service or part of the tinyDB on the device, if cloud is inaccessible at the point of use. ODS must be able to store image data (supports binary blog column type) alongside time-index numerical and character data. It is intended to only serve “hot” (nascent) data to the application. Older data may be evicted (batch uploaded to cloud managed facilities) to optimize on-device service and prevent data amplification. ODS is tuned for fast random reads and serves requests made by mobile app when users view recent uploads and additional metadata about those uploads, including “risk scores”. ODS is optimized for fast writes and high efficiency time-series queries.

\textsuperscript{35} ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
**Feature Extractor**

Extracts additional metadata from images/data uploaded from the mobile app (uploads it to the long-term system of record\(^{36}\) which includes raw data uploaded from the application, similar to “master data” in ERP\(^{361}\)). Feature Extractor may convert the uploaded image into a numeric matrix\(^{362}\) or create hash table or representation of a region\(^{363}\) and correct for differences in resolution (for example, variation due to pixel density of cameras on different smartphones). Feature\(^{364}\) vectors\(^{365}\) may be maintained in the long-term system of record. It may be written to the operational data store to enable extraction/selection\(^{366}\) of incoming data (uploads from message broker) relevant to these feature vectors.

**Long-term System of Record**

Mobile applications may never access data directly from this data store\(^{367}\). Interactive-speed queries to this data store may not be supported. When necessary, objects stored in this “record” may be extracted and the data is loaded into an analytical data store. For object stores, this operation may be accomplished using query-over-files engines\(^{368}\). The thorniest problem that ferments within long-term data record is the inaccuracy of “accurate” data and the diabolical mayhem from “big data” if it is sourced and stored.

**Analytical Data Store**

Scientists and data experts will need historical data (from uploaded samples) to train task-specific\(^{369}\) machine learning (ML) models to assign risk scores to samples. Analytical data store (ADS database\(^{370}\)) may be populated with data from the long-term system of record using scheduled batch data uploads.

**Model Training**

In model training\(^{371}\), a statistical model is built from historical data. Models should be serializable\(^{372}\) representations of the program generated by ML training. Serialization is essential for interoperability on different platforms. It is key to create composable models where models from different groups can be deconstructed to sub-elements which can be reconstructed to compose a new model (which may be greater than the sum of parts). Serialization enables the process of translating a data structure or object state into a format that can be stored or transmitted and reconstructed. Proprietary software vendors obfuscate or encrypt serialized data to prevent access. Standard architectures such as CORBA\(^{373}\) define the serialization formats in detail to enable open access.

**Model Scoring**

In model scoring, a model is called on input data, the model processes the input data and generates a prediction. The structure of this code depends on the design choices made during model training. For reliability of deployment, model scoring may run in a container\(^{374}\) (an unit of software) which contains code (and all its dependencies) that uses a model to produce predictions on new input data. If model scoring runs in a container then the model can be arbitrary code in the developer’s language\(^{375}\) of choice. Model scoring requires features created previously by the feature extractor (feature selection is critical).
APPENDIX II – THE PROPOSED “SENSIBLE” SYSTEM AND APTAMER-AS-A-DRUG (AAAD)

There is nothing new in proposing that we apply a patent-free 30 year old idea (aptamers) to a nascent problem (SARS-CoV-2). The expectation is that a credible scientific investigation may reveal if aptamers may indeed serve as an alternate to conventional wisdom (immune response). The scientific strength of this idea is based on a rigorous tenet of molecular biology which has repeatedly demonstrated that proteins bind to nucleic acids, as a fundamental mechanism of action in biological regulation. If this approach succeeds, it may help and save lives in the less affluent nations (80% of the global population). The R&D pillars in the process are as follows (a few may be pursued in parallel):

[1] (Pharmaco)Dynamics of DNA aptamers that bind with efficiency (target access?), specificity and reproducible (quantitative) affinity. Which SARS-CoV-2 proteins are targets? (Pharmaco)Kinetics of binding, in terms of equilibrium dissociation constant \((K_D)\), where a smaller \(K_D\) (the ratio \(k_{OFF} / k_{ON}\)) indicates greater binding affinity of the ligand for its target (SARS-CoV-2 target protein), between aptamers and preferred target vs “nearest neighbor” protein competitor, based on shape and amino acid sequence (larger \(K_D\) value indicate weaker binding), must show reproducible and statistically significant difference of at least one order of magnitude. The design of the binding assay remains to be determined but must include kinetic data using three different principles (surface plasmon resonance\(^{376}\), biolayer interferometry\(^{377}\) and electrochemical impedance spectroscopy\(^{378}\)) to enable incisive data analysis.

\[
K_d = \frac{[A][B]}{[AB]} \quad K_a = \frac{[AB]}{[A][B]}
\]

Rigorous determination of dissociation constant and association constant (reciprocal of dissociation constant) is the bedrock of biochemistry (nauseatingly quantitative, repeatedly reproducible, criticism from real experts) at the heart of chemical equilibrium\(^{379}\) which is the essential pharmacokinetic pillar to determine which aptamers and protein targets may be potentially useful for which purpose. It remains to be explored if lectins\(^{380}\) and glycan specificity (Figure 20) must inform our data outcome. Will including lectin sensors\(^{381}\) (parallel positive control) enhance data accuracy (reduce false negatives) if Spike glycan shield adds to data uncertainty by (occasionally) perturbing epitope specific binding with aptamers?

[2] Test DNA aptamer binding to sensor material (laser inscribed graphene, Au/Pt nano-materials, etc). Establish metrics for stability of binding, using linkers to attach aptamers (covalent), using lectins and/or chemical tails, eg, poly(N-isopropylacrylamide).


[4] Repeat [3] with aptamers linked/adsorbed on sensor material surface to choose the best outcome from materials and transduction. Is the signal vs noise quantitative or qualitative? Prefer quantitative because qualitative offers only “yes/no” (with reservations) and influenced by limit of detection (LoD).
[5] Combine outcomes. If signal over noise is statistically significant ($P < 0.001$) after data acquisition by a mobile platform, then we have accomplished the scientific rigor to fuel the engineering basis for creating tools and applications for detection/diagnostics/screening/surveillance. The latter may involve innovation in engineering design to determine form factors, product development (breathalyzer) and imagination to transform the idea of USB connected modular, mobile, adaptable, sensor-on-a-chip to link to smartphones (any USB port) to create the (hypothetical) surveillance tool: molecularphone.


Can aptamers serve as alternates or supplements to traditional vaccines? Small molecule-like "inhibition" by aptamers (in vivo) may offer low-cost (?) therapeutic paths for less affluent nations. However, the general mechanism of action of small molecules versus aptamers may be quite different. Aptamer binding to a specific region of a target protein may induce some changes, perhaps a change in conformation, but it may not disable the ‘active site’ of the target protein (if it has enzymatic functions) or dissuade the protein from its usual activities even if it suffices to reduce its efficiency and/or efficacy.

Figure 24: Natural and synthetic organic small molecules are important pharmaceuticals (top panel). SARS-CoV-2 Nsp1 (Thoms et al, see Fig 7) acts as a small molecule to “fit” in the “groove” of the 40S ribosome (bottom, left) to arrest host translational systems, as shown in the models (bottom, right).
Figure 25: RdRp (left panel) may be the primary target for aptamer-as-a-drug (AAAD). Remdesivir (middle panel) is ineffective for SARS-CoV-2 RdRp (other small molecules are being tested). Additional targets (bottom panel) may include proteases and the individual proteins of the RdRp (RNA-dependent RNA polymerase) complex. Source: Bradner, James (2020).
Uncompromising rigor of (pharmaco)dynamics and (pharmaco)kinetics [step 1] tested under a spectrum of conditions (pH, salinity, salivary enzymes, blood, plasma) will determine if AAAD is even a possibility. A positive outcome (small number of aptamers binding select viral proteins with stability, epitope specificity and selectivity in a variety of body fluids, pH range) may be the first indication that it may be worthwhile to test AAAD candidates for in vivo activity (assay to test strength of inhibition?).

Initial interactions in buffer and laboratory conditions are good indicators but it cannot predict what will happen in vivo because aptamer-protein binding is determined by the 3D shape (secondary structure) the aptamer (single stranded DNA string) will assume (under a set of conditions) and protein (epitope) binding, as a consequence of that structure (shape). To partially mimic in vivo conditions, testing aptamers in cell culture may be the ‘quick and dirty’ first choice. The dynamic conformation of aptamers under various conditions influences the binding to target proteins. It is a source of uncertainty.

Extrapolating results from in vitro tissue culture and in vivo animal models (next mandatory step) to humans is neither prudent nor a bona fide scientific process. But step-wise success may help to justify the path forward. Any one of many factors could be the nail on the coffin of the AAAD idea. These factors include but may not be limited to stability, bio-availability, delivery, absorption, permeability, distribution, metabolism, elimination, cross-reactivity, general cytotoxicity, organ specific toxicity (for example, cardio toxicity, renal toxicity, neurotoxicity, blood-brain barrier). Finally, only unequivocal success in most stringent human clinical trials may help to transform the idea of AAAD into reality.

Figure 26: AAAD could use lipid nanoparticle delivery as in the Moderna mRNA-1273 (1273 amino acid SARS-CoV-2 Spike protein) vaccine protocol\(^3\) (LNP formulation by Langer et al). Delivery of AAAD using LNP may improve absorption, for example, if used as a topical nasal spray to prevent spread of infection. All mRNAs in human cells are encoded by 2% of the entire genome. 98% of the genome is transcribed into cellular RNAs whose activities still remain to be discovered (new RNA therapies?). Thus, most of the human genome does not code for proteins, most of the disease-causing genetic variants are located in non-coding regions, most transcripts produced by genes are non-coding and non-coding genes may not have a functional annotation (their biological role is unknown). But, clues about the biological role of long non-coding RNAs in the human genome are beginning to emerge\(^5\).
APPENDIX III – THE MOLECULAR BASIS OF DISEASE

Variability in Immune Response to SARS-CoV-2 Infection: Conundrum, Conjecture, Common Themes

The spectrum of immunological diversity\textsuperscript{386} presented by CoVID-19 reinforces the value of thinking broadly\textsuperscript{387} and thinking differently. It may not be unwise to forward hypothesis or conjectures which may or may not provide clues to understand or unravel the biological basis of this conundrum. On the other hand, based on common themes in molecular biology and genetics, perhaps what we are observing is not a conundrum at all. This discussion brings together what we think we may know.

