

Open camera or QR reader and
scan code to access this article
and other resources online.



COSPAR Sample Safety Assessment Framework (SSAF)

Gerhard Kminek,¹ James N. Benardini,² Frank E. Brenker,³ Timothy Brooks,⁴ Aaron S. Burton,⁵
Suresh Dhaniyala,⁶ Jason P. Dworkin,⁷ Jeffrey L. Fortman,⁸ Mihaela Glamoclija,⁹ Monica M. Grady,¹⁰
Heather V. Graham,¹¹ Junichi Haruyama,¹² Thomas L. Kieft,¹³ Marion Koopmans,¹⁴ Francis M. McCubbin,⁵
Michael A. Meyer,¹⁵ Christian Mustin,¹⁶ Tullis C. Onstott,¹⁷ Neil Pearce,¹⁸ Lisa M. Pratt,¹⁹
Mark A. Sephton,²⁰ Sandra Siljeström,²¹ Haruna Sugahara,²² Shino Suzuki,²²
Yohey Suzuki,²³ Mark van Zuilen,^{24,25} and Michel Viso²⁶

Abstract

The Committee on Space Research (COSPAR) Sample Safety Assessment Framework (SSAF) has been developed by a COSPAR appointed Working Group. The objective of the sample safety assessment would be to evaluate whether samples returned from Mars could be harmful for Earth's systems (*e.g.*, environment, biosphere, geochemical cycles). During the Working Group's deliberations, it became clear that a comprehensive assessment to predict the effects of introducing life in new environments or ecologies is difficult and practically impossible, even for terrestrial life and certainly more so for unknown extraterrestrial life. To manage expectations, the scope of the SSAF was adjusted to evaluate only whether the presence of martian life can be excluded in samples returned from Mars. If the presence of martian life cannot be excluded, a Hold & Critical Review must be established to evaluate the risk management measures and decide on the next steps. The SSAF

¹European Space Agency, Mars Exploration Group, Noordwijk, The Netherlands.

²NASA Headquarters, Office of Planetary Protection, Washington, DC, USA.

³Goethe University, Department of Geoscience, Frankfurt, Germany.

⁴UK Health Security Agency, Rare & Imported Pathogens Laboratory, Salisbury, UK.

⁵NASA Johnson Space Center, Astromaterials Research and Exploration Science Division, Houston, Texas, USA.

⁶Clarkson University, Department of Mechanical and Aeronautical Engineering, Potsdam, New York, USA.

⁷NASA Goddard Space Flight Center, Solar System Exploration Division, Greenbelt, Maryland, USA.

⁸Security Programs, Engineering Biology Research Consortium, Emeryville, USA.

⁹Rutgers University, Department of Earth and Environmental Sciences, Newark, New Jersey, USA.

¹⁰The Open University, Faculty of Science, Technology, Engineering & Mathematics, Milton Keynes, UK.

¹¹NASA Goddard Space Flight Center, Astrochemistry Laboratory, Greenbelt, Maryland, USA.

¹²Japan Aerospace Exploration Agency (JAXA), Institute of Space and Astronautical Science (ISAS), Chofu, Tokyo, Japan.

¹³New Mexico Institute of Mining and Technology, Biology Department, Socorro, New Mexico, USA.

¹⁴Erasmus University Medical Centre, Department of Viroscience, Rotterdam, The Netherlands.

¹⁵NASA Headquarters, Planetary Science Division, Washington, DC, USA.

¹⁶Centre National d'Études Spatiales (CNES), Nancy, France.

¹⁷Princeton University, Department of Geosciences, Princeton, New Jersey, USA.

¹⁸London School of Hygiene & Tropical Medicine, Department of Medical Statistics, London, UK.

¹⁹Indiana University Bloomington, Earth and Atmospheric Sciences, Emeritus, Bloomington, Indiana, USA.

²⁰Imperial College London, Department of Earth Science & Engineering, London, UK.

²¹RISE, Research Institutes of Sweden, Department of Methodology, Textiles and Medical Technology, Stockholm, Sweden.

²²Japan Aerospace Exploration Agency (JAXA), Institute of Space and Astronautical Science, Sagami-hara Kanagawa, Japan.

²³University of Tokyo, Graduate School of Science, Tokyo, Japan.

²⁴Université de Paris, Institut de Physique du Globe de Paris, Paris, France.

²⁵European Institute for Marine Studies (IUEM), CNRS-UMR6538 Laboratoire Geo-Océan, Plouzané, France.

²⁶Conseiller Scientifique, Innovaxiom, France.

starts from a positive hypothesis (there is martian life in the samples), which is complementary to the null-hypothesis (there is no martian life in the samples) typically used for science. Testing the positive hypothesis includes four elements: (1) Bayesian statistics, (2) subsampling strategy, (3) test sequence, and (4) decision criteria. The test sequence capability covers self-replicating and non-self-replicating biology and biologically active molecules. Most of the investigations associated with the SSAF would need to be carried out within biological containment. The SSAF is described in sufficient detail to support planning activities for a Sample Receiving Facility (SRF) and for preparing science announcements, while at the same time acknowledging that further work is required before a detailed Sample Safety Assessment Protocol (SSAP) can be developed. The three major open issues to be addressed to optimize and implement the SSAF are (1) setting a value for the level of assurance to effectively exclude the presence of martian life in the samples, (2) carrying out an analogue test program, and (3) acquiring relevant contamination knowledge from all Mars Sample Return (MSR) flight and ground elements. Although the SSAF was developed specifically for assessing samples from Mars in the context of the currently planned NASA-ESA MSR Campaign, this framework and the basic safety approach are applicable to any other Mars sample return mission concept, with minor adjustments in the execution part related to the specific nature of the samples to be returned. The SSAF is also considered a sound basis for other COSPAR Planetary Protection Category V, restricted Earth return missions beyond Mars. It is anticipated that the SSAF will be subject to future review by the various MSR stakeholders. Key Words: Mars Sample Return—Sample Safety Assessment—Planetary Protection. *Astrobiology* 22, Supplement, S-186–S-216

1. Introduction

ANALYZING MARTIAN SAMPLES in terrestrial laboratories would advance our understanding of Mars in multiple ways that are impossible when using *in situ* missions or martian meteorites alone. Most recently, the Mars Sample Return (MSR) Science Planning Group 2 (MSPG2) produced an up-to-date status of MSR science planning (Kminek *et al.*, 2022; Meyer *et al.*, 2022).

With the expected benefits of MSR, however, come responsibilities. If life is present on Mars, then samples from Mars could be a source of extraterrestrial biological contamination for Earth. In line with Article IX of the United Nations Space Treaty (UN Space Treaty, 1966), a range of measures that are described by the Committee on Space Research (COSPAR) Policy on Planetary Protection would have to be employed to prevent undesirable consequences for Earth's systems (DeVincenzi *et al.*, 1998; COSPAR, 2021). One of these measures is to conduct a timely safety assessment of any unsterilized material from Mars. The first step to develop such a safety assessment began under the leadership of the National Aeronautics and Space Administration (NASA), with contributions from the Centre National d'Études Spatiales (CNES), in 2000 with a series of five workshops that led to a Draft Test Protocol (Rummel *et al.*, 2002). An important recommendation of this earlier work was to periodically review and update the Draft Test Protocol by taking into account new scientific findings and advances in instrumentation. As an intermediate step and response to this recommendation, NASA and the European Space Agency (ESA), in coordination with COSPAR, organized a life detection conference and workshop in 2012 to discuss the latest concepts and methods to search for life and identify relevant elements for a safety assessment (Allwood *et al.*, 2013; Kminek *et al.*, 2014).

With an increased interest in a joint NASA-ESA MSR Campaign and associated planning activities under-way (Meyer *et al.*, 2022), the need to produce an updated version of the safety assessment became evident. This is reflected in

one of the recommendations of the International Mars Architecture for the Return of Samples (iMars) Phase II Working Group (Haltigin *et al.*, 2016): “A Planetary Protection Protocol should be produced as soon as it is feasible by an international working group under the authority of COSPAR or another international body.” This need is also described with additional contextual information in the work of Rummel and Kminek (2018).

COSPAR swiftly reacted and established a Sample Safety Assessment Protocol (SSAP) Working Group in 2018. This Working Group had the mandate to review existing literature and the planned MSR Campaign architecture to produce a sample safety assessment protocol. The mandate for the SSAP Working Group specifically excluded biosafety control and management aspects, that is, sterilization of material from Mars, environmental and health monitoring, containment elements, and contingency planning. The Working Group had members covering the relevant expertise in life detection, public health, infectious diseases, physical and chemical composition of expected material from Mars, extraterrestrial sample analysis, sample curation, and statistical analysis. Additional experts have been invited to participate in specific meetings. In particular, a team from the US Centers for Disease Control and Prevention (CDC) was invited to comment on the draft and final versions of this report. Collectively, the external input has produced substantial added benefits in the SSAP Working Group's deliberations.

Toward the end of the Working Group's term, the name for our product was reconsidered. It was felt that it would be more appropriate to call this a *framework* rather than a protocol (or a draft protocol) to better represent the content. A detailed (or even draft) protocol will need to be developed once a number of open issues addressed throughout this report and summarized in Section 5 are resolved and more information about the samples is available.

Some general remarks to better understand this Sample Safety Assessment Framework (SSAF) include the following:

- For the purpose of formulating the SSAF, we considered the NASA-ESA MSR Campaign (Meyer *et al.*, 2022).
- We are using the term sterilization in a generic way to include both overkill (*i.e.*, a process with substantial margin that does not require viability testing after application but would typically render samples useless for further biochemical investigations) and inactivation (*i.e.*, a process with less margin that requires viability testing after application and would likely allow for conducting certain biochemical investigations after it has been applied).

The following sections describe the scope, structure, and content of the SSAF. Section 5 describes the key elements of the SSAF. These key elements are not independent and must be taken together with the remainder of the report for context and for additional and essential information. For elements of the SSAF the Working Group considered mandatory or very important, we use the term “must”, while for elements that only have an indirect effect (*e.g.*, making the assessment faster or using less material) we use a conditional form.

2. Objective and Scope of the SSAF

The objective of the safety assessment is to assess whether martian life is present that would pose a risk for Earth’s systems (*e.g.*, environment, biosphere, geochemical cycles) in samples intentionally returned from Mars. Traditionally, risk is defined in terms of probability of occurrence and consequences. In our case, we do not know and could only speculate about the consequences of releasing potential martian biology on Earth. For the purpose of the SSAF, we use the term *risk* not in relation to consequences but to the release of active martian biology exclusively. The associated risk mitigation is based on two pillars: performing a safety assessment and/or sterilizing the material from Mars.

One of the assumptions we use in terms of potential martian biology is that it is based on carbon chemistry. The likelihood that extraterrestrial life is carbon based has been suggested and discussed in various publications with arguments focused on the versatility and abundance of carbon in our Solar System and beyond (*e.g.*, NRC, 2002; Allwood *et al.*, 2012; Craven *et al.*, 2021). It is worth noting that even theoretical concepts of silicon-based life still employ organic moieties (Petokowski *et al.*, 2020). Another assumption used is that any potential life on Mars utilizes soluble organic compounds. Organic molecules used by terrestrial life are soluble in either polar or nonpolar solvents. Any potential life based on insoluble organic molecules would be unlikely to cause harm to terrestrial systems (Dirk and Irwin, 2005). In addition, solid-solid reactions are very slow compared to those in solution. An intractable solid may be hazardous, though without the capability of interacting in fluids it would not be able to replicate on timescales that compete with those of terrestrial biological systems.

We agree with the National Research Council (NRC) Committee on the Review of Planetary Protection Requirements for Mars Sample Return Missions that “the potential risks of large-scale effects arising from the intentional return of martian materials to Earth are primarily those associated with replicating biological entities, rather

than toxic effects attributed to microbes, their cellular structures, or extracellular products” (NRC, 2009). In addition to replicating biological entities, we consider it prudent to include biologically active molecules in the sample safety assessment (ESF, 2012; Craven *et al.*, 2021). This expansion of the SSAF covers the incorporation of potential martian non-self-replicating biological agents that could lead to a redirection of life processes on Earth (*i.e.*, virus-like, stray RNA or DNA-like, and prion-like entities) and even theoretical concepts of propagating catalytic reactions that may directly precede *de novo* life. Although it might be easier to find life that is the producer or host for such agents, the Mars returned sample safety assessment must have the capability to detect biologically active molecules independently as well. Throughout the SSAF, we use the term martian life to include both *de facto* martian life and biologically active molecules produced by martian life. There is also a very real possibility that there are toxic compounds present in the samples, for example, inorganic species such as perchlorates. Toxic effects that originate from the samples are not covered in the SSAF because they are limited to an occupational hazard and can be managed accordingly.

In the case that martian life is found in samples returned from Mars, large-scale negative effects on Earth’s systems are not expected (NRC, 1997; NRC, 2009; ESF, 2012). However, it is impossible to exclude absolutely such consequences. Thus, a prudent and conservative approach is the most appropriate response—be ready for the unexpected (NRC, 1997; NRC, 2009; ESF, 2012).

There are many ways an alien life form could be harmful to Earth’s systems. The possible interactions could include not only direct effects on humans, animals, and plants or their associated beneficial microbes, but also indirect effects of a competitive interaction with various terrestrial species. Examples abound of the detrimental effects that result from terrestrial invasive species being introduced into new environments (*e.g.*, van der Putten *et al.*, 2007; Litchman, 2010; Randolph and Rogers, 2010). There are also more subtle effects that can be imagined. Some microorganisms, while not obviously beneficial to humans, are actually keystone species, the loss of which could cause irrevocable harm to an ecosystem (*e.g.*, Mills *et al.*, 1993) or disrupt essential biochemical cycles (*e.g.*, Jardillier *et al.*, 2010). Unfortunately, we have only a limited ability to predict the effects of terrestrial invasive species, emerging pathogens, and uncultivated microbes on Earths’ ecosystems and environments. This is true even for cultured and fully genome-sequenced terrestrial organisms and more so for potential extraterrestrial life. Thus, conducting a comprehensive sample safety assessment with the required rigor to predict harmful or harmless consequences of potential martian life for Earth is currently not feasible. This situation is not likely to change substantially within the next decade. On the contrary, the increased knowledge accumulated over the last decades has shown many more unexpected effects and dependencies in the various ecosystems of Earth (*e.g.*, Pejchar and Mooney, 2009). Therefore, the scope of the SSAF is limited to evaluating whether the presence of martian life can be excluded in the samples without pretending to assess the potentially hazardous nature of the samples—except that if there is no life, there is no biological hazard. This position is in line with the NRC Committee on Mars Sample Return

Issues and Recommendations: “Evaluation of the sample for potential hazards should focus exclusively, then, on searching for evidence of living organisms, their resting states (*e.g.*, spores or cysts), or their remains in the sample” (NRC, 1997).

Although this approach might lead to an impression that the SSAF is essentially a life detection framework, this impression would be incorrect. There are very important and clear distinctions between the general search for martian life in returned samples for purely scientific purposes and the assessment to exclude the presence of martian life in them. The SSAF is starting from the positive hypothesis—“there is martian life in the samples.” Testing this hypothesis, that is, excluding the presence of martian life, is complementary to the scientific objective to search for martian life. Science investigations and the sample safety assessment use the same scientific methodologies, though the purpose and associated burden of proof is reversed. Disproving either the positive (safety) or null (science) hypotheses to a certain level of confidence can only be accomplished by collecting sufficient statistical data. Meeting the objective to disprove the null scientific hypothesis is typically constrained by the available resources in terms of budget and time. The constraints of disproving the positive safety-relevant hypothesis is much less dependent on resources and more dependent on the acceptable risk or the acceptable level of assurance that a risk will be avoided. As a consequence, the search for martian life science objective will benefit from the increased rigor required by the safety assessment, given that it will utilize the same scientific methods and tests required to address the search for martian life objectives. Thus, all samples used for the safety assessment and all tests done on these samples will have a scientific value.

To emphasize again, the SSAF is not a life detection framework. There are life detection frameworks in discussion and under development (*e.g.*, Green *et al.*, 2021; Graham *et al.*, 2021). These are timely efforts to assess the validity and confidence for evidence of extraterrestrial life and ways to communicate this information effectively. Finding evidence for life typically follows an incremental path until definitive evidence is reached by a consensus in the scientific community (Green *et al.*, 2021). Due to the reversed burden of proof for the safety assessment, any ambiguous results (*e.g.*, maybe abiotic, maybe terrestrial contamination, maybe masking martian life) would not disprove the positive hypothesis until a clear root cause is identified and confirmed. Any step toward an agreed upon framework for life detection established by the science community would certainly help to reduce some uncertainties in the safety assessment and is therefore encouraged.

The following principles, derived from a Life Detection Conference & Workshop (Allwood *et al.*, 2013; Kminek *et al.*, 2014), reflect the interplay of science and sample safety assessments and provide the basis for the SSAF:

1. Use of a hypothesis-driven approach in the development of life detection investigation strategies and measurements for science (null hypothesis) and sample safety assessment (positive hypothesis).
2. The same types of scientific measurements inform the scientific understanding of the samples and their safety assessment.

3. A sample safety assessment must be data-driven, *i.e.*, responsive to the results of individual or combined investigations.
4. The distinction between the scientific objective to search for martian life and the sample safety assessment is mainly the degree of rigor and supervision applied, which is described in this framework.

Unlike a scientific objective to search for life on Mars, the scope of the SSAF is limited to exclude the presence of martian life in the samples from Mars. Taking into account the diversity of samples and the microscopic distribution of potential life in macroscopic samples, every sample tube is considered a separate sample. A negative result (*i.e.*, no martian life) associated with samples from a sample tube would provide a certain pre-defined level of assurance that there is no life and therefore no hazard for Earth in that sample tube. Such a determination cannot be extrapolated to other sample tubes nor can it be extrapolated to the planet Mars. A positive result for one or more samples would not necessarily mean they are hazardous for Earth. Any positive result would lead to a Hold & Critical Review (see Section 3.4). A deeper understanding of how any newly discovered biology works and what kind of capabilities it has would require detailed understanding of the metabolism, informational macromolecules, and replication of this extraterrestrial life. As on Earth, it is unlikely that life is represented only by one of its members, that is, if we discover a single martian life form, we would possibly discover more than one member of a martian biology. This, together with the fact that we do not even know how to cultivate most terrestrial microorganisms makes it essential to manage expectations in terms of the possibility to conduct a proper hazard assessment. This aspect is further detailed in the implementation part of the SSAF (Section 4).

There is one open parameter that must be introduced to the sample safety assessment—the level of assurance required to declare a sample safe. This parameter would describe the stopping threshold, that is, level of confidence in the statement “the presence of martian life is excluded in this sample.” Setting such a level is important to avoid open-ended discussions and better estimate the efforts and resources necessary to conduct the sample safety assessment. For the purpose of running simulations and test cases, we have taken a value of “1 in a million chance of failing to detect life if it is there.” For details on the background of this canonical value, the reader is advised to consult the ESF Study Group Report on MSR Planetary Protection Requirements (ESF, 2012).

3. Elements of the SSAF

There are four elements in the SSAF. Each is necessary, though on its own not sufficient to qualify for a safety assessment. The four elements are (Fig. 1):

1. Bayesian statistics;
2. Subsampling strategy;
3. Test sequence;
4. Decision criteria.

These elements will be described in the next few sections. The test sequence section includes a description of the types

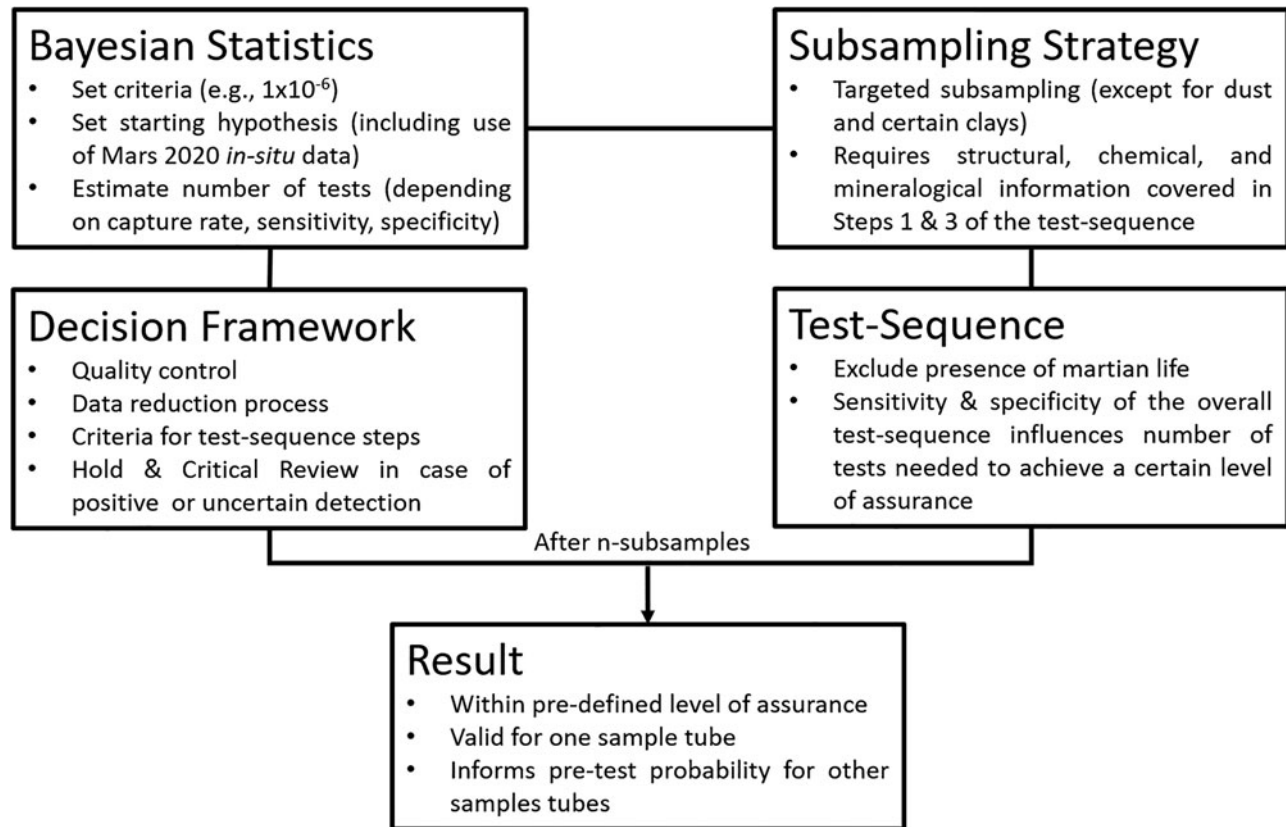


FIG. 1. The four elements of the SSAF. There are multiple interdependencies between the various elements. The major external input parameter required is the level of assurance that something is safe. Some of the parameters need informed judgements based on Mars 2020 *in situ* data and tailored analogue test programs.

of techniques or instruments that could provide the information required by the SSAF. This list of candidate instruments is not a set of required or endorsed instruments but has been established for planning purposes.