Let us commence with the observation that bacteria belonging to even one strain, for example, *Escherichia coli* (O104:H4, O157:H7, O121) if sequenced (DNA genome), will reveal that their genomes, in terms of DNA sequence are not exactly identical. One explanation based on the molecular biology of CRISPR (clustered regularly interspaced short palindromic repeats) indicates acquisition of new spacer sequences\textsuperscript{388} from foreign DNA necessary to adapt CRISPR-Cas\textsuperscript{389} system to confer adaptive immunity. The human genome\textsuperscript{390} revealed our genomes\textsuperscript{391} are similar but not identical (even between twins), due to unequally distributed single nucleotide polymorphisms (SNPs) in coding and non-coding sequences.

Figure 27: Acquired spacer sequences in bacteria is one reason why genomic sequences differ even within the same strain. What is the impact of the integration and the heterogeneity of the type and number of spacers on bacterial gene expression, protein expression, physiology and metabolism?
The conundrum about the variation in human immune response to CoVID-19 may be natural if we consider that each human is genetically unique. We have known for 2 decades that approx. 3 million nucleotides (including about 2 million SNPs) are different between our genomes (why we are genetically unique) which contain ~6.4 billion base pairs\(^\text{392}\) (6.4 billion nucleotides - A, T, G, C – in diploid human genome or 3.2 billion base pairs, monoploid). These differences are significant when evaluating drugs. Hence, genetic\(^\text{393}\) stratification of humans in clinical trials is now routine. The efficacy, effectiveness or resistance\(^\text{394}\) of the same drug may be quite different between individuals. The latter partially explains the observed variation\(^\text{395}\) in immune\(^\text{396}\) responses\(^\text{397}\) to CoVID-19. SARS-CoV-2 induces a multi-factorial\(^\text{398}\) physiological\(^\text{399}\) cascade of events\(^\text{400}\) involving systems\(^\text{401}\) and network of factors\(^\text{402}\) linked to genetic predispositions and co-morbidities which may influence phenotypic expression, to different degrees, directly or indirectly, in each human. The 823 epitopes\(^\text{403}\) mapped in the SARS-CoV-2 proteome, were not all equally recognized by antibodies in all individuals, indicating the complexity of stratification.

It is not only genomics but regulation of gene\(^\text{404}\) expression\(^\text{405}\) (transcriptomics), proteomics and metabolism (rates of anabolism and catabolism, metabolomics). Omics may be affected by epigenetic factors (food, air, water, environment\(^\text{406}\)) and immune cell dynamics are modulated by microbiomes\(^\text{407}\) (viromes). Taken together, these factors are likely to affect detectable symptoms and clinical outcomes.

In humans, multiple processes and DNA sequences flanking the immunoglobulin genes (V, D, J) influences the genetic rearrangement of the gene segments followed by somatic hypermutation\(^\text{408}\) which contributes to the great diversity of our immunoglobulin repertoire. It is one of the key tools available to the immune system to design antibodies and respond appropriately upon presentation of an antigen. Somatic mutations vastly differentiate and enhances the scope of response which may be otherwise restricted if the system were to depend only on the inherited genetic components (germ line theory\(^\text{409}\)).

The machinery available to antibody-producing cells for executing somatic changes in genes and gene expression is an evolutionary process. Creative application of this machinery may generate quite a variation in phenotypic response in CoVID-19. Somatic reshuffling in combination with differences between SNPs may result in an inordinate number of different permutations and combinations. Hence, the spectrum of CoVID-19 symptoms. SARS-CoV-2 proteins\(^\text{410}\) may induce somatic hypermutation in cells and tissues to result in perturbation of homeostasis. Epigenetic modifications triggered by sequence cassettes\(^\text{411}\) may affect basic processes (transcription, translation, post-translational modifications, etc.).

**Figure 28:** Discovery\(^\text{412}\) of transposons in maize\(^\text{413}\) revealed that segments of genes “jump” from one genome to another (see left, kernel colors\(^\text{414}\)). Variations of this “dynamic” concept are found in influenza\(^\text{415}\), Trypanosomes\(^\text{416}\), Plasmodium\(^\text{417}\) and other organisms. SARS-CoV-2 may hijack this mechanism, create *ad hoc* changes and alter therapeutic targets.
Figure 29: Variations in antibodies and binding affinities are due to VDJ (upper, left) recombination (combinatorial diversity $\sim 2.5 \times 10^6$ & junctional diversity $>10^{14}$) and somatic mutations (Tonegawa et al, 1974) mainly in B cells ($\sim 10^{11}$ in humans). Understanding recurrent features of antibodies binding to SARS-CoV-2 Spike protein (lower, left) helps to identify antibodies with therapeutic potential. Four classes of human neutralizing monoclonal antibodies are shown in the cartoon (lower, right panel).
Molecular dissection of the SARS-CoV-2 genome to delineate the functional role of each viral protein is fundamental. Testing infectivity of single gene knockouts (and multi-gene combinations) may be one essential step. Physiological exploration of the >332 human proteins (Gordon et al., 2020) which may interact with SARS-CoV-2 proteins is crucial. A global collaboration may be necessary to analyze the value of each nucleotide in the ~30KB single stranded RNA genome of SARS-CoV-2. The analytical rigor of this approach may mimic the incisive minutiae of mutagenesis as applied to beta lactamase. Thoughtful design and detailed execution of this megaproject may provide clues to what constitutes virulence. Metrics of virulence is pivotal to deconstructing its cryptic complexity and reconstructing the role of molecular medicine in healthcare for humans and animals infected with virulent agents.

In the interim, undergraduates in molecular biology may undertake the theoretical analysis of SARS-CoV-2 proteins. For each known viral protein, it may be useful to list expected modifications in primary amino acid sequence (if any) due to changes in the 3rd position of the RNA codon (Fig 26, L). For example, if AGU mutates to AGC, the amino acid Serine is still the same (silent mutation) due to the degeneracy of the triplet genetic code. If AGU/AGC mutates to AGA/AGG then Serine is replaced by Arginine. Using the Ramachandra Plot (Fig 26, R) students may explore which values of the $\psi$ and $\phi$ angles are possible for that amino acid residue which changed in the viral protein. Can the change in the codon create an amino acid substitution which can influence the conformation of the viral protein? Structure and function are inextricably integrated in biological activity. This exercise may uncover targets for experimental analysis and predict which changes in the codon and primary sequence, may be of consequence with respect to interaction between viral proteins and their putative human targets.

![Figure 30: RNA Codon table (L). Ramachandran Plot of allowed values of the $\psi$ and $\phi$ angles (R).](image-url)
The diversity and acuity of symptoms suggests SARS-CoV-2 proteins may have access to genetic circuitry\textsuperscript{429} of developmental\textsuperscript{430} clocks\textsuperscript{431} and may be “playing” with master\textsuperscript{432} switches\textsuperscript{433} or re-wiring\textsuperscript{434} circuits or “time\textsuperscript{435} spoofing” the expression in some form\textsuperscript{436} to modulate\textsuperscript{437} differentiation. It is possible that developmentally re-programmed genetic circuits or regressively differentiated cells may express proteins and/or other molecules which are not in our ‘data dictionary’ because it is not a part of our differentiated physiology. Therapeutic targets and approved drugs\textsuperscript{438} may not be effective\textsuperscript{439} because the virus may be creating their own decoy\textsuperscript{440} molecules to wreak havoc. The virus appears to be changing our targets to escape host defense (immune system) and offense (drugs) as well as perturbing cellular signals for biomarkers associated with CoVID-19 mortality (IFN-α\textsuperscript{441}, IL-18\textsuperscript{442}, IL-10\textsuperscript{443}).

Figure 31: Immune evasion strategies\textsuperscript{444} by coronaviruses include antagonization/disruption of various pathogen recognition receptors, TLRs\textsuperscript{445} (TLR3, TLR4, TLR7, TLR8; blue) and RLRs\textsuperscript{446} (RIG-I, MDA5; purple), transcription factors nuclear factor kappaB (NF-κB\textsuperscript{447}) and interferon regulatory factors 3 and 7 (IRF3, IRF7) which are (normally) supposed to stimulate the production of pro-inflammatory cytokines and type I and III interferons (IFNs), respectively. IFNs (autocrine and paracrine secretion) induce expression of interferon-stimulated genes (ISGs) via the JAKSTAT signaling pathway.
Figure 32: Promise of interferon and its role as a potential therapeutic agent. A reason to believe? Hypothetical Interferon Therapy illustration provided by Kizzmekia Corbett, NIH.

Figure 33: With >50 million infected by SARS-CoV-2 (actual number may be >500 million) the clinical deluge to deal with CoVID-19 patients may leave little time to pursue molecular stratification. Access to vaccines, neutralizing monoclonal antibodies and promise of interferon therapy (Figure 32) could make this discussion moot. But, currently the best we can expect is cluster treatment (top, right). If optimism begins to fade, then research must go beyond the hierarchical model (bottom, right) to understand the molecular basis of disease.
Investing in basic science to probe the molecular basis of virulence may be complex and tedious. The answers may not inform us sufficiently. It may leave room for doubt but it may also create room for consilience and build extensions for imagination and innovation. Science, engineering and economics may instruct us to “build back better” the predictive compass we may need for our tryst with destiny. Whether we can reach that fateful destination or not may be shaped, in part, by the plight of our ability or inability to reduce the incredible inequity which surrounds access to health science and healthcare.