3.1. Bayesian statistics

Bayesian reasoning and methods of statistical analysis are standard approaches with which to address complex statistical issues (Greenland, 2021) and are widely used in medical decision making (Hunink *et al.*, 2014). Bayesian statistics can accommodate various forms of information and help to optimize limited resources, like sample material or time. Therefore, Bayesian statistics is considered an appropriate tool for the SSAF.

When little prior information is available, the Bayesian and frequentist statistics will generally yield very similar results (Rothman and Lash, 2021). When prior knowledge can be incorporated, whether for decision making in medicine or assessments of Mars samples, and a series of tests are to be used with the results being updated after each test, Bayesian statistics is more applicable and appropriate (Hunink *et al.*, 2014; Greenland, 2021). In our case, it is necessary to specify an *a priori* probability that there is martian life in a sample tube. The information acquired by the NASA Mars 2020 mission (Farley *et al.*, 2020) can be used to make an informed judgement about the *a priori* probability of finding martian life in a sample tube before actually starting any testing. This informed judgement must reflect the conservative posture of a

positive hypothesis. The results of applying tests on one sample tube can also inform, together with the other Mars 2020 information, the *a priori* probability of finding martian life in subsequent sample tubes. Recall, however, that a sample safety determination cannot be directly extrapolated from one sample tube to another one.

3.1.1. Sensitivity and specificity. In addition to establishing a pre-test probability (*a priori*), the other quantities that need to be estimated before Bayesian statistics can be applied are the sensitivity (S_n) and specificity (S_p) of the test. One complication is that terrestrial biological contamination would impact the specificity of the test, that is, leading to a false positive. There is also another complication: even if there is life somewhere in the sample tube, there is no guarantee that there will be life in the subsamples that are examined. Thus, the sensitivity of the test for the sample of a specific sample tube depends on both the sensitivity of the test and the capture rate, that is, the probability of finding martian life in a subsample, if there is in fact martian life somewhere in the sample material inside the sample tube, which is certainly less than 1.0 (*i.e.*, less than 100%). The effective sensitivity (ES_n) is the product of the sensitivity and the capture rate.

3.1.2. Driving factors for the safety assessment. If the presence of martian life in one of the subsamples cannot be excluded, then in terms of the safety assessment, we need to assume that there is a high probability that life is present.

Note that this conservative approach, taking the precautionary principle into account (Pearce, 2004), assumes that the specificity of the test is 1 (i.e., 100%), i.e. that a false positive cannot occur. Although the test would certainly be chosen to minimize the chance of a false positive, there is still the issue of terrestrial biological contamination that could at least bias the results (see Section 4.3 for more details). With this background information, how many negative test results are actually required before it can be concluded that the positive hypothesis (i.e., that there is martian life in the sample tube) has been “refuted” (i.e., the probability of life being present is less than a pre-defined level of assurance)? A theoretical example can illustrate this, including the dependency of the required number of negative tests on the various parameters. More information about the relationship of samples and subsamples is described in Section 3.2.

Definitions:
Pre-test (a priori) probability: the pre-test probability that a sample tube contains martian life
Capture rate (Cr): the probability that a subsample contains martian life, given that life is somewhere in the sample tube
Sensitivity (Sn) or True Positive Rate (TPR): true positives/(true positives + false negatives); the probability that the test sequence will detect martian life in a particular subsample, given that there is life in the subsample
Effective Sensitivity (ESn): Sensitivity corrected for the capture rate
Specificity (Sp) or True Negative Rate (TNR): true negatives/(true negatives + false positives); the probability that the test sequence will not detect martian life in a particular subsample, given that there is no life in the subsample
False Positive Rate (FPR): 1 - Sp
False Negative Rate (FNR): 1 - ESn
Positive likelihood ratio (PLR): ESn/(1 - Sp)
Negative Likelihood Ratio (NLR): (1 - ESn)/Sp
Pre-Test Odds: Pre-Test Probability/(1 - Pre-Test Probability)
Post Negative Test Odds: Pre-Test Odds x NLR
Post Negative Test Probability: Post Negative Test Odds/(1 + Post Negative Test Odds)

Using the following assumptions:

Pre-test probability = 0.50

Sensitivity = 0.99

Capture rate = 0.75

Specificity = 0.99

The following can be derived:

Effective sensitivity = $0.99 \times 0.75 = 0.7425$

Pre-test odds = $0.50 / (1 - 0.50) = 1.00$

Positive Likelihood Ratio (PLR) = $0.7425 / 0.01 = 74.25$

Negative Likelihood Ratio (NLR) = $(1 - 0.7425) / 0.99 = 0.26$

If then the test is applied to the first subsample, and the results are negative, then the post-test probability can be calculated as follows:

Post-negative-test odds = pre-test odds \times NLR = $1.00 \times 0.26 = 0.26$

Post-negative-test probability = $0.26 / (1 + 0.26) = 0.206$

Having one negative test reduces the probability that there is life in the sample from 0.50 (pre-test) to 0.206 (post-test); this value is now used as pre-test probability for a second test; a second negative test reduces the probability further to 0.063, etc.

Table 1 shows the number of sequential negative tests required for the post-test probability to become less than 1×10^{-6} (i.e., 1 in a million), under a variety of assumptions.

This exercise shows two important results (Table 1):

- The capture rate is crucial to this process. With a capture rate of only 0.25 (25%), a pre-test probability of 0.95 (95%) and both sensitivity and specificity at 0.95 (95%), it would require at least 77 negative tests (and no positive test) before one could conclude that the post-test probability is less than 1×10^{-6} . On the other hand, if the capture rate is 0.75 (75%), then only 15 negative tests (and no positive test) are required. For an unrealistic capture rate of 1 (100%), only 6 tests would be required.
- The sensitivity and specificity of the overall test sequence is important as well but to a lesser extent than the capture rate. Only when values go much below 0.9 (e.g., 0.7) would this markedly increase the number of negative tests required before one could conclude that the post-test probability is less than 1×10^{-6} .

The capture rate for a natural sample and the sensitivity and specificity of a real test sequence can only be estimated by an informed judgement. The elements that are necessary to enable such an informed judgement are described in Sections 3.2 and 4.2.

3.2. Subsampling strategy

As described in the previous section, the capture rate has a major impact on the number of negative tests required before the samples from a sample tube can be declared safe with a pre-defined level of assurance. The number of negative tests (as defined in Section 3.3) required is equivalent to the number of subsamples of a sample in a sample tube that need to be tested. As some parts of natural samples are

TABLE 1. DEPENDENCIES OF THE NUMBER OF NEGATIVE TESTS (I.E., NUMBER OF INDEPENDENT SUBSAMPLES) NECESSARY FOR THE SAMPLE OF ONE SAMPLE TUBE TO REACH A POST-TEST PROBABILITY FOR LIFE IN THE SAMPLE OF LESS THAN 1×10^{-6}

Pre-test Probability	Capture Rate	Sensitivity	Specificity	Number of Tests
0.5	0.25	0.95	0.95	63
	0.5	0.95	0.95	24
	0.75	0.95	0.95	12
		0.70	0.70	36
0.95	0.25	0.95	0.95	77
	0.5	0.95	0.95	29
	0.75	0.95	0.95	15
		0.70	0.70	44

more likely to contain life than the rest (*e.g.*, Onstott *et al.*, 2019), it is typically not appropriate to apply random sampling. To maximize the probability that a subsample will contain martian life (*i.e.*, increase the capture rate) if there is martian life somewhere in the sample of a sample tube, informed, targeted sampling is required. Random sampling or a poorly informed targeted sampling will reduce the effective sensitivity and thus would lead to a substantially higher number of negative tests (*i.e.*, number of subsamples) required before a sample from a sample tube could be declared safe with a pre-defined level of assurance. This is illustrated in Table 1: a poor capture rate of 0.25 (25%) compared to a good one of 0.75 (75%) would require more than 60 additional subsamples to be processed (each with a negative result) before reaching the same level of assurance. Thus, an informed targeted sampling strategy needs to be applied to reach capture rate levels, ideally, above 0.5 (50%). Such a strategy requires a focus on the areas, characteristics, and features of the samples that are likely to contain martian life, taking into account the type of sample and the expected distribution and patchiness of life in the samples associated with fractures, veins, and general interconnected spaces as well as chemical interfaces and boundaries (*e.g.*, Gorbushina, 2007; Cockell *et al.*, 2019; Onstott *et al.*, 2019; Brady *et al.*, 2020; Suzuki *et al.*, 2020). Many sample tubes will undoubtedly contain samples with diverse features that, depending on their categorization, could number from a few to a large number of distinct sections of each sample. However, it should be noted that such targeted sampling needs to be balanced, for example, by containing an appropriate mix of high-probability and medium-probability sites, rather than solely sampling from those sites with features that are considered to have the highest probabilities of containing life.

The first step in this process is to obtain information about the 3-dimensional (3-D) morphological characteristics of the external and internal structures of the sample at a micrometer scale (*i.e.*, 1×10^{-6} meter) resolution. Though this spatial resolution is not necessarily sufficient to find morphological evidence of life, it is sufficient to image physical features that could contain such information (*e.g.*, fractures, veins, and general interconnected spaces). To select optimal targets and establish priorities for subsampling, spatially correlated chemical and mineralogical information is required as well (*e.g.*, Onstott *et al.*, 2019).

Airfall or windblown dust is a special case in this sampling context. Although dust might be sorted to a certain degree during sample acquisition and transport, random sampling for dust samples is an adequate approach as long as the dust samples are homogenized (*i.e.*, mixed) before subsamples are taken. It is worth noting that the serendipitous dust on the sample tubes is likely not of sufficient quantity to perform a sample safety assessment. This would be the case as well for any dust components inside the various sample tubes; hence, it is questionable whether small quantities of dust can be declared safe (*i.e.*, free of martian life) based on a sample safety assessment. Clays are another special case. There are clays formed by local aqueous alteration (*e.g.*, smectite coatings on weathered feldspars), and there may be clay-rich mudrocks that are typically homogenous in terms of the distribution of the clays. For determining the right subsampling approach, in-

formed targeted subsampling is appropriate for clays formed by local alteration associated with distinct features (*e.g.*, fractures) within lithified clay rocks. By contrast, clay-rich “muds” would be more suited for random subsampling. This informed approach can be generalized for various types of fine-grained minerals, that is, targeted subsampling for localized fine-grained alteration products or localized features, such as fractures in lithified fine-grained rocks, and random subsampling for unconsolidated fine-grained sediments. These approaches must be tested and confirmed by using terrestrial analogue material (see Section 4.2).

A further consideration in selecting subsamples is that each subsample must be independent (conditional on the targeted sampling strategy), that is, each subsample needs to be from different parts of the sample. If this is not done, for example, if all subsamples are selected from the same section of the same crack, the subsamples would not be independent, and the assumptions of the Bayesian analysis would no longer be valid (and neither would the assumptions of standard “frequentist” statistics). The independence of sampling is not an issue for dust samples since these are assumed to be homogenous.

The information about the sample, however, is only one element in developing a credible and robust targeted sampling strategy. The specific martian sample information must be linked with a knowledge base, that is, experience with similar terrestrial sample types, including dust samples. To establish such a knowledge base requires an analogue test program tailored to the expected types of samples from Mars (*i.e.*, information from Mars 2020) and the kind of measurements that will be used to establish the information for deriving the capture rate (*i.e.*, 3-D structural information and the spatially associated chemistry and mineralogy). For more information see Section 4.2.

Bayesian statistics provide an estimate of the number of subsamples necessary to reach a pre-defined level of assurance that the sample in a sample tube is safe. This is a very important aspect of the sample safety assessment because it facilitates planning with regard to the resources (*e.g.*, time, number of subsamples) required for individual sample tubes. It also helps to establish a strategy that optimizes the sequence of investigations required for analyzing the available sample tubes. The amount of material for each subsample depends on the sensitivity of the test in relation to the resolution required for the sample safety assessment. Thus, any available technique that has been properly vetted and meets or exceeds the measurement requirements should be considered for use.

3.3. Test sequence

In the previous sections about Bayesian statistics and the subsampling strategy, the term “test” is used in a very generic form. Unfortunately, there is no single “test” that can be applied to acquire all of the information necessary to perform a sample safety assessment. What is actually required is a set of investigations in a specific logical order that will inform the sample safety assessment. This set of investigations, referred to specifically as the “test sequence,” is focused on the type of information necessary for sample safety assessment rather than the specific techniques and instruments required to gather it. Any statements made about a “test” in the previous sections (*e.g.*, number of

negative tests, dependencies of tests, etc.) are actually applicable to the test sequence. For example, the number of negative tests required to reach a pre-defined level of assurance that there is no martian life in a sample tube is actually the number of negative test sequences applied on the equivalent number of subsamples, that is, 3 negative tests means that the test sequence was applied on 3 different subsamples, each time with a negative overall result. It should be emphasized that all of these tests will produce scientifically important data.

3.3.1. Test sequence steps. The first investigations in the test sequence are derived directly from the need to support the targeted subsampling (see Section 3.2).

Three-dimensional structural information—Step 1. This investigation is focused on the 3-dimensional (3-D) morphological characteristics of the external and internal structural features of interest at micrometer-level spatial resolution while the samples are still in the sealed sample tubes. The importance and usefulness of this kind of analysis has been previously described by McLennon *et al.* (2011) and Tait *et al.* (2021). Together with the information provided by the Mars 2020 mission (Farley *et al.*, 2020), this information will enable the assignment of a risk level (*i.e.*, chance to find martian life, pre-test probability) to each sample tube. The use of Bayesian statistics and considerations of the dependencies of the various parameters (pre-test probability, capture rate, sensitivity, specificity) makes it obvious that the samples with the highest risk index (*i.e.*, highest chance to find martian life) should be opened and analyzed first because these samples will provide the highest confidence in the results—positive or negative. They would also produce the most robust pre-test probability for any subsequent samples and altogether optimize the use of sample material. After having selected a specific sample tube for further analyses, the 3-D structural information forms the basis of a risk-informed subsampling strategy.

Gas analysis—Step 2. There are many scientific reasons for analyzing the headspace gas inside a sample tube (Swindle *et al.*, 2022; Tosca *et al.*, 2022; Velbel *et al.*, 2022). In fact, analysis of the sample tube headspace gas is likely to be one of the first direct measurements of martian material beyond the indirect investigation described in Step 1. At the most basic level, analysis of the headspace gas might inform whether a tube has leaked (Parai *et al.*, 2021). Having been sealed on Mars, the tubes will be at negative pressure relative to ambient Earth pressure, so immediate leakage will be of terrestrial atmosphere into the sample. If terrestrial atmosphere has leaked into the tubes, then it would have occurred during atmospheric entry or when the capsule was on the ground awaiting recovery. In either case, bacteria, dust, or other air-borne particulate matter may have been carried into the sample tubes as well, depending on the nature of the leak. Deposition of such matter on the martian samples has the potential to create false positives, assumed-to-be martian species, or overprint a true positive signal of martian life (Milam *et al.*, 2021). The gas analysis would be important for planning the sequence of operations for opening the individual tubes, as well as for the interpretation of the data to know early on which tubes might be com-

promised by terrestrial contamination. Knowing which tubes are compromised will also be a key factor in determining the extent to which contamination knowledge samples will be required to deconvolve any terrestrial life signals in a sample from any potential martian signals that are also present (refer to details in Section 4.3).

Chemistry and mineralogy associated with the 3-D structural information—Step 3. These investigations focus on the acquisition of information about the chemistry and mineralogy associated with 3-D structural features of interest in a sample (*e.g.*, fractures, veins, and general interconnected spaces). This could be done at the same time that 3-D structural information is acquired on samples while still in their respective sealed sample tubes (Step 1) or, subsequently, once sample material is removed from the sample tubes. The benefit of the latter approach is that the quality and spatial resolution of the chemical information acquired on sample material removed from the sample tubes might be better. Such chemical and mineralogical information is essential to refine the subsampling strategy (Tait *et al.*, 2022; Carrier *et al.*, 2022), which is based on the 3-D structural information, and of particular importance to optimize the subsequent use of sample material.

To put these first investigations in the proper context, it is useful to describe briefly the expected initial sample characterization steps in the frame of the sample curation activities (Tait *et al.*, 2022). The initial sample characterization covers three distinct phases: Pre-Basic Characterization (Pre-BC), Basic Characterization (BC), and Preliminary Examination (PE) (Fig. 2). The first investigation required in the SSAF, 3-D morphological characterization of the external and internal structures at micrometer-level spatial resolution (Step 1) while the samples are still in sealed sample tubes, overlaps with the Pre-BC investigations. Step 3—chemical and mineralogical information associated with features of interest in the sample structure—overlaps with the BC and PE investigations (Fig. 2). These overlaps are beneficial because the set of investigations serve three functions—curation, science, and sample safety assessment.

Steps 1 and 3 of the test sequence provide information about how many and which subsamples to take from the sample of a sample tube. Products and effects of life in a host rock are generally volumetrically more significant than life itself (Onstott *et al.*, 2019). Therefore, it is possible that the results of these first steps would provide initial indications of life, in addition to refining the targeted subsampling. Morphological indications consistent with life are a special case in this context. Independent of the analytical process used, morphology alone can be misleading. There is a long history of incorrect interpretations of cell-like morphologies as evidence of fossilized life (see Section 3.3.2). What actions follow, in particular for Step 4, depend upon the associated chemistry and whether any morphological feature of interest is unique and an isolated observation or a common constituent of a sample. Unlike the scientifically relevant null hypothesis, the sample safety assessment is focused on the positive. Therefore, targeted investigations for Step 4 require morphological and chemical information that exclude a martian biological origin of common and unique features in the samples rather than just attempt to confirm their potential biological origin.

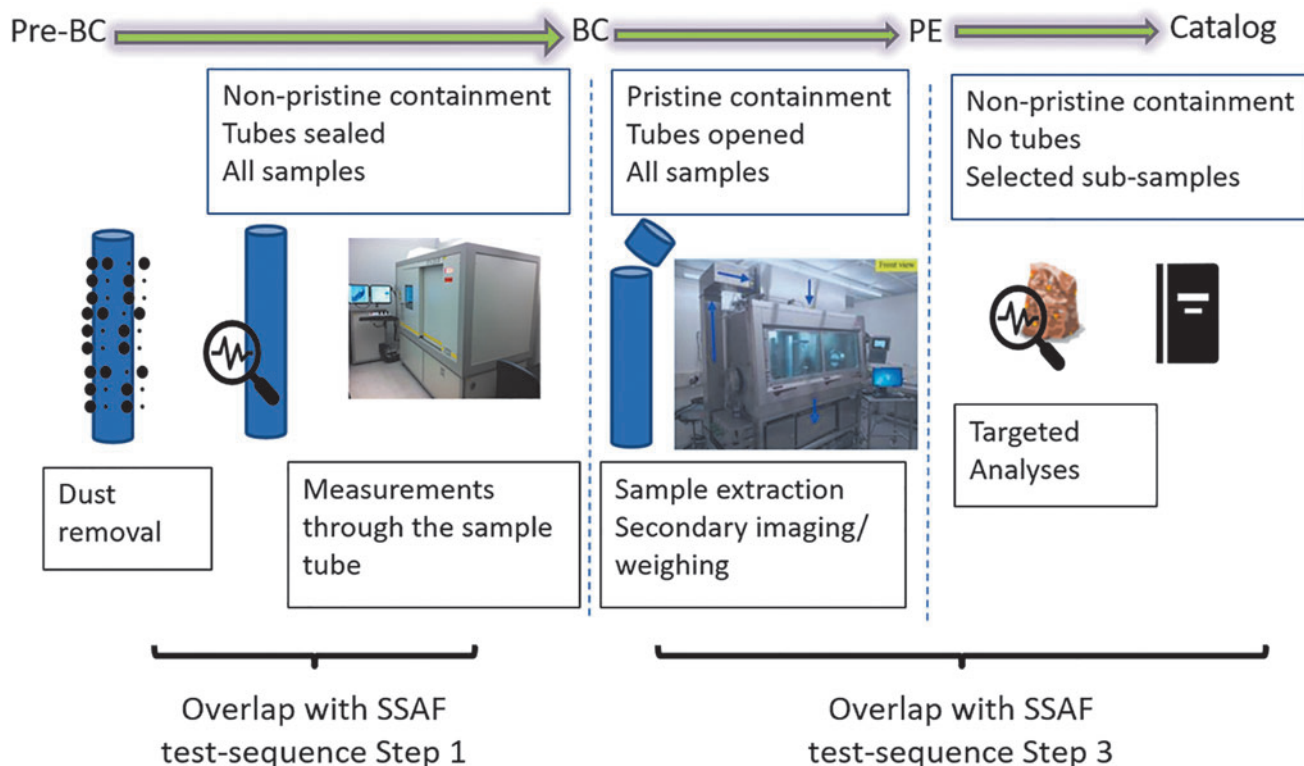


FIG. 2. Overlap of planned initial sample characterization activities for curation and the SSAF test-sequence. Graph modified from Tait *et al.* (2022).