To improve our (global one health) preparation for future epidemics and pandemics we need metrics (quantitative comparisons) to better grasp the variations in molecular structure and function associated with physiological dysfunctions, degree of virulence with respect to infections, rate of replication of infectious agents and factors affecting mortality. Genetic perturbation screens and GWAS (genome-wide association studies) are already generating molecules of interest for further analysis.

These are tasks for dedicated bench scientists who may toil for long hours to contribute even an infinitesimal iota of data to inform our understanding. Science in the service of society is a purveyor for the progress of civilization. Credible advances in science may not be achieved by flaky, fluffy and fuzzy flash of publicity, chicanery and malarkey.

Figure 34: Let’s do the numbers: can we explain this incredible difference? True, false or artifact?
APPENDIX IV – APTAMERS as ADJUVANTS and/or PARALLEL ALTERNATIVE to ANTIBODIES

Global public health and healthcare needs immune-agnostic and temperature-agnostic low-cost therapy

Dedicated bench scientists from diverse fields in molecular biology and biochemical genetics worked for years before the successful \textit{in vitro} transfection\footnote{\textsuperscript{463}} of nucleic acids (DNA, RNA) into human cells in tissue culture. It was an even greater ordeal to establish that \textit{in vivo} only those messenger RNAs (synthetic mRNAs) with a modified nucleoside\footnote{\textsuperscript{464}} may avoid immune recognition. Free RNA in higher animals rapidly activate host immune response. SARS-CoV-2 replicates its RNA inside coated vesicles\footnote{\textsuperscript{465}} in infected cells to evade immune surveillance. The promise of synthetic mRNA immunotherapies\footnote{\textsuperscript{466}} and its success as a CoVID-19 vaccine displays excellence in intellectual pursuit, plight and perseverance.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure35.png}
\caption{Synthetic messenger RNA encodes the target antigen which is packaged in a lipid nanoparticle and injected into individuals. Following uptake by cells, the mRNA\footnote{\textsuperscript{467}} is translated in the cytosol\footnote{\textsuperscript{468}}. The foreign protein may be released (induces B cell immune responses) and/or the antigen-derived epitopes are presented on the cell surface by major histocompatibility complex proteins (MHC) class I and II.}
\end{figure}
The title of this sub-section springs from the chasm between invention and implementation. The latter is a combination of resources and logistics. Implementation depends on cost of goods, production, facilities, distribution, personnel to facilitate access and administration. mRNA vaccine is a milestone for science but is it accessible, globally? Are aptamers affordable as a global alternative or wishful thinking?

One could have left this discussion at the doorstep of social disequilibrium if not for recent developments. Emerging reports suggests that the virus seems quite eager to take a bite out of the vaccine strategy. The usual suspect is antigenic drift, which has hindered the creation of a successful vaccine against the influenza virus, despite the devastation due to the 1918 global flu pandemic.

This usual (antigenic drift) occurrence may be unusual in case of SARS-CoV-2 (and coronavirus family) because error correction in coronaviruses (during viral RNA replication) should, biologically, reduce the frequency of mutations which contributes to antigenic drift (see Figure 10). However, antigenic and epitope evolution is not solely due to antigenic drift (mutations). Assuming the latter still holds true for recent SARS-CoV-2 mutants, then, once again, we may have stepped into yet another unknown abyss in our effort to understand the molecular evolution of antigenic drift in SARS-CoV-2 and implications for future vaccines.

If this paradox (antigenic drift despite error correction) is a sign of what may eventually become a paradigm for SARS-CoV-n then the mRNA vaccine strategy may be forced to continually invent and reinvent itself to keep up with the antigen/epitope evolution if it chooses to serve as an effective vaccine. It is a mimicry of “flu vaccines” which are of questionable efficacy.

SARS-CoV-n may evolve as the aetiologic agent of future CoVID-yyyy and start trending in the annals of the 21st century medical anthropology, with usual public health and wellness concerns. But, if the mortality and morbidity from SARS-CoV-n/CoVID-yyyy approaches SARS-CoV-2/CoVID-2019, then the economics of re-inventing mRNA vaccine cycles (for antigenic drift) and vaccination, may be prohibitive even for the general population in most affluent nations. Pursuing multiple targets for vaccines is prudent. Including aptamers as an alternate may be even wiser.

Despite its existence for about thirty years, patent-free classic aptamers are still uninfluential in molecular medicine except for one FDA-approved drug. Aptamers in molecular diagnostics including proteomic profiling (patented modified aptamers) may not have reached beyond R&D domains, yet. This discussion about classic aptamers for non-invasive rapid detection and prevention (as an “agile” drug which may be configured and re-configured/re-composed with relative ease compared to syn-mRNA vaccines) builds on one success, the RNA aptamer based FDA-approved drug Pegaptanib (Macugen) as a treatment for choroidal neovascularization associated with age-related macular degeneration. However, progress in the use of aptamers are not in short supply. But, it should be no surprise if critics choose to label the advocacy of aptamers as a case of chacun voit midi à sa porte.
The short-term rebuttal to immune-agnostic suggestions (including the current advocacy for aptamers) is the fact that epitope evolution due to antigenic drift in SARS-CoV-2 variants may not affect the molecular strategy adopted by the synthetic mRNA vaccine directed toward the SARS-CoV-2 Spike protein (including the receptor binding domain, RBD). Epidemiology of S protein variants supports\(^{481}\) this short-term view. The long term risk mitigating strategy calls for alternative therapies.

Why are we observing an increasing number of variants in a virus family supposedly equipped with error correction\(^{482}\) system? Is the replication coupled error correction\(^{483}\) erroneous? Inaccuracies during replication by the RNA-dependent RNA polymerases (RdRp, nsp12, Fig 8) are error-corrected by nsp14, a bifunctional enzyme possessing RNA cap guanine N7-methyltransferase (MTase) and 3'-5' exoribonuclease (ExoN) activities (proofreading function). Mutated\(^{484}\) or inactivated nsp14 may be lethal for viral replication\(^{485}\). Error correction is the reason for RNA genome size expansion (Figure 16). Nsp14 mediated excision of erroneous mutagenic nucleotides inserted by nsp12 is also the molecular basis\(^{486}\) for potential nucleoside drug resistance exhibited by SARS-CoV-2 family of pathogens\(^{487}\).

With accumulating genome wide errors\(^{488}\) in coding and non-coding sequences of emerging\(^{489}\) variants, the impact of errors on SARS-CoV-2 proteins may influence the mortality and morbidity due to SARS-CoV-2 infection. Spike protein RBD is an element in this set and may remain as the prominent antigen of choice for mRNA vaccines, albeit in the short-term. The nsp14 protein works with the non-enzymatic\(^{490}\) nsp10\(^{491}\) in an error correction complex to restore replication fidelity. Both are susceptible to mutations. Reduced efficacy of either protein may deliver deleterious outcome for humans if the new mutations introduce changes in the progeny virus which may be harmful to humans.

Structure and function (genomic/proteomic) comparison of nsp14 and nsp10 (independently and in complex) between sister clades of SARS-CoV-2 variants (D614G\(^{492}, 493\), N501Y\(^{494}\), etc.\(^{495}\)) may offer clues, if we can identify specific mutations in the parent genome which affected nsp14 (exoribonuclease for proofreading) and/or nsp10, resulting in progeny which causes more harm.

The generality of evolution\(^{496}\) and its punctuated equilibrium\(^{497}\) introduces uncertainty in the sudden/dynamic appearance of variants which manifests\(^{498}\) differently depending on the host. The latter warrants development of a \textit{portfolio} of therapies. Creating a portfolio of synthetic mRNA vaccines is perhaps a better and/or efficient immune-dependent pharmacological strategy compared to the classical approach of using attenuated virus in a viral vector\(^{499}\). But, high energy\(^{500}\) consuming products (vaccines) are incompatible with distribution to energy-deprived regions, nations and continents.

Appendix II outlines the rigor of kinetics necessary to establish the criteria for DNA aptamers. The foundation of the science in terms of molecular biophysics is without question. The silent success of SomaLogic\(^{501}\) may be a “radio silence” because it was not intended for publicity (for example, the media-cultivated baseless drum beat of “Smart Cities” that is about to bite the dust\(^{502}\)).
Trials and tribulations\textsuperscript{503} involved in the mRNA vaccine strategy may have a parallel story\textsuperscript{504} in the innumerable\textsuperscript{505} failures (200+ experiments) that preceded \textit{one} successful protocol for creating LNP (lipid nanoparticles, see Figure 26) to deliver charged macromolecules to cells, \textit{in vivo}. Use of aptamers as an alternate path for 80% of the global population may have to endure pain, shame and ignominy on the fatiguing climb, hopefully, to claim success, albeit partial. This “\textit{dismissed tool}” (DNA aptamers) may lead to alternative diagnostics and therapeutics to serve the globally underserved but it will not be easy.

Unique and degradation-resistant aptamers with enhanced binding affinities may unleash new approaches. But attempts to expand our genetic vocabulary has met with limited success. Modified\textsuperscript{506} nucleotides “artificial” to nature are available. Will the inclusion of modified bases in aptamers result in favourable changes in terms of kinetics and dynamics?