Organic molecules—Step 4. This step initiates the search for molecular evidence of martian life in targeted subsamples. The search strategy is based on the assumption that potential martian life is based on carbon chemistry. Therefore, the subsequent investigations (Steps 4, 5, and 6) must include a focus on locating, identifying, and characterizing organic compounds in the subsamples. Steps 1 and 3 of the test sequence are critical in the search for any organic molecules that might be associated with life because, like on Earth, it is expected that life is spatially clustered and not homogeneously distributed in the host rock, and that the bulk organic content of the host rock is not necessarily correlated with the presence or absence of life (*e.g.*, Onstott *et al.*, 2019; Suzuki *et al.*, 2020).

For the purpose of the SSAF, organic molecules are defined as a group of covalently bonded molecules that contain carbon and at least one other element. We exclude CO, CO₂, CO₃²⁻, carbides, graphite, and steel from this functional definition of organic. Insoluble organic matter (IOM), as delivered by meteorites, and kerogen that originated from extinct life are also excluded because such substances consist of molecular compounds that are not soluble in polar or non-polar solvents. Examples of included organic compounds are mellitic acid, urea, CS₂, CCl₄, methane, carbon suboxide, Prussian blue, polycyclic aromatic hydrocarbons, and obviously organic species like lipids, amino acids, aldehydes, etc. To properly characterize any specific organic compounds in returned samples from Mars, it is necessary to use destructive techniques. The decision about whether to apply *in situ* techniques or bulk extraction-based techniques will require information from the previous investigations

(*i.e.*, Steps 1 and 3). *In situ* based techniques are less likely to conclusively identify any specific organic molecule because they rely on identification of only one type of information—either mass (*e.g.*, Matrix Assisted Laser Desorption/Ionization Mass spectrometry (MALDI-MS), Time of Flight Secondary Ion Mass Spectrometry (ToF-SIMS)), or functional group (*e.g.*, Raman spectroscopy, infrared spectroscopy, deep ultraviolet fluorescence). For some techniques, a substantial interference from the mineral matrix is expected. The advantage of *in situ* based techniques is that a result can be spatially associated with observed features. *In situ* analysis is the preferred approach in those cases where compelling morphological or chemical evidence of life are detected. Bulk extraction-based techniques (*e.g.*, Liquid Chromatography Mass Spectrometry (LC-MS), Gas Chromatography Mass Spectrometry (GC-MS)) provide two types of information for identification of organic molecules—the time it takes for a compound to pass through a chromatographic column (retention time), and the mass and fragmentation behavior of the molecule as measured by the mass spectrometer. Although these two types of information improve the reliability of identifying specific organic molecules, extraction-based techniques eliminate the direct spatial association with structural features of the sample and could dilute a localized low biomass signal (*i.e.*, reduce the sensitivity). However, in those cases when the evidence for possible life is more widespread in a sample, extraction-based techniques could also increase the sensitivity because they typically sample a larger volume. It is acknowledged that organic molecules occupy a wide range of polarity space, and thus no single solvent will extract all

compounds. This affects the amount of sample material that needs to be used for bulk analysis of each subsample. Further, any binding of life and organic compounds to mineral surfaces will require additional steps (such as hydrolysis) to release them (Mitra, 2004). Sample extracts from a specific subsample could be split for analyses by multiple complementary techniques. It is very important that all blanks be processed in the same manner as a sample of interest.

Molecular patterns—Step 5. If regions of organic-rich material are identified in a subsample, it is necessary to characterize the molecules present. Compound-specific measurements are required to search for molecular patterns. The targets of interest are small organic compounds, such as those found in biological monomers or biochemical intermediates, while the characterization of larger molecules is covered in Step 6. Molecular patterns are defined as a limited suite of organic compound abundances distinct from what would be produced abiotically with respect to structural diversity, chirality, and stable isotopes. For example, abiotic reactions tend to produce organic compounds at decreasing abundance with increasing molecular weight and show a lack of chemical specificity (*e.g.*, biological *vs.* meteoritic amino acid abundances or Fischer-Tropsch hydrocarbons *vs.* even-numbered biological fatty acids). With the exception of certain meteoritic compounds, molecules produced from abiotic reactions show no chiral preference (Glavin *et al.*, 2019). Glavin *et al.* (2019) provided a framework for using structural diversity, chirality, and stable isotopes together to evaluate possible biological origins of a compound, and they cautioned that any one of these indicators would be insufficient to indicate biology. Though this framework is science driven, it should be acknowledged that, for sample safety assessment purposes, the aim is to exclude biological origin. It is difficult to generate a predetermined life detection or life exclusion test from molecular patterns because of the likely co-existence of mixtures of several end member organic compounds that include those from active and prolific biology, degraded biological compounds, degraded abiotic organic compounds, and abiotic chemistry. As in the previous step, sample extracts could be split for analysis by multiple complementary techniques, and blanks must also be analyzed in parallel. Molecules most likely to be detected in this step are amino acids, nucleobases, sugars, lipids, and pigments.

Macromolecules—Step 6. The next step in the SSAF is designed to target polymeric or other large molecules to search for patterns in order to differentiate abiotic macromolecules, such as meteoritic insoluble organic material, from biological molecules, including, but not limited to, deoxyribonucleic acid (DNA), ribonucleic acid (RNA), proteins, and polysaccharides. For the purpose of the SSAF, a macromolecule is defined as an organic compound with molecular weight greater than 2500 Da (Dalton). This limit is derived from taking one half of the mass of the smallest known functional macromolecules in terrestrial biology. For example, the smallest prion is 300 kDa (Silveira *et al.*, 2005), the smallest enzyme is 66 residues, or 6811 Da (Chen *et al.*, 1992), and the smallest ribozyme is 16 nucleotides, or 5233 Da (Scott *et al.*, 1995). This limit is also smaller than

the smallest of the well-studied RNA *in vitro* replicating systems with a mass of about 15 kDa (Oehlschläger and Eigen, 1997). It should be noted that the smallest amyloid is an 8-residue domain, or 800 Da (Gazit, 2007; Sabate *et al.*, 2015), and that amyloids, transmissible epigenetic regions in a larger protein, need to be in high enough concentrations to form fibrils (Sabate *et al.*, 2015). Such a concentration of peptides would be strong evidence for life, but defining a macromolecule so broadly is likely to generate more false-positive detections than is useful. Metabolic only hypercycle-like life (Eigen and Schuster, 1997) would lack informational macromolecules but is unlikely to be able to outcompete terrestrial biology and pose a threat. Nevertheless, such a biological system would show a strong positive signal for the previous investigations but fail the current investigation step, and must be investigated further for the potential for life. Similar to the previous steps, sample extracts could be split for analysis by multiple complementary techniques, and blanks must also be analyzed in parallel. Molecules likely to be detected in this step include proteins.

Life as we know it—Step 7. Hallmarks of terrestrial life include ATGC-based DNA, AUGC-based RNA, proteins comprising 20 L-amino acids, lipids (*i.e.*, fatty acids, phospholipids, etc.), and glycopeptides, such as peptidoglycan and polysaccharides (*e.g.*, cellulose). Detecting life as we know it assumes that, if there is a living organism, it relies on the same chemical processes as terrestrial organisms and thus differs from the agnostic approach described in Step 8. To improve the sensitivity in what is expected to be a low biomass scenario requires the use of amplification steps (see Section 3.4). There are two types of life detection techniques that amplify specific targets of interest: cultivation-dependent and cultivation-independent, both with varying degrees of sensitivity and specificity (Table 2). The most useful techniques would be highly sensitive and have low specificity. Cultivation techniques theoretically have extremely high sensitivity in that one can grow a culture from a single cell, but the narrow bandwidth of any one combination of culture medium and growth condition makes culture-based approaches unpractical (see also Section 3.3.2). Alternatively, one can apply cultivation-independent techniques with amplification steps, like the polymerase chain reaction (PCR) for the amplification of nucleic acids. PCR sensitivity is high because billions of copies of a gene of interest can be derived from as little as one template copy. The usual target gene encodes small-subunit ribosomal RNA (SSU rRNA), which is a component of all terrestrial cells. PCR can also be very non-specific in that primers for SSU rRNA genes have been designed to have homology to all, or nearly all, members of each of the three evolutionary domains of life. In fact, these universal primers are routinely used to characterize microbial communities on Earth, including those in extreme environments. Sequencing of PCR-amplified SSU rRNA genes has revealed many new phyla (*i.e.*, taxonomic rank in biology) of previously unknown life as we know it (*e.g.*, Lloyd *et al.*, 2018).

If life as we know it is detected in samples from Mars, the most likely explanation for this would be contamination from a terrestrial source. Contamination can occur during the assembly of the spacecraft and proceed all the way to

TABLE 2. CURRENT TECHNIQUES FOR DETECTION OF LIFE AS WE KNOW IT

<i>Method</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Sample Quantity</i>	<i>Notes</i>
Cultivation	High	High	Small	Not recommended
Polymerase chain reaction (PCR)	High	High	Small	Degenerate or random primers could be used to detect ATGC DNA with unknown sequences. Some alternative nucleic acid base pairs (life as we don't know it) might be amplifiable by PCR.
High-throughput metagenomic sequencing of all genes	Low	Low	Large	
Proteins analyses by MS	Low	Low	Large	
Membrane phospholipid analyses	Low	Low	Large	
Fluorescence <i>in situ</i> hybridization (FISH), and variations, <i>e.g.</i> , CARD-FISH	High	High	Large	High sensitivity requires cell separation and concentration.
Single-cell amplified genome (SAG) analysis	High	Low	Large	Requires amplification by multiple displacement amplification (MDA), which is non-specific, but biased.
Flow cytometry and fluorescence-activated cell sorting (FACS)	Low	Low	Large	Requires separation of cells from a large volume of sample.
ATP assays	High	Low	Large	Chemical assay is sensitive, but requires separation and concentration of the cells from the sample
Activity assays, <i>e.g.</i> , enzyme activity, respiration assays, nano-SIMS	Low	Low	Large	Tests the ability to metabolize a substrate. Specificity depends on the substrate used.
Bio-orthogonal noncyclic amino acid tagging (BONCAT)	Low	Low	Large	
Direct microscopic counting using DAPI, acridine orange or other nucleic acid stain	Low	Low	Moderate	Direct microscopic counting using DAPI, acridine orange or other nucleic acid stain.
Viability stains: SYBR, Bac-Light Live/Dead, Bac-Light Redox Sensor, FISH, AO/PI	Low	Low	Moderate	Requires separation of cells from a large volume of sample.

The search for life as we know it is facilitated by a vast knowledge of terrestrial life and the development of powerful tools for life detection and characterization. All techniques require an extraction step (*e.g.*, solvents and/or physical agitation) to release the target of interest (*i.e.*, life-form) from the mineral matrix and are destructive for the potential life-form under investigation (except for successful cultivation). Compatibility of using aliquots of one extract for more than one technique might only be possible for a few cases.

analyses of returned material (McCubbin *et al.*, 2019; Chan *et al.*, 2020). Sequences of any PCR-amplified SSU rRNA genes derived from such samples could easily be compared to those from known spacecraft and spacecraft assembly facility contaminants (*e.g.*, La Duc *et al.*, 2014; Moissl-Eichinger *et al.*, 2015; Koskinen *et al.*, 2017; Regberg *et al.*, 2020). Slim as the possibility is, there could exist life as we know it that is otherwise very different from known life on Earth, *e.g.*, that evinces nucleic acid sequences and protein sequences that are so distinct from those in existing databases that one might conclude that they represent a life form that evolved on Mars rather than Earth. Such a conclusion would have to be made with the utmost care and with the hindsight that we are continually discovering novel terrestrial life forms and an increasing body of unannotated sequences in metagenomic datasets (*i.e.*, microbial dark matter, Rinke *et al.*, 2013). This has occurred partly through the development of new tools. Sequencing approaches revealed the existence of a third branch of life, the Archaea, only about 40 years ago; high-throughput DNA sequencing continues to unearth new microbial

phyla. Also, the exploration of new, extreme habitats such as the deep ocean and the continental subsurface has greatly expanded our datasets. In other words, life as we know it is much more diverse than we knew just a few decades ago and may encompass even more forms by the time extraterrestrial samples are examined on Earth. Beyond self-replicating life, there are new viruses discovered on a monthly basis. This includes some very different classes of viruses, such as the giant viruses found across widespread habitats and ecological systems (*e.g.*, Brandes and Linial, 2019) and newly confirmed bacteriophages that employ an alternate nucleobase (2-aminoadenine) in the genome (Zhou *et al.*, 2021; Sleiman *et al.*, 2021).

Another consideration if life as we know it is detected will be to ask whether it is alive. This is especially important for the sample safety assessment but also impacts the science. A range of analytical methods is available for determining microbial viability, each with its own sensitivity and specificity (*e.g.*, Emerson *et al.*, 2017). Each method uses a single criterion for determining life *vs.* death along what is actually a continuum, given that cells proceed from active to

inactive and subsequently senesce and eventually disintegrate. Besides cultivation, currently available viability assessments are made on the basis of metabolic activity, positive energy status, and the detection and abundance of ribosomes, RNA transcripts, or intact membranes.

Viruses and other infectious nucleic acids do not have any universal genes, and hence, they have no non-specific PCR primers or genetic probes. PCR of the main functional motifs (*e.g.*, polymerases, helicases, receptor binding domains) that are most conserved among virus families could be used to look for viral signatures. Virus-like particles can also be stained with general nucleic acid stains and viewed by epifluorescence microscopy (Suttle and Fuhrman, 2010), though results can be ambiguous. Prions would likely not be distinguished by mass spectrometric analyses since they are misfolded versions of naturally occurring host proteins, yet a sensitive cyclic amplification of a protein folding assay that tests for protein misfolding of common proteins (*e.g.*, Saborio *et al.*, 2001) could be applied.

Life as we know it is quite varied, and the full range of possible lifeforms and their structure or the range of conditions within which they can survive remains unknown. In assessing the possibility of life on another planet, it is necessary to take into account the possibility that alternative nucleic acids, amino acids, electron transfer systems, and high energy bonds for driving metabolic activity could exist. Investigating such additional considerations is described in the next step.

Agnostic life detection—Step 8. Analytical methods that do not presuppose knowledge of the chemistry of a target life form (agnostic approaches) are especially useful for analyzing samples that contain unanticipated complexity. There are different metrics for complexity in chemistry that are typically associated with specific analytical techniques. Detection of complexity not seen in controls or anticipated by statistical models developed for agnostic analytical methods are interpreted as an indication of potential life that requires further study.

Earlier steps in the test sequence, especially Step 5, include analytical techniques that may inform an agnostic approach with regard to such features as particular classes of molecules, patterns within the molecular weights, or even intrinsic molecular complexity. There is a distinct need for novel techniques specialized for biochemical systems that do not share a chemical heritage with life on Earth. An expanded agnostic search for life could include molecules that are sufficiently complex but not associated with life on Earth (*e.g.*, Marshall *et al.*, 2021), discrete metastable accumulations of elements or isotopes that are not typical of abiotic geological or mineralogical process (*e.g.*, Kempes *et al.*, 2021), and disequilibrium redox chemistries that are not consistent with abiotic redox reactions (*e.g.*, Frank *et al.*, 2013).

To cast the widest possible net for life detection, the range of allowable interpretations for life must broaden. In addition to the expanded interpretive frameworks for typical methods, we present two concepts for agnostic life detection (see Section 3.3.3). Both require amplification and sequencing and explore the possibility for novel metabolisms that would not be detectable by typical biological methods (*i.e.*, Step 7) yet also identify particles with surface chemistry characteristics typical of living organisms. These concepts could be used to recognize organic or inorganic evidence of life. Any

concepts to be used, like those presented here, must address different forms of complexity (*e.g.*, molecular vs surface binding complexities) and use orthogonal techniques in a sense that they use different interactions of analytical technique and sample. The logical consequence of this is also that one agnostic life detection methodology is not sufficient.

3.3.2. Diagnostic elements not explicitly used in the test sequence

Carbon. Life on Earth is based on carbon, which is present as a mixture of simple and complex organic molecules. As a guide to the search for life on Mars, it was assumed that carbon plays a similarly significant role. So, the search for life (extinct or extant) on Mars could be cast as the search for carbon. The rationale for searching for organic molecules is described in the work of Neveu *et al.* (2018). This search should be performed at the detection limits of available instrumentation, though it is acknowledged that the organic compounds released from a single cell in a given sample tube would be below the limit of quantitation (*i.e.*, as required by Neveu *et al.*, 2018) of the most likely instrumentation and that, in many cases, the detection of compounds of interest also means destroying them and disrupting any life present. A cell contains about 40 fg (femto-gram, 10^{-15} gram) of organic molecules (Braun *et al.*, 2016), and even the most sensitive technique likely requires at least 100s-1000s of cells in the sampled volume in order to be detected (*e.g.*, Summons *et al.*, 2014; Bhartia *et al.*, 2010; Braun *et al.*, 1999). Thus, the corollary, that if no carbon is detected there is no life, does not hold true. Hence, no lower limit for carbon detection is set for the test sequence in the SSAF.

Stable isotopes. The stable carbon isotopic compositions of living organisms on Earth are determined by the metabolic pathways that operate in the organisms. However, there is such a wide diversity of carbon isotopic compositions, and no single diagnostic composition or defined fractionation between nutrients and organisms, that the use of carbon isotopic composition as a diagnostic tool for life is substantially compromised. Combinations of isotopic compositions, for example, carbon, nitrogen, and sulfur, might help improve these limitations, though without knowing all the abiotic sources, sinks, and fractionation processes possible, this approach is still considered a weak diagnostic tool. Given that there are distinct isotopic differences between martian geological materials and geological materials from Earth (*e.g.*, Barnes *et al.*, 2020; Franchi *et al.*, 1999; Franz *et al.*, 2017; Füre and Marty, 2015; Shaheen *et al.*, 2015), it is logical to presume that similar isotopic differences might persist between possible martian organisms and terrestrial organisms. Furthermore, it may be tempting to use isotopic composition as a means to differentiate between terrestrial and martian organisms. However, organisms often acquire the isotopic composition of their primary energy sources (*i.e.*, their food) (*e.g.*, Berry *et al.*, 2015; Boschker and Middelburg, 2002; Jennings *et al.*, 2017; Tykot, 2003); so terrestrial organisms that have subsisted on the elements in martian rock would likely inherit an isotopic composition like that of its environmental components. For these reasons, it is not advisable to rely primarily on the isotopic composition of potential biological material to identify martian life or differentiate whether any given life form discovered had a martian vs a terrestrial origin.

Solubility. Solubility is an important aspect in evaluating the potential harmful consequence of martian life on Earth. Terrestrial biology is solution based. Biochemical intermediates and macromolecules are soluble (or can be dispersed) in water or lipids. Cells and viruses can also be dispersed in water. Exceptions are various types of (naked) viruses or biological systems that could have minerals that cover the outside of a cell. With regard to the sample safety assessment, therefore, the concern is whether martian life is, and martian organic molecules are, soluble under physiological conditions. What is important for the test sequence is whether there are soluble organic molecules in extracts that can be detected and characterized by the analytical tools to discern what they are (*i.e.*, *via* Steps 4–7). A separate investigation to assess the solubility of (organic) material in a sample is not considered required and would also unnecessarily consume sample material. For these reasons, solubility is not considered a standalone diagnostic tool but is indirectly addressed by the extraction processes for some chemical and biological analyses.

Metals. Living systems on Earth interact with a range of metals, including those that act as cofactors with enzymes. Roughly one-third to one-half of all known enzymes depend upon metal ions (*e.g.*, Mounicou *et al.*, 2009; Banci and Bertini, 2013). The most common metallic cofactors are Mg, Ca, Zn, Mn, and Ni; Fe and Cu are commonly redox-active; and Co and Mo interact with coenzymes (Banci and Bertini, 2013; Madigan *et al.*, 2019). The concentration of metals in cells and their association with cellular organic compounds suggest that metal profiles might be useful biosignatures. Indeed, the systematic biological study of these metal profiles has been termed “metallomics,” and the suite of metals associated with a cell is known as the “metalome.” Problems with relying on metal data for a sample safety assessment include:

- Their occurrence in concentrated form due to purely abiotic processes;
- The collection of cellular metal proportions, which varies among phylogenetically diverse microbial cells and in response to environmental parameters;
- The evolution of life on Earth to use particular metal ions based, in part, on their availability suggesting that life on Mars could potentially select for utilization of an entirely different set of metal ions than its counterparts on Earth.

For these reasons, metallomics is not considered a strong diagnostic tool for the SSAF.

Morphology. Morphological evidence of life can compound the challenges of life detection as cell-like forms can easily be produced by non-biological processes. The self-arrangement of lipid molecules with hydrophilic heads and hydrophobic tails in water is an example of how molecules with cell-like morphologies can be formed abiotically (*e.g.*, Dworkin *et al.*, 2001; Jordan *et al.*, 2019). The chemical behavior and relative size of the hydrophobic heads of lipid molecules causes them to pack into a cell-like arrangement called a micelle, in which the hydrophilic heads face outwards toward the water and the hydrophobic tails are positioned toward the center of the 3-D micellar structure.