![Table II: This table (Biondi and Benner, 2018) shows DNA aptamers containing artificial nucleotides in “red” letters (P and Z). Aptamers with the lowest $K_d$ (highest affinity, slow turnover) contained Z and/or P residues (LZH1 $K_d$ 14 nM) while natural AGCT-only aptamers had $K_d$ ranging from 326 nM (LZH13) to >1 \mu M (LZH16, LZH17). Percentage in the pool indicates relative amounts of a particular sequence to the total sequences analyzed from cycle 13. Six\textsuperscript{507} nucleotide DNA alphabet containing AGCT (Adenine, Guanine, Cytosine, Thymine), Z \{6-amino-5-nitro-3-(1′-β-D-2′-deoxyribo-furanosyl)-2(1H)-pyridine\} and P \{2-amino-8-\{(1′-β-D-2′-deoxyribofuranosyl)imidazo\[1,2-a\]-1,3,5-triazin-4(8H)-one\}\} was guided by Watson-Crick base pairing ability. Perhaps a larger number of “artificial” DNA and RNA alphabets may be available for use in single stranded aptamers to improve specificity and sensitivity.]

---

\textsuperscript{503} ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
APPENDIX V – ENGINEERING RNASE RESISTANT RNA APTAMERS: A PARALLEL APPROACH?

DNA aptamers may be easier to work with compared to RNA aptamers. Which one is better? There is no single answer. Unless created and evaluated, we do not know which synthetic aptamers may be lower cost alternatives for diagnostics and therapeutics for 80% of the less affluent global population. One shoe does not fit all. We need portfolio of tools from an epidemiologic perspective because bias is omnipresent in science and variations in human physiology are far beyond the grasp of medicine.

Ubiquitous presence of copious amounts of mammalian (extracellular) ribonucleases (RNase A, RNase P) and to a lesser degree the intracellular RNase L, if present in cell-free extracts, induces fear in scientists. It makes it harder to work with RNA tools for diagnostics and therapeutics. On the other hand, the configurational versatility of RNA makes it an efficient “lasso” to bind proteins. We focused on ssDNA aptamers because the deoxyribonuclease families may be less of a problem.

Thirty years ago this conundrum was visited by none other than Paul Zamecnik. More than a decade ago the seminal paper by Katalin Karikó (Karikó et al 2005) was the key which later enabled Moderna and Pfizer/BioNTech’s mRNA vaccine. Advances in therapeutics over the past few years provide optimism. Use of modified pyrimidines and purines to create synthetic RNase-resistant single stranded RNA aptamers may not be an insurmountable challenge for nucleic acid chemists.

As outlined in Appendix II, establishing rigorous pharmacokinetics and pharmacodynamics of synthetic RNase-resistant single stranded RNA aptamers for specific protein targets (communicable diseases and non-communicable diseases) may be similar but with at least one additional qualification. The pre-requisite is a two part pre-phase where the range of resistance must be determined (profile of RNase attack on naked RNA aptamers in cell-free extracts in vitro) to serve as a qualitative indicator. Verification of the stability of the RNA aptamer to remain intact (structure, function) in cell-free extracts in vitro compared with saliva samples (from uninfected/normal human and animal) may be critical as a metric or index of comparative stability to serve as a quantitative guide. These tasks are not easy.

APPENDIX VI – SCIENCE, SOCIETY VS SENSITIVITY, FREQUENCY OF TESTING AND COST

Discussion with global experts (see “acknowledgements”) points to the relative naïveté (?) of this proposal with respect to aptamers as potential diagnostic (ADD) and therapeutic (AAAD) tools in detection and prevention of SARS-CoV-2. Thirty years of advances in aptamers may appear to suggest that this proposal is inadequate to approach, meet or exceed the existing excellent body of science which has been accomplished with respect to specificity and sensitivity of aptamer-protein interactions.

Sub-attomolar detection limits are unlikely to be attained by ADD tools. Do we need such sensitivity for ADD as a global public health tool? Is detection at 10⁻¹⁸ moles per liter (concentration) necessary to detect SARS-CoV-2 in (asymptomatic) individuals to reduce transmission? What is more important: frequent, low-cost, low sensitivity tests or infrequent, high-cost, highly sensitive qRT-PCR?
How many viable SARS-CoV-2 must invade a human in order for the individual to be infected? *The answer is unknown.* In ferrets and cats the infectious dose is $10^5$ colony forming units (cfu)\(^{524}\) which suggests that attomolar, femtomolar or nanomolar levels may be unnecessary for detection of viral RNA or proteins. Can micromolar sensitivity (*twelve orders of magnitude* lower than attomolar) suffice to detect infection by viruses, such as SARS-CoV-2? Just because high performance Koenigsegg *Agera* is the vehicle of choice for a few in Chicago or Cologne or Cannes, does not mean that it a global standard for automobiles or must be *the* adopted make and model for driving in Calcutta or Cairo or Cartagena\(^{525}\).

Figure 36: Sensitivity is *not* the Holy Grail. The evidence to maximize sensitivity may not be the golden rule\(^{526}\). Sensitivity is a *shoe where one size does not fit all.* The difference in virus concentration (limit of detection, LOD) between high and low sensitivity tests may be two orders of magnitude (TOP panel, note red zone, horizontal bar). The time for the viral load to leap from 10 or 100 ($10^1$ or $10^2$) to 1,000 or 10,000 ($10^3$ or $10^4$) is a matter of hours (*viral lifecycle*). What is detectable by gold standard qRT-PCR ($10^1$ or $10^2$) is *also* detectable with a lower sensitivity test ($10^3$ or $10^4$ virus concentration), a few hours later. For control of transmission, degree of sensitivity is neither critical nor rate limiting with respect to LOD (BOTTOM panel). The *frequency*\(^{527}\) of testing is key to controlling the pandemic. Low-cost, rapid, lower-sensitivity tests are essential to establish the epidemiology\(^{528}\) of the population already\(^{529}\) infected.
Therefore, the notion that attomolar or femtomolar is the “best” or the “standard” of sensitivity for enabling access to public health is a misguided bias in the Western world. For a different purpose, a different level of sensitivity may work for a community if the tool delivers value by reducing mortality and morbidity, perhaps not for all, but may be for 80% of the population. ADD tool will deliver value.

The criteria of good and best is subject to modification by socio-economic yardsticks when the “best” is out of reach of the people and government healthcare budgets. The value of the service must not be compromised just because it is used by people who are less litigious and less affluent. ADD is also expected to serve the underserved nations where 80% of the global population resides. This is not only about science in pursuit of the truth but also the role of science as a service to society to lift many boats, not just a few yachts.

Quote [A]: Three decades of grand scientific accomplishments in the field of aptamers may induce a novice to hear the echo of the quote “who am I that I should do better” attributed to Alfred North Whitehead. Similar presuppositions may have fueled attempts by Harvard (established 1636) to “absorb” MIT (established 1861) in six failed merger attempts (1862 through 1905). Unquestionable excellence in the domain of aptamers may also serve the underserved in the post-pandemic world but it may not happen if we are trying to gild the lily or without the guidance and support from the experts.

Comparing sensitivity of detection between tools may lead to imprecision sensitivity (the conventional wisdom surrounding “apples versus oranges” problem). The elements that may be measured in an ADD-type detection tools shares few common features with tests using quantitative RT-PCR (reverse transcriptase – polymerase chain reaction). Perhaps comparing the outcome of the tests may offer a common feature as another dimension of sensitivity. If two tools were used to test the same group of individuals and the tests were scored, with respect to the ratio of positive over false positive versus negative over false negative, then this ratio (or an appropriate combination of positive, negative, false positive, false negative) may serve as an indicator of meaningful sensitivity (compared to the gold standard).

54 • ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
The caveat in this framework is that the notion of false negative as a function of the true viral load is a difficult metric. Nevertheless, the distribution of the outcome data (several tools compared to the gold standard) may reflect the range or the spectrum which is useful for diagnostic purposes. The latter is the key to determining the efficiency of each tool, for example, if the score falls within a certain range then they are useful. For ADD, if optical signal transduction from quantum nano dots (to smartphone) is inefficient in terms of signal versus noise, then other options are available (electrochemical impedance spectroscopy, surface plasmon resonance, etc.) to obtain the data.

Figure 37: “Frameworks” (Eric Lander). False positives and false negatives as a function of true viral load are mired in numerical controversy due to LOD and inherent errors in biological measurements.
Specificity may turn out to be equally critical in the context of testing individuals who may be affected by multiple viruses at the same time. Patients with persistent infections may be exposed to yet another virus. Context-aware diagnostic specificity must differentiate between viruses in the event that the recognition element shares homology or something related.

Epidemiologic indicators (for example, R0) are statistical tools which are entirely dependent on correct data. The quagmire of data sharing, fake data, suppressed data and data-driven misinformation makes a mockery of such KPIs (key performance indicators). Using qRT-PCR as the ultra-sensitive gold standard tool to provide data for R naught estimation of transmission in local communities may be limited to a few cities in the world while the pandemic continues its march of unreason, globally.

CoVID-19 pandemic may morph into a globally endemic infectious disease (EID). The resulting situation may be akin to a social IED (improvised explosive device) which could transform geo-political economy. Demand for alternatives to vaccines may be served by small molecules, including aptamers, for diagnostics and therapeutics. The grave disequilibrium facing 80% of the global population calls for tools to deliver value rather than pursue six sigma R&D to reach the luminous summit where “best, highest or lowest” resides within the hall of fame for metrics of excellence with respect to characteristics, attributes and features.

Contrary to what one might conclude, this proposal neither promotes nor includes that which is shoddy and second grade masquerading as good enough. Re-visiting aptamers for ADD and AAAD with “new” eyes is a structured exploration and convergence of tools and protocols to deliver value optimized to help billions of users. But, it may fail to be useful. Yet, the pursuit must continue “not because the work is easy but because the work may be hard”. The plight of ADD and AAAD may be a tryst with destiny.

APPENDIX VII – ROOM FOR IN SILICO CREATIVITY IN PREDICTING APTAMER DESIGN?