Similar structures can be generated by polymers in water where a dense phase forms droplets within a more dilute phase and the droplets represent cell-like compartments. These entities, known as coacervates, were implicated in early origin of life models proposed by Alexander Oparin, who hypothesized that coacervates could have operated as protocells. Spontaneously formed cell-like structures can also leave residues that can be misinterpreted as life, what J.D. Bernal called “jokes of nature” (Urey, 1962). The early 1960s saw reports of “organized elements” in carbonaceous meteorites derived from asteroids. Claus and Nagy (1961) believed that these entities could be microfossils indigenous to the meteorite. Subsequent studies revealed that these entities were either exogenous materials, such as pollen and fungal spores that had contaminated the sample, or endogenous materials such as olivine crystals (Fitch and Anders, 1963). Observations of cell-like morphologies have also been used to suggest evidence of life in meteorites from Mars. Scanning electron microscope (SEM) images of ALH84001 revealed segmented tubular structures that were interpreted as fossil nanobacteria (McKay *et al.*, 1996), though later work implied such features were related to crystalline pyroxene and carbonate growth steps (Bradley *et al.*, 1997). Cell-like morphologies have also led to misinterpretations of evidence for early life on Earth. The 3.5 Ga Apex Chert in Western Australia contains filament structures that were once interpreted as oxygen-producing cyanobacteria (Schopf, 1993), yet modern interpretations of the host rocks suggest that the structures originated in a hydrothermal vent rather than the originally proposed shallow sea floor setting (Brasier *et al.*, 2002). The Apex Chert filament morphologies that were assigned to a biological origin have also been reinterpreted as carbon that may be organic compounds generated by Fischer-Tropsch-type reactions during hydrothermal serpentinization of ultramafic rocks (Brasier *et al.*, 2002) and as organic molecules that adsorbed onto self-organized crystal aggregate biomorphs (Garcia-Ruiz *et al.*, 2003) or exfoliated phyllosilicates (Wacey *et al.*, 2015). The filamentous morphologies have also been reinterpreted as aggregates of hematite microcrystals (Marshall *et al.*, 2011). It is worth noting that a biologic origin of the filamentous microstructures has not been demonstrably excluded, since they could represent remnant chemolithoautotrophs that lived in a hydrothermal setting (Schopf *et al.*, 2018). In general, cell-like morphologies remain controversial because there are many processes in nature that generate life-like microscale objects that include tubular, filamentous, framboidal, and dendritic structures (*e.g.*, Cosmidis and Templeton, 2016; Garcia-Ruiz *et al.*, 2009; Kotopoulou *et al.*, 2020; Muscente *et al.*, 2018; Rouillard *et al.*, 2018; 2021; McMahan *et al.*, 2021). Given the extensive history of incorrect interpretations for life based on morphological evidence alone, morphology is not considered a reliable stand-alone criterion for or against life, though it may be useful when associated with chemical information or to inform subsequent steps in the test sequence (*e.g.*, Step 4).

Cultivation. The SSAF is in agreement with the position of the NRC Committee on Mars Sample Return Issues and Recommendations that “Attempts to cultivate putative organisms, or to challenge plant and animal species or tissues,

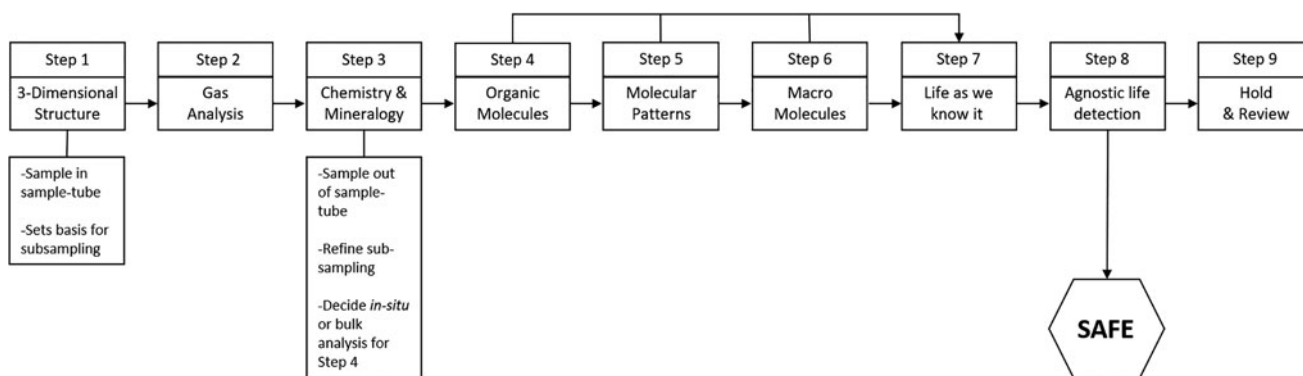


FIG. 3. Overview of integrated test-sequence. The test-sequence is a set of sequential investigations (*i.e.*, steps), each one responsive to the previous steps. There is only one real gate—Step 8—in terms of stopping any further investigations and declaring a sample tube safe within the pre-defined level of assurance. Step 9 establishes a Hold & Critical Review for any sample investigations and executes a set of activities to evaluate all relevant data and the risk management measures, before deciding on the next steps.

are not likely to be productive” (NRC, 1997). The major limitations of this approach are that cultivation is not even possible for most terrestrial organisms and challenge tests are typically tailored to one or a few targets of interest. In addition, it is not considered advisable to multiply viable organisms that could have unknown and potentially harmful consequences. Therefore, cultivation is not considered a diagnostic tool used by the SSAF. As an indirect consequence and due to the limited diagnostic scope that covers the potential avenues of causing harm, animal and plant inoculation are ruled out as well.

3.3.3. Integrated test sequence and candidate instruments. Figure 3 describes the test sequence, and Fig. 4 explains the nomenclature used in the context of the test

sequence. Rather than applying a scattergun approach (*i.e.*, using all techniques available) or a piecemeal approach (*i.e.*, focusing on individual steps or using a particular technique), it is critical to establish an ensemble of techniques and instruments capable of producing the information required for the safety assessment. Table 3 includes a number of techniques and instruments that could provide this information. The list of analytical instrumentation draws heavily on the list prepared by the MSPG2 (Carrier *et al.*, 2022). Some techniques are complementary and overlap with other techniques, which, from a science point of view, is advantageous. From a safety assessment point of view, complementary or overlapping information acquired with different levels of sensitivity and specificity could lead to challenges in its interpretation if this is not considered in advance. In

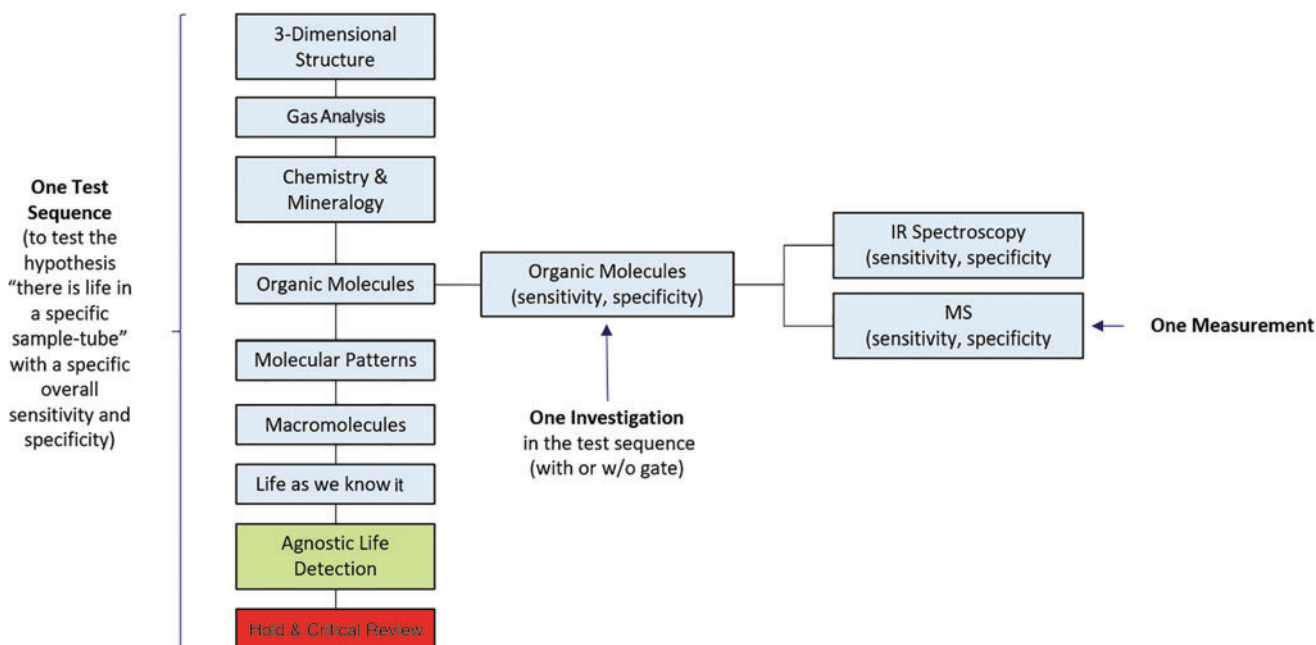


FIG. 4. Nomenclature used in the context of a test-sequence. The elements of the test-sequence (*i.e.*, investigations) address individual questions. Each investigation includes typically more than one measurement technique or instrument. The measurements provide the data that are discussed at the level of investigations. The safety assessment for one sample tube is based on the scientific assessments of the individual investigations carried out on the subsamples.

TABLE 3. DESCRIPTION OF TEST-SEQUENCE

<i>Step #</i>	<i>Investigation</i>	<i>Information</i>	<i>Possible Technique</i>	<i>Sample Preparation</i>	<i>Comments</i>
1	3-D Structure	<ul style="list-style-type: none"> External and internal structures at micron level Informs pre-test probability Informs subsampling approach 	High-resolution X-ray computed tomography (HR-XCT)	No	Samples still in their respective sample tubes
2	Gas Analysis	<p>Volatile organic components</p> <p>Organic molecules, including isomers; isotopic compositions</p>	<p>Selected ion flow tube mass spectrometry (SIFT-MS) or photon transfer reaction mass spectrometry (PTR-MS)</p> <p>Gas chromatography–isotope ratio mass spectrometry (GC-IRMS), plus high temperature conversion elemental analyzer (TC/EA)</p>	<p>No</p> <p>No</p>	Head-gas of each sample tube or gas in dedicated atmospheric sample
3	Chemistry and Mineralogy	<p>High spectral resolution imagery</p> <p>High spatial resolution imagery; elemental analysis (mineralogy)</p> <p>High spatial and spectral resolution characterization of mineralogy</p>	<p>Multispectral/Hyperspectral imager</p> <p>Variable pressure–scanning electron microscopy (SEM)</p> <p>Fourier transform infrared (FTIR) spectrometry</p>	<p>No</p> <p>Uncoated chip or polished block</p> <p>Uncoated chip or polished block</p>	Samples removed from the sample tubes
4	Organics	<p>Bulk organic abundance</p> <p>Spatial distribution of organic compounds</p> <p>High spatial and spectral resolution characterization of organics</p>	<p>Oxidation/pyrolysis gas chromatography–mass spectrometry (GC-MS); thermogravimetric analysis–evolved gas analysis (TGA-EGA)</p> <p>Deep ultraviolet fluorescence</p> <p>Confocal Raman spectrometry, Fourier transform infrared (FTIR) spectrometry, imaging mass spectrometry</p>	<p>No</p> <p>No</p> <p>Uncoated chip or polished block</p>	
5	Molecular Patterns	Identify organic molecules, including isomers; isotopic compositions	<p>Gas chromatography–isotope ratio mass spectrometry (GC-IRMS)</p> <p>Imaging mass spectrometry</p>	<p>Wet chemistry</p> <p>Uncoated chip or polished block</p>	
6	Macromolecules	Characterization of polymers	Liquid chromatography tandem mass spectrometry (LC-MS/MS)	Wet chemistry	

(continued)

TABLE 3. (CONTINUED)

Step #	Investigation	Information	Possible Technique	Sample Preparation	Comments
			Capillary electrophoresis–mass spectrometry (CE-MS)	Wet chemistry	
			Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), Electrospray ionization mass spectrometry (ESI-MS)	Wet chemistry	
			Imaging mass spectrometry	Uncoated chip or polished block	
7	Life as we know it		Many	Wet chemistry	
8	Agnostic Life Detection	Chemical complexity that is beyond what is expected from abiotic process		Wet chemistry	

Wet chemistry involves, for example, solvent and acid extraction, sonication, derivatization.

this context it is considered essential that, regardless of the instrumentation or techniques that are ultimately selected, their limitations are well-understood and their performance is known and dependable (see chapter 4.2 for a way to address these issues).