Nucleic acid polymers binding to a protein or peptide with specificity in body fluids is at the scientific heart of the principle salient both for detection (ADD, binds in an external environment) and therapy (AAAD, in vivo anti-viral activity). To create a set of these highly specific aptamers, the usual approach may start with a library of aptamers (say \(10^{19}\) or \(10^{20}\)) and through sequential enrichment (principle of column chromatography, one molecule is immobilized) the aptamer pool may be reduced \((10^2 < x < 10^3)\) using a range of \(K_d\) values as a metric (see Table II in Appendix IV).

Molecular design in silico may be worth exploring to predict aptamer design and consequently focus on testing a smaller pool of synthetic aptamers. The caveat in this modus operandi is the departure from pristine principles of chemistry (the assumptions in the design process). The criteria for binding is determined by shape and may be similar in principle to some variation of “induced fit” which is the underlying mechanism for enzyme-substrate interactions prior to catalysis.

---

56 • ADD is in the MIT Library [https://dspace.mit.edu/handle/1721.1/128017](https://dspace.mit.edu/handle/1721.1/128017) and ChemRxiv [https://doi.org/10.26434/chemrxiv.13102877](https://doi.org/10.26434/chemrxiv.13102877)
The shape or configuration of a single stranded DNA aptamer is guided by base-pairing between distant stretches which could generate “hairpin” structures. The *inter*-hairpin hydrogen bonding\textsuperscript{543} between base pairs and/or other electrostatic interactions\textsuperscript{544} may generate tertiary structures. Can a pool of ssDNA give rise to tessellated, concatenated or catenated DNA? Are the structures/hairpins stable enough to interact with specificity with the protein target and generate a signal upon binding? Can the *in silico* approach predict the structure, a measure of its stability and the probability of binding?

As mentioned in the previous paragraph, the attributes associated with these *in silico* models assumes that nucleotides are interacting in a *standard* state. The *in silico* approach largely depends who programmed the principles of physics and chemistry associated with each molecule. One assumes the interaction is occurring in some form of optimized buffer with known hydrophobic and hydrophilic profiles, specified van der Waals radii, hydration values, pH, ionic content and a mixture of salts.

The situation is exponentially complicated when we move from ssDNA aptamers in “ideal” solutions to aptamers co-existing with large/small proteins/peptides in a colloidal solution of variable pH with different factors (for example, mucus in saliva). The amino acid sidechains play a defining role in binding. Factors that perturb secondary and tertiary structure of protein/peptide will influence binding (specificity, $K_d$). Protein structures are dynamic (think resonance structures of organic molecules) and the zwitterionic status of amino acid sidechains may be sensitive to presence of other charged molecules (aptamers), pH, salinity and viscosity.

Viscosity is not so subtle an influencer because the state of fluidity (fluid dynamics, kinetics) influences Brownian motion of molecules in solution and the frequency of collision between molecules. The effect of temperature on kinetics is an equally important parameter we are avoiding, at present, because the environment of external interaction (ambient temperature at the point of use) and *in vivo* applications (body temperature in humans/animals) are unlikely to fluctuate beyond a few degrees.

Taken together, predictions from the *in silico* model under “ideal” conditions may *immediately* fall apart when aptamers (which were predicted to be “good” *candidates in silico*) are evaluated for binding strength (aspirational low $K_d$) and specificity in body fluids, for example, saliva and blood, which may diverge from the computational “ideal” state (fluids may also vary between individuals).

But, this is not a fact. This is a pessimistic *hypothesis* that the *in silico* prediction may not even come close to mimicking the experimental milieu when evaluating aptamer binding to target proteins in saliva, blood or other fluids (*in vivo*). Nevertheless, *in silico* modeling and prediction is a scientific design tool worth evaluating before we conclude that the “ideal” state predictions are at complete odds when testing aptamer binding in body fluids.
For SARS-CoV-2, there is a “baseline” interaction which could serve as a template for *in silico* design and prediction tool. ssDNA aptamers were identified to bind to the Spike (S) protein RBD (Song, Yanling, *et al.*, 2020) and nucleocapsid (N) protein (Chen, Z. *et al.*, 2020) with affinity below 5 nM (one $K_d$ 0.49 nM in Zhang, Liyun, *et al.*, 2020). The *in silico* approach may model the parameters of the N protein using data/attributes relevant to the aptamers identified. The model must include primary sequence of the binding region on the N protein followed by the putative secondary structure of the region (both as a peptide and as a region in the intact protein). Linking *in silico* models with protein structure data is necessary to understand how the secondary structure may change due to mutations in the primary structure. This connectivity is required to test if variants of SARS-CoV-2 may be defective in binding certain aptamers. The ADD “sensible” system will encounter individuals harboring variants other than the “model” virus on which the *in silico* approach may be trained.

The *in silico* logic tools must connect with the knowledge base with attributes of the “alphabets” (DNA alphabets A, T, G, C) which are *featured* in the sequence of the ssDNA. Displaying the theoretical structure of that sequence in solution (hairpins) is a required step. Then proceed to predict which part of the ssDNA aptamer secondary structure interacts with the secondary structure of SARS-CoV-2 protein. The “*glaçage sur le gateau*” will be the ability of the *in silico* tool to display the tertiary structure of the specific region of the SARS-CoV-2 N-protein bound by the aptamer. Creating a numerical framework of values must accompany the predictions. The tool needs a knowledge base to import $K_d$ values and other equilibrium constants associated with this interaction for quantitative predictive analytics.

Can this “modelling” predict negative changes in $K_d$ (higher $K_d$ value) if we change the ssDNA aptamer sequence to unfavorably perturb the “model” secondary structure or change one or more of the amino acids at the “active” site of the aptamer-protein tertiary structure? If this *in silico* tool can provide the anticipated response (in this “known” scenario) then it suggests that the model is using a relatively reasonable set of attributes. The data store of attributes may be enriched by connecting or feeding the *in silico* system with feature-specific data to populate selected features in the model. These characteristics of feature engineering are not essential because data volume related to aptamer-protein binding is limited.

However, what is much more important, for the tool to be immediately useful, is the ability of the model to predict the impact of changes in aptamers and aptamer-protein *binding* due to changes in protonation and deprotonation from fluctuations in pH. The pKa of amino acids are susceptible to pH changes and that could modify configuration which could affect binding, either positively (lower $K_d$) or negatively (higher $K_d$). Creating model dependencies between pKa, $K_d$ and Ramachandran parameters (see Figure 30) may be an important part of the analytical engine in this *in silico* model/tool. Experts may point to many other factors we have omitted. But, if the issues in the above discussion are addressed with some degree of confidence, then the *in silico* model may be ready for a design challenge. The overarching question is the ability of the tool to predict the design of aptamers which could bind target proteins with high specificity.
Using the SARS-CoV-2 N protein and the ssDNA aptamers which binds to the N protein as the base model in the *in silico* tool, could we substitute the N protein with the SARS-CoV-2 Spike (S) protein RBD (receptor binding domain) and then ask the *in silico* tool to predict, based on its learning, which ssDNA *aptamer shapes* are appropriate candidates for binding to the S protein RBD (S1)? The next expectation is a list of ssDNA *sequences* that could generate the aptamer *shapes* (predicted *a priori*) with an accompanying list of predicted Kd for each sequence-shape combination.

Comparing the outcome of this challenge with the data from the reported (Song, Yanling, *et al*., 2020) binding of DNA aptamer with the S protein RBD, we can estimate the fit and reliability of the *in silico* model which used the data from the N protein as the template. What if the prediction of the *in silico* tool is incompatible with the reported aptamer that binds with the S protein? Perhaps there are at least three options: [i] abandon the *in silico* approach or [ii] start over with a different template which has rigorous data or [iii] use the untrustworthy, trial and error prone back propagation algorithm, an error correction routine licensed to practice arbitrary numerical malfeasance.

In the name of supervised learning the attempt is to “train” artificial neural networks (ANN) to deliver what the trainer wants the output to deliver. For example, if the “training” failed to recognize a hot dog, after training with millions of photographs of hot dogs (if the ANN output is a baseball bat), then the algorithm, goes back through each step to push and pull and adjust arbitrary “weights” associated with each step, in order to drive the ANN to output “hot dog” as the outcome of the supervised learning.

To apply the algorithm in this case, the gulf between the predicted versus observed S protein binding aptamer needs to be narrowed down by pushing and pulling the *in silico* machine learning training tool to deliver the reported (Song, Yanling, *et al*., 2020) S protein binding aptamer sequence (based on the initial training of the *in silico* model using the N protein binding aptamer). Die-hard believers observe that “force fitting” in back propagation algorithm (BPA) is sort of similar to curve fitting of data to models and mimics the induced fit mechanism of enzyme action (Koshland, 1958) representing biomimicry.

*If “made” useful*, this machine learning embedded *in silico* analytical tool may be trained to be deployed in a *retrosynthetic* mode. The key variables (aptamer and protein) leads to the Kd value. Using an oversimplified analogy, Ax (aptamer) + By (protein) = Cz (Kd) in some form or the other. If we “fix” the desired value of Kd (or a range or set of values) and pre-set the target protein, then, can we ask the tool to predict the best-fitting shapes and corresponding optimized ssDNA sequences? Can we simulate scenarios using a drag and drop menu (Kd, protein, etc.) and enter values in the *in silico* simulation tool in a manner similar to variant configuration in computer-assisted design?
Automation of small molecule discovery is the bread and butter of drug design in the pharma industry. Convergence of bio-physics, chemistry and computer science embedded as principles in logic tools are trained to predict and guide the synthesis of designer molecules. There is a healthy place and need for some forms of machine learning techniques which may be valuable. Not all routines are as ill-conceived and over-hyped as back propagation algorithm and its recent attempt to re-invent itself.

The learning that repurposed the antibiotic Halicin, is one good example. Training the MPNN (message passing neural networks) was structured and deployed hyperparameter optimization, without any artificial intelligence, as explained elsewhere. Ensembling was applied to improve outcomes but predictions were biologically tested. Even after steps to minimize errors, the authors were cognizant to note: “It is important to emphasize that machine learning is imperfect. Therefore, the success of deep neural network model-guided antibiotic discovery rests heavily on coupling these approaches to appropriate experimental designs.” (Stokes et al, page 698)

Keeping in mind these cautionary notes, can the in silico tool reliably predict sequence of ssDNA aptamers which may bind with low Kd? Of course, the predicted aptamers must be rigorously tested in saliva samples to answer the question. But, if we can translate the prediction to reality (testing in saliva using an aptamer-based sensor), then we may accomplish an in silico milestone. Even if it is only partially effective, this tool may still predict, guide and design the synthesis of aptamers which may make it less arduous to explore the therapeutic application of aptamer-as-a-drug (AAAD). The demand for binding specificity for AAAD is not only more stringent but is intensely complicated because the binding must remain functional in vivo. Training the in silico tool to include P and Z (Table II) as “alphabets” in the prediction engine may enable the tool to predict ssDNA aptamers containing artificial nucleotides, which appear to improve binding (see aptamers with lowest Kd in Table II).

Thus far we have discussed binding as if the specific viral proteins are available to bind. The latter may be true when using purified recombinant protein as test molecules. Figure 20 removes all doubt that bio-availability of the exposed protein for binding is far from what a deterministic model could predict. Therefore, a negative outcome does not necessarily indicate absence of the target protein (by extrapolation, absence of the virus). The inability of an aptamer to bind to the target protein in an actual test with body fluids (saliva sample with live virus) may be obstructed by the glycan shield which prevents the protein from being exposed or available to the aptamer for binding.

Could we alleviate this problem by using ENGases (see Section 2 and Fairbanks, Antony J., 2017)? Experiment alone can shed light whether ENGases can punctuate glycan shields sufficiently to enhance the ability of the aptamers to bind and elicit a signal (over noise) which can be transduced for acquisition of data, as proof of binding. Although the focus is on the SARS-CoV-2 Spike protein it behooves us to explore the parallel use of one or more of the other external proteins of SARS-CoV-2 (nucleocapsid [N], membrane [M] and envelope [E] proteins) as targets for detection.
APPENDIX VIII – FIELD TESTING THE APTAMER SENSOR, ALONE, MAY NOT BE PRUDENT

The rubber meets the road in the outcome, that is, the data, which must indicate with confidence if an infectious agent (for example, SARS-CoV-2) is present or absent. For return on investment (ROI), this data is the Holy Grail. Ambiguity, uncertainty and false negatives will obliterate the value of this data and may annihilate the ROI. False positives may hurt the reproducibility and reliability of ADD. Selling for-profit tests lacking the foundation of robust kinetic data and scientific field testing without proper controls are criminal practices. University of Oxford considered forty devices but only nine met some of the performance criteria, and only one was advanced enough for assessment.565

<table>
<thead>
<tr>
<th>Positive Control</th>
<th>Negative Control</th>
<th>Aptamer Test</th>
<th>Confidence in Outcome (Data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>B</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>D</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>E</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>F</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table III: Design of field trial may start with independent scoring of each category before the combined scoring. Positive control is tested with 100 samples which are qRT-PCR positive for SARS-CoV-2. Let us assume that the score from the positive control is 90% (compared to qRT-PCR which has a lower LOD). The same 100 samples are tested with the negative control and scores 99% (1/100 samples “detected” what the negative control is supposed to detect, but generally the target is never found in human saliva). The same 100 samples are tested with aptamer test with a score of 75%. Therefore, is the “ambiguous” in row B [table] less ambiguous than the “ambiguous” in row C, if we take into account the positive control is correct 90% of the time versus the aptamer test which is correct 75% when compared to qRT-PCR? Are we comparing apples vs oranges? Can this test serve as a low-cost tool to control transmission? In the real world, tests must be performed daily to isolate individuals who test positive, early enough, at least to reduce transmission in the absence of treatment (TETRIS – TEst, TReat, ISolate). Criticism of the aptamer test may ask whether the aptamer is functional. Positive aptamer function test (AFT) may include an aptamer which binds to a known component of saliva (for example, salivary amylase). As a control for AFT, inclusion of an aptamer for the vasoactive peptide hormone BNP (Sudoh et al 1988, Sudoh et al 1990) and/or the human ACE2-short isoform (Blume et al 2021) may be of clinical interest. Inclusion of the former (BNP) may uncover physiological co-morbidities independent of SARS-CoV-2 and/or it may be significant in view of cardiovascular (CVD) symptoms due to CoVID-19.
Proper implementation of field trials are crucial. Aptamer tests (ADD) must include positive and negative controls. A positive control must use a completely different medium to test the same sample for the presence of the infectious agent. A negative control should test the sample using the same medium as the positive control which serves as a control to confirm that the “medium” is not faulty and recognition is functional. The negative control tests for an agent that may not be found in saliva. Future testing may be able to create a 3-in-1 single test system with different modes of signal transduction\(^{568}\) generating different streams of data from optical signal, electrochemical impedance spectroscopy and surface plasmon resonance. The data can be analyzed and visualized on a smartphone.

For resource unlimited scenarios, testing kits may include another set of positive and negative controls. The positive control (P2) medium remains identical to P1 but tests for an agent not uncommon in the oral microbiome\(^{569}\). The second negative control (N2) medium is identical (as in P1 and N1) but tests for a rare but not impossible agent, for example, HIV1. In a completely different approach, we can use immobilized hACE2 on the sensor surface as the medium (SARS-CoV-2 positive) and use hACE2-short isoform\(^{570}\) as a control because hACE2-short isoform lacks the virus binding domains.

THE CASE FOR LECTIN SENSORS AS CONTROL

To change the medium of sensing we turn from nucleic acids (DNA aptamers) binding to target proteins (for example, Spike protein in SARS-CoV-2) to proteins (lectins) binding to target sugar groups (glycans) which may be on the same target protein (for example, glycan shield on the Spike protein of SARS-CoV-2).

Lectins are carbohydrate-binding proteins that are specific for sugar groups on other molecules. Host-derived sugar moieties on viral proteins are tools for immune evasion\(^{571}\) whereas lectins (opsonin) trigger immune recognition\(^{572}\). The glycan shield on SARS-CoV-2 Spike protein (Casalino et al, 2020) is an excellent target for lectins, for example, mannose-binding lectin (MBL), which was reported a decade ago, to neutralize SARS-CoV mediated viral infection\(^{573}\). Macrophage galactose binding lectin (MGL) belonging to the Calcium-dependent C-type lectins\(^{574}\) and subsets of I-type lectins (sialic acid-binding immunoglobulin-type lectins), Siglec-9 and Siglec-10, specifically binds\(^{575}\) to SARS-CoV-2 Spike protein.

The role of lectins and the science behind mammalian and non-mammalian lectins are well\(^{576}\) studied. Lectin-based adjuvants to vaccines\(^{577}\), lectin antibodies\(^{578}\) (lectibodies\(^{579}\)) and lectin-based\(^{580}\) biosensors\(^{581}\) are well documented\(^{582}\). Creating SARS-CoV-2 specific lectin sensors may not be trivial but certainly not an insurmountable barrier. Use of a lectin sensor as the positive control (Table III) needs no further emphasis. The negative control may be a mannose-binding lectin (MBL) sensor or something even more specific for detecting Mycobacterium leprae\(^{583}\) which prefers to grow in cooler extremities of the human body (around \(\sim 30^\circ\text{C}\)) and unlikely to be present in human saliva samples.
Table IV: The choice of lectin sensors as controls to improve confidence in data from aptamer tests. Non-mammalian lectins with anti-viral activities (top). Use in biosensors (bottom). Man, mannose; GlcNAc, N-acetylglucosamine; Glu, glucose; Gal, Galactose; LAP, lectin affinity plasmapheresis.

<table>
<thead>
<tr>
<th>Lectin</th>
<th>Source</th>
<th>Sugar affinity</th>
<th>Anti-SARS-CoV-2 action</th>
<th>Anti-SARS-CoV action</th>
<th>Anti-MERS-CoV action</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRIL</td>
<td><em>Lablab purpureus</em></td>
<td>Man/Glu</td>
<td>In vitro</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>APA</td>
<td><em>Allium porum</em></td>
<td>Man</td>
<td>—</td>
<td>In vitro</td>
<td>—</td>
</tr>
<tr>
<td>Morniga M II</td>
<td><em>Morus nigra</em></td>
<td>Man</td>
<td>—</td>
<td>In vitro</td>
<td>—</td>
</tr>
<tr>
<td>EHA</td>
<td><em>Epipactis helleborine</em></td>
<td>Man</td>
<td>—</td>
<td>In vitro</td>
<td>—</td>
</tr>
<tr>
<td>UDA</td>
<td><em>Urtica dioica</em></td>
<td>GlcNAc</td>
<td>—</td>
<td>In vivo and in vitro</td>
<td>—</td>
</tr>
<tr>
<td>NICTABA</td>
<td><em>Nicotiana tabacum</em></td>
<td>GlcNAc</td>
<td>—</td>
<td>In vitro</td>
<td>—</td>
</tr>
<tr>
<td>Con A</td>
<td><em>Canavalia ensiformis</em></td>
<td>Man/Glu</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>HHA</td>
<td><em>Hippeastrum hybrid</em></td>
<td>Man</td>
<td>—</td>
<td>In vitro</td>
<td>—</td>
</tr>
<tr>
<td>GNA (or GNL)</td>
<td><em>Galanthus nivalis</em></td>
<td>Man</td>
<td>—</td>
<td>In vitro</td>
<td>—</td>
</tr>
<tr>
<td>Cyanovirin-N</td>
<td><em>Nostoc ellipsosporum</em></td>
<td>Man</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Griffithsin</td>
<td><em>Griffithia sp.</em></td>
<td>Man</td>
<td>—</td>
<td>In vivo and in vitro</td>
<td>In vitro</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lectin</th>
<th>Source</th>
<th>Sugar affinity</th>
<th>Type of application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con A</td>
<td><em>Canavalia ensiformis</em></td>
<td>Man/Glu</td>
<td>Biosensor for viral detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Biosensor for detection of serum glycoproteins</td>
</tr>
<tr>
<td>Cramoll</td>
<td><em>Cratyli mollis</em></td>
<td>Man/Glu</td>
<td>Biosensor for detection of serum glycoproteins</td>
</tr>
<tr>
<td>BmoLL</td>
<td><em>Bauhinia monandra</em></td>
<td>Gal</td>
<td>Biosensor for detection of serum glycoproteins</td>
</tr>
<tr>
<td>GNA (or GNL)</td>
<td><em>Galanthus nivalis</em></td>
<td>Man</td>
<td>—</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

The author is thankful to Brittany Newell (section 2), Larry Gold (section 5), Andrew Fire and Greg Winter (for section 6) as well as John Rossi, Britt Glaunsinger and Sanjay Sarma (for general review). The critical review by Jack Szostak expressed skepticism about the sensitivity of the optical signal (from a nano quantum dot conjugated DNA aptamer bound to the target viral protein in a sample of saliva from humans or animals). Jack’s skepticism focused on the comparative sensitivity between the proposed hypothetical optical signal versus signal from traditional qRT-PCR (reverse transcriptase-polymerase chain reaction) of viral RNA genome. (Jack Szostak is Chairman of the Scientific Advisory Board of Moderna Inc., Cambridge, MA)
More than 39 million doses of vaccine have now been administered in at least 49 higher-income countries. Just 25 doses have been given in one lowest-income country. Not 25 million; not 25 thousand; just 25.

This note is a suggestion and a proposal from Shoumen Datta (SD) based on published scientific research. It is not new and copyright free. It may be used by anybody for any purpose without any need to cite or credit the author. If you have questions please email shoumen@mit.edu and/or sdatta8@mgh.harvard.edu
REFERENCES


https://doi.org/10.3390/diagnostics10010022

“SNAPS TRILOGY” MIT Libraries https://dspace.mit.edu/handle/1721.1/123983

https://news.mit.edu/2020/kinsa-health-0821

15 Lamb, James (2020) ADD Appendix – Description of Major Components (personal communication)
https://github.com/jameslamb


18 APPENDIX Figure 1: Description of Components https://github.com/shoumendatta/ADD-DIGITAL

https://emclamor.wixsite.com/mclamorelab

https://www.health.harvard.edu/blog/which-test-is-best-for-covid-19-2020081020734

21 Classification: True vs. False and Positive vs. Negative
https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative


23 World Economic Forum (2016) 4 billion people still don’t have internet access.
https://www.weforum.org/agenda/2016/05/4-billion-people-still-don-t-have-internet-access-here-s-how-to-connect-them/

24 Likert Scale https://www.sciencedirect.com/topics/psychology/likert-scale


26 CDC - 2019-nCoV Real-Time RT-PCR Diagnostic Panel. www.fda.gov/media/134920/download


33 Data Distribution Service https://www.dds-foundation.org/what-is-dds-3/
https://www.nature.com/articles/s41467-020-18148-7.pdf
https://science.sciencemag.org/content/sci/369/6509/1318.full.pdf
40 Gold, Larry (2020) Personal Communication https://amzn.to/2QGSz1i
www.thelancet.com/action/showPdf?pii=S2352-3964%2820%2930229-2
42 https://somascandiscovery.com/publications/
https://www.pnas.org/content/114/14/3619.full.pdf

ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


https://pdfs.semanticscholar.org/aad2/985a4bc40430260e3570683cb2b3b79f7e7.pdf?ga=2.2426903061804402263.1599417152-1424810635.1595859001


https://link.springer.com/content/pdf/10.1007/s00253-020-10747-0.pdf

79 Sumedha D Jayasena, Aptamers: An Emerging Class of Molecules That Rival Antibodies in Diagnostics, Clinical Chemistry, Volume 45, Issue 9, 1 September 1999, Pages 1628-1650.
https://doi.org/10.1093/clinchem/45.9.1628


https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2651625/pdf/bc8001248.pdf


71 • ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
https://pubs.rsc.org/en/content/articlepdf/2009/an/b906788d
https://science.sciencemag.org/content/369/6507/1041/tab-pdf
doi:10.1016/0003-9861(64)90322-4
93 Nucleocapsid Phosphoprotein: Severe Acute Respiratory Syndrome Coronavirus - Protein - NCBI. https://www.ncbi.nlm.nih.gov/protein/YP_009724397.2
94 Nucleocapsid Protein: Severe Acute Respiratory Syndrome-Related Coronavirus - Protein - NCBI. https://www.ncbi.nlm.nih.gov/protein/NP_828858.1
https://pubs-rsc-org.libproxy.mit.edu/en/content/articlepdf/2020/cc/d0cc03993d
ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


NSF Award Abstract #2028531(RAPID) Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)-Prevention: Multiple-Site Binding with Fusing Aptamers to mitigate Coronavirus Disease 2019. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2028531&HistoricalAwards=false


Zhou, DJ (2012) Quantum dot-nucleic acid/aptamer bioconjugate based fluorimetric biosensors. Biochemical Society Transactions 40 635 – 639 ISSN 0300-5127 https://doi.org/10.1042/BST20120059

https://www.microsoft.com/en-us/hololens/hardware


https://www.aptamergroup.co.uk/aptamers-in-the-battle-against-covid19/
ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877

114 https://en.wikipedia.org/wiki/Kinect


117 https://emclamor.wixsite.com/mclamorelab


74 • ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


www.ncbi.nlm.nih.gov/pmc/articles/PMC1287568/pdf/1240-05.pdf


https://jvi.asm.org/content/jvi/79/24/15189.full.pdf


75 • ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


150 Streisinger G, Okada Y, Emrich J, et al. (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb Symp Quant Biol. 1966; 31:77-84. doi:10.1101/sqb.1966.031.01.014 http://symposium.cshlp.org/content/31/77.long


78 • ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877

https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2770945


https://www.biorxiv.org/content/10.1101/2020.04.29.069054v2.full.pdf

ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


https://www.thelancet.com/action/showPdf?pii=S0140-6736%2820%2931866-3


https://science.sciencemag.org/content/early/2020/09/18/science.abd7343


The Global HIV/AIDS Epidemic


https://science.sciencemag.org/content/368/6494/948/tab-pdf


Our World in Data
https://ourworldindata.org/grapher/daily-covid-cases-7-day?time=2020-03-31..latest


Fauci, Anthony (2020) MIT Biology course number 7.00 (COVID-19, SARS-CoV-2 and the Pandemic)

Baltimore, David (2020) MIT Biology course number 7.00 (COVID-19, SARS-CoV-2 and the Pandemic)


https://www.medrxiv.org/content/10.1101/2020.05.22.20110403v1.full.pdf


ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


---

82 • ADD is in the MIT Library [https://dspace.mit.edu/handle/1721.1/128017](https://dspace.mit.edu/handle/1721.1/128017) and ChemRxiv [https://doi.org/10.26434/chemrxiv.13102877](https://doi.org/10.26434/chemrxiv.13102877)


217 Torrente-Rodriguez, Rebeca et al., *SARS-CoV-2 RapidPlex: A Graphene-Based Multiplexed Telemedicine Platform for Rapid and Low-Cost COVID-19 Diagnosis and Monitoring*. Matter (Oct 1, 20)
https://doi.org/10.1016/j.matt.2020.09.027 ; www.cell.com/matter/pdfExtended/S2590-2385(20)30553-1

https://cocoa.ethz.ch/downloads/2014/06/None_MIT-AUTOID-WH-014.pdf

https://www.rfidjournal.com/the-5-cent-challenge

https://www.medrxiv.org/content/10.1101/2020.05.31.20115196v1.full.pdf

https://www.medrxiv.org/content/10.1101/2020.05.31.20115154v1.full.pdf

https://www.wired.com/story/could-breathalyzers-make-covid-testing-quicker-and-easier


doi:10.1021/acs.nano.0c05657. https://pubs.acs.org/doi/pdf/10.1021/acs.nano.0c05657

https://pubs.acs.org/doi/pdf/10.1021/acs.nanolett.5b03052


https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4548194/pdf/srep13368.pdf


SARS-CoV-2 (COVID-19) Detection Using the Breath Analyzer TeraSystem

https://clinicaltrials.gov/ct2/show/NCT04497610

Parinaz Fozouni, Sungmin Son, María Díaz de León Derby, Gavin J. Knott, Carley N. Gray, Michael V. D’Ambrosio, Chunyu Zhao, Neil A. Switz, G. Renuka Kumar, Stephanie I. Stephens, Daniela Boehm, Chia-Lin Tsou, Jeffrey Shu, Abdul Buuiya, Max Armstrong, Andrew Harris, Jeannette M. Osterloh, Anke Meyer-Franke, Charles Langelier, Katherine S. Pollard, Emily D. Crawford, Andreas S. Puschnik, Maira Phelps, Amy Kistler, Joseph L. DeRisi, Jennifer A. Doudna, Daniel A. Fletcher and Melanie Ott (2020) *Direct detection of SARS-CoV-2 using CRISPR-Cas13a and a mobile phone*. Published 30 Sep 2020

https://doi.org/10.1101/2020.09.28.20201947


“The Nobel Prize in Chemistry 2020.”


https://ia800904.us.archive.org/9/items/p1decayrestorati00schwuo/p1decayrestorati00schwuo.pdf


https://jamanetwork.com/journals/jama/fullarticle/2771764

240 [https://www.worldometers.info/gdp/gdp-by-country](https://www.worldometers.info/gdp/gdp-by-country)


https://engineering.purdue.edu/~landgreb/whitepaper.pdf
http://cba.mit.edu/events/01.11.retreat/Butera/Butera_files/v3_document.htm

266 Raffi C. Krikorian (2004) Internet 0. Submitted to the Program in Media Arts & Sciences, School of Architecture and Planning in partial fulfillment of the requirements for the degree of Master of Science in Media Arts & Sciences at the Massachusetts Institute Of Technology, August 2004 https://dspace.mit.edu/bitstream/handle/1721.1/28866/60412683-MIT.pdf?sequence=2&isAllowed=y


268 Hiroshi Ishi https://tangible.media.mit.edu/person/hiroshi-ishii/


272 Origins of IoT https://autoid.mit.edu/


274 https://autoid.mit.edu/people-2

275 https://openlearning.mit.edu/about/our-team/sanjay-sarma


283 https://www.iotone.com/term/decreasing-cost-of-storage/t172

284 Golubchik, Leana, Khuller, Samir, Mukherjee, Koyel and Yao, Yuan. (2013). To send or not to send: Reducing the cost of data transmission. IEEE Infocom. 2472-2478. Doi: 10.1109/INFCOM.2013.6567053


https://pdfs.semanticscholar.org/88b4/a255082d91b3c88261976c85a24f2f92c5c3.pdf
https://ptolemy.berkeley.edu/books/leeseshia/releases/LeeSeshia_DigitalV2_2.pdf
https://ieeexplore.ieee.org/document/5523280
https://steinhardtapps.es.its.nyu.edu/create/courses/2015/reading/Horn.pdf
https://epdf.pub/queue/information-design.html
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6042346/pdf/12936_2018_Article_2408.pdf
https://www.wired.com/2016/05/project-ara-lives-googles-modular-phone-is-ready/
299 NK Labs, Cambridge, Massachusetts https://www.nklabs.com/ara-knaian


Glucosense Monitoring. University of Leeds, UK https://www.youtube.com/watch?v=j--utQE9Pz8


Datta, Shoumen. APPENDIX Figure 1: Description of Major Software Components [https://github.com/shoumendatta/ADD-DIGITAL](https://github.com/shoumendatta/ADD-DIGITAL) and James Lamb [https://github.com/jameslamb](https://github.com/jameslamb)


Amazon S3 Glacier [https://aws.amazon.com/glacier/](https://aws.amazon.com/glacier/)


Common Object Request Broker Architecture (CORBA) https://www.corba.org/

Container: Docker https://www.docker.com/resources/what-container
Developer’s language of choice may include (but the “menu” is not limited to):

Java jar (https://en.wikipedia.org/wiki/JAR_(file_format)),
Python pickle file (https://docs.python.org/3/library/pickle.html),
R rds file (https://stat.ethz.ch/R-manual/R-devel/library/base/html/readRDS.html) or a precompiled executable which can read in input data from “stdin” (standard input is a stream from which a program reads its input data) or from a file, created with C/C++ or language-agnostic description of a model: Predictive Model ML (https://en.wikipedia.org/wiki/Predictive_Model_Markup_Language) or Portable Format for Analytics (http://dmg.org/pfa/)


https://www.cell.com/action/showPdf?pii=S0092-8674%2800%290206-8

https://www.pnas.org/content/pnas/36/6/344.full.pdf

“The Nobel Prize in Physiology or Medicine 1983.” NobelPrize.Org 

“Happy Birthday, Barbara McClintock.” Wired. 
https://www.wired.com/2012/06/happy-birthday-barbara-mcclintock/


438 https://www.fda.gov/media/137564/download


https://science.sciencemag.org/content/sci/370/6521/eabe9403.full.pdf


https://www.thelancet.com/action/showPdf?pii=S2589-7500%2820%2930142-4


https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9208795

462 https://coronavirus.jhu.edu/map.html

https://www.annualreviews.org/doi/pdf/10.1146/annurev.genom.2.1.177

https://www.cell.com/action/showPdf?pii=S1074-7613%2805%2900211-6

https://www.futuremedicine.com/doi/pdfplus/10.2217/fvl-2020-0124

https://science.sciencemag.org/content/359/6382/1355/tab-pdf


https://www.nature.com/articles/nrd4278.pdf


---

104 • ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877
https://doi.org/10.1016/j.vaccine.2007.07.027


https://patents.justia.com/patent/20190330634


480 Collected papers related to progress in aptamers with respect to stringency of binding assays (kinetics) and potential for use as therapeutics. https://bit.ly/APTAMER-PROGRESS


https://www.biorxiv.org/content/10.1101/2020.08.13.248211v1.full.pdf


501 https://somalogic.com/


531 Dialogues of Alfred North Whitehead as recorded by Lucien Price. Little, Brown & Co. Jan 1, 1954
543 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3175338/pdf/IRV-6-71.pdf
549 http://www.tulane.edu/~biochem/lecture/gbch701/kosland2.pdf
https://science.sciencemag.org/content/371/6525/160/tab-pdf


http://www.mcb5068.wustl.edu/MCB/Lecturers/Baranski/Articles/RegressionBook.pdf

547 Machine Learning for Pharmaceutical Discovery and Synthesis. MIT https://mlpds.mit.edu/papers/


552 Simulation Tools www.ansys.com/products/3d-design


https://doi.org/10.1038/ncomms13276 and https://www.nature.com/articles/ncomms13276.pdf


---

111 - ADD is in the MIT Library https://dspace.mit.edu/handle/1721.1/128017 and ChemRxiv https://doi.org/10.26434/chemrxiv.13102877


Lars Kai Hansen and Peter Salamon (1990) Neural Network Ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No. 10, October 1990. [https://pdfs.semanticscholar.org/257d/c8ae2a8353bb2e86c1b7186e7d989fb433d3.pdf](https://pdfs.semanticscholar.org/257d/c8ae2a8353bb2e86c1b7186e7d989fb433d3.pdf)


Deeks, Jonathan J., and Angela E. Raffle. “Lateral Flow Tests Cannot Rule out SARS-CoV-2 Infection.” British Medical Journal [https://doi.org/10.1136/bmj.m4787](https://doi.org/10.1136/bmj.m4787) (Published 11 December 2020) [https://www.bmj.com/content/bmj/371/bmj.m4787.full.pdf](https://www.bmj.com/content/bmj/371/bmj.m4787.full.pdf)


https://www.nature.com/articles/s41588-020-00759-x.pdf


Cartoon for Reference #371 (A more complete discussion may be found in “HIP” – https://bit.ly/EXPLORE-HIP)

Machine Learning Algorithms – ML makes better sense without the misnomer of AI since there is no canonical ‘intelligence’ in ‘artificial intelligence’. AI was erroneously named due to historical accidents. At best AI may be referred to as artificial reasoning tools (ART) but there is nothing about reasoning that is “artificial” because the logic and rules in any analytics approach or analytical technique is programmed by humans in the loop. To be mathematically correct, analysis is a term applied to calculus and all higher mathematics that uses calculus. Logic, rules and reasoning tools (LRRT) devoid of calculus is likely to be hand-waving subjectivity of limited value, if at all. Similarly, machine “learning” also takes “artistic” license by creating the illusion that machines are learning, when in reality machines are applying stored logic and rules, programmed by humans, to data and information that is supplied by the structures created by humans in the loop. Neither ML nor AI creates anything new or novel but uses programmed logic and rules in all possible and “allowed” permutations and combinations to data. There aren’t any “magic inside the black box” but “machinery” which supplies correlations using correspondence rules (provided by equations) that govern the function. There is no “intelligence” or anything “artificial” because the machinery is the relation between variables determined by functions. Function is a relation between two variables which maps to values given by domain, range, Cartesian coordinates (x,y) or polar coordinates (r,θ). Values or sets of values and limits are deduced, derived, formulated and programmed by humans (algorithms) at the heart of the engine in any learning machinery. Much to the chagrin of buzz-word peddlers (consulting firms) and marketing teams (“sound bite” manufacturers), the purpose of percolating the term “AI” is to deliberately distract us from facts and truth in order to catalyze collusive strategies for snake-oil sales. ML is tolerable but presents illusions of grandeur when learning refers to a mathematically informed ensemble of logic, rules and reasoning tools (LRRT). Pedantically speaking “LRRT” are machinery applications of logic (LO), rules and reasoning (RE) tools (TO). LRRT or LORETO are not glib and smug acronyms or sound-bites but conveys the unvarnished concept. In a world where polishing the chrome is valued higher than tuning the engine, any effort to reduce bias and increase credibility (what is truly deliverable), is an exercise in futility by a scripturient fool (referring to SD).
Katalin Karikó spent the 1990s collecting rejections. Her work, attempting to harness the power of mRNA to fight disease, was too far-fetched for government grants, corporate funding, and even support from her own colleagues. By 1995, after six years on the faculty at the University of Pennsylvania, Karikó got demoted. She had been on the path to full professorship, but with no money coming in to support her work on mRNA, her bosses succumbed to myopia.

In time, those better experiments came together. After a decade of trial and error, Kariko and her longtime collaborator at Penn — Drew Weissman, an immunologist with a medical degree and Ph.D. from Boston University — discovered a remedy for mRNA's Achilles’ heel.

“We have to be willing to embrace unreasonable propositions and unreasonable people in order to make extraordinary findings, because the notion that utterly reasonable things will produce massive breakthroughs, it just doesn’t compute to me.”

— Noubar Afeyan, CWT, Episode 113