Steps 1 to 3 of the test sequence are concerned with identification of features that are not associated specifically with living entities, although they may have been formed by life. Measurements focused on analyses of textures, mineralogy, chemistry, and gases are the same types of analyses that are currently used to identify possible biological and biogenic features in geological materials. These first 3 steps employ routinely tested and well-understood analytical techniques, such as microscopy, spectroscopy, and chromatography. For some steps, only one technique might be applicable, though it is inevitable that, for many of the steps, several different instruments could deliver the required results. For example, given appropriate calibrations, the mineralogy of a sample could be determined by optical or electron microscopy, IR spectroscopy, Raman spectroscopy, or X-ray diffraction. It is also the case that the same instrument could deliver required information for several steps. For example, Raman spectroscopy can identify the mineralogy of a specimen (Step 3) and the types of organic molecules (Step 4) that it contains.

Steps 4 to 6 of the test sequence cover analysis of organic material, including organics that are not necessarily of biological origin. There is a wide variety of techniques and instrumentation available for the required analyses. Step 4 is a measurement of the presence or absence of organic compounds. Moving from Step 4 to Step 6 employs techniques of increasing specificity to enable acquisition of the required information: if organic material is present, what are its characteristics? The information includes recognition of molecular patterns and isomeric variations associated with individual

species (e.g., amino acids, lipids) as well as the presence of macromolecules (which may, or may not, be polymeric). When using the techniques available at the time of this writing, progression to Step 6 requires increasing invasion of the selected subsample through treatment with a sequence of solvents (i.e., polar, non-polar, acidic, alkaline) to produce solutions for introduction of the processed sample into appropriate analytical instrumentation. The main technique for analysis of organic species is mass spectrometry, though the differing chemistries and molecular masses of the components require specific methods to introduce samples into the analyzer. Examples in use currently include Gas Chromatography (GC), Capillary Electrophoresis (CE), and High Performance Liquid Chromatography (HPLC). Alternatively, high molecular mass compounds can be analyzed by imaging mass spectrometry techniques (e.g., MALDI-MS, Laser Desorption/Ionization Mass Spectrometry (LDI-MS), Desorption Electrospray Ionization Mass Spectrometry (DESI-MS), nano-DESI-MS and ToF-SIMS). These techniques enable *in situ* molecular analysis at high spatial resolution when coupled to optical and electron microscopy and constitute an area of research that is rapidly developing. Depending on the ionization method, the techniques can analyze a wide mass range (1-100,000 Da) with a spatial resolution down to less than 1 micrometer with minimal sample preparation (Watrous *et al.*, 2011; Heeren, 2015; Bodzon-Kulakowska, 2016).

At Step 7, the question changes from how best to identify the characteristics of organic material to whether the material has come from a living (or dormant) biological form of life. The equipment proposed for Step 7 assumes that any organisms present have characteristics that produce analogous signals to those that we observe on Earth, and hence, they can be detected by the same instruments used for determination of terrestrial evidence of life. Step 7, then, is looking to sequencing

techniques for amplification of genetic material. Nucleic acids are relatively easy to detect, and moreover their genetic sequences can reveal a vast amount of information about the life forms that synthesized them. Variations of PCR can provide further information. For example, qPCR can quantify gene copy number (and thus estimate cell number), and reverse transcriptase PCR (RT-PCR) can bias the assay in favor of active, rRNA-rich cells. High throughput metagenomic and transcriptomic sequencing is increasingly being used to more fully characterize microbes and their activities, which requires greater amounts of nucleic acids for analysis since there is no amplification step. The technology has now advanced such that as few as 50 cells may be fully characterized (Minich *et al.*, 2018). Additional analytical methods include single cell genomics (Woyke *et al.*, 2017) and a mini-metagenomics approach, which can characterize the genomic features of 5-10 cells (Yu *et al.*, 2017). Both single-cell genomics and mini-metagenomics require the amplification of DNA from cell(s) and are designed for samples with low cell abundance. These detection and characterization techniques are relatively mature, such that while we anticipate incremental improvements in the coming decade, the fundamental principles will likely still apply.

Step 8 is going beyond the familiar, that is, agnostic life detection with an amplification step and minimal assumptions. From a sample safety assessment framework point of view, this is the most important element in the test sequence and the only one with a clear gate. At the same time, it is the least defined step in terms of techniques and robustness that can only be addressed by targeted developments. Two concepts are described that could benefit from such targeted developments. The first concept is focused on the identification of non-canonical information polymers. Current nanopore-based sequencing technology is well suited to expanding the search for informational molecule patterns beyond the specific amino acids and nucleotides conserved in contemporary extant life on Earth. This technique is amenable to analysis of the diversity of informational polymers that might have been common in a pre-RNA or RNA world, before the diversification of life and dominance of DNA- and protein-based life (Joyce, 2002). Polymerase evolution and design experiments have found six additional possible RNA alternatives and precursors, such as threose nucleic acid (TNA), hexose nucleic acid (HNA), and other xenonucleic acids (XNAs, which are nucleic acids not found in nature), all of which can store and transmit genetic information (Pinheiro *et al.*, 2012). Strands of XNAs can also bind to target ligands with high affinity and specificity, which demonstrates the capacity for preferential folding that is associated with Darwinian evolution. While this study is speculative about the nature of RNA alternatives and precursors, there are many examples of DNA and RNA alternatives used in nature that include methylated forms of DNA (Moore *et al.*, 2013), a 2,6-diaminopurine found in the DNA of bacteriophages (Sleiman *et al.*, 2021), and over 120 modified forms of RNA found in ribosomal and transfer RNAs (Schaefer *et al.*, 2017). These exceptions to the highly conserved structures of DNA and RNA only strengthen the need for a capability that extends beyond characterization of the standard forms of DNA, RNA, and proteins when searching for unfamiliar life. Life as we know it is generally based on multiple classes of polymers with conserved sets of monomer units. The differences in number and sequence order of these monomers—their informational content—are what

distinguish the structure and function of these types of polymers. Repetitive polymers are not necessarily informational or biological molecules, however; and abiotic polymers of carboxylic acids and amines (*e.g.*, nylon or polyester) represent a case where neither is true. Biology with a unique origin may capitalize on the informational capability of unique semi-repetitive polymers based on alternative genetic alphabets (monomer chemical structures), which would require analysis of any polymer that contains a set of semi-repeating monomers. This type of sequencing is possible with nanopore (electrochemical) devices that can detect a broad range of water-soluble, charged molecules (nucleic acids, proteins, polyions, etc.) with simpler and faster sample preparation than required by other commercially available sequencing platforms. Nanopore sensing is “agnostic” in that it analyzes any linear polymer that enters the pore. Nanopore devices can distinguish between monomers with slight differences in shape, volume, or polarity and only require a template to tune for the voltage-driven translocation rate for identification (Branton *et al.*, 2008). Nanopore analyses have been used to sequence RNA (Garalde *et al.*, 2018), inosin-bearing oligonucleotides (Carr *et al.*, 2017), methylated nucleobases (Rand *et al.*, 2017; Simpson *et al.*, 2017), and even proteins (Ouldahli *et al.*, 2020). The proposed concept can be used to interrogate returned samples for non-canonical polymers that could indicate a novel informational or catalytic polymer distinctive from those used by biology on Earth.

The second concept is focused on randomly generated oligonucleotides to build an informatics fingerprint that represents the binding complexity of a particle surface. The patterns of nucleic acid binding to surfaces, independent of their biological function, can be used to probe and report on any chemical environment, which opens up a new way to detect evidence of life. This concept (Johnson *et al.*, 2018) targets the secondary and tertiary structures that oligonucleotides naturally form that can have affinity and specificity for a variety of molecules—from specialized biomolecules, such as peptides and proteins (*e.g.*, Jayasena, 1999), to non-linear polymers, and even to inorganic substrates such as mineral and metal surfaces (Cleaves *et al.*, 2011; Ye *et al.*, 2012). Short DNA sequences (~15 nucleotides) or “aptamers” will bind to all types of chemical structures in complex samples, similarly to how antibodies bind to analytes. Unlike antibodies, however, aptamers are agnostic in that they comprise a nearly unlimited variety of binding specificities, whereas antibodies have been selected for recognition of limited types of biomolecules. Aptamer binding is driven by the surface chemistry of the analyte and limited only by chemical characteristics that discourage DNA binding, such as occurs in those regions of strong negative charge or when there is a deficit of aromatic or hydrophilic moieties. By accumulating large numbers of binding sequences that reflect different compounds in a mixture, statistical data analyses of aptamer motifs and sequence counts generate patterns associated with increasing levels of complexity that distinguish biological surfaces to be analyzed. This pattern recognition, known as “chemometrics,” represents a set of protocols that can be applied to find patterns in chemical data sets (Nie *et al.*, 2015), which in turn can be used to fingerprint agnostic evidence of life. The statistically derived level of complexity in aptamer sequences can be analyzed to generate high-dimensionality chemometric score plots that reflect the complexity and assumed biogenicity of the resulting pattern.

To optimize the use of sample material, it may be possible to use a sample for more than one investigation or analysis. In the context of the SSAF, this approach would only be acceptable if it is shown that multiple uses of sample material cannot lead to an increased false-negative rate in the overall assessment.

3.4. Decision criteria

Figure 3 shows the entire test sequence. Investigations in Steps 1 and 3 inform two kinds of decisions:

- Sequence of opening and investigating the individual sample tubes from Mars.
- Number, type and locations for subsampling the sample in each sample tube.

There are no yes/no criteria or specific thresholds levels to reach a decision for these two steps. The decisions will need to be based on informed judgements. A positive test for organic compounds in Step 4 is suggestive of the potential for biology, although abiotic chemistry (*e.g.*, that found in carbonaceous chondrite meteorites) or terrestrial contamination can result in the presence of organic compounds as well. A negative test for organic compounds in Step 4 does not necessarily indicate the complete absence of organic molecules. Rather, it would indicate that—if any molecular evidence of biology is present in the sample—it is in very small concentrations that are below the level of detection or strongly bound to the substrate. A positive test for molecular patterns (Step 5) should be viewed as highly suggestive of the potential for active or recent biology. The abundance and the signal-to-noise of the patterns (for example homochiral in all species *vs.* 20% enantiomeric excess in some species) must be compared to plausible abiotic formation and preservation processes for such compounds and the best current knowledge of the samples and martian environment. A negative test for molecular patterns with a positive test for organic compounds suggests that if biology is present, it is overwhelmed by organics or degraded organic material or that biology is absent. A positive test for macromolecular patterns should be viewed as highly suggestive of the potential for active or recent biology or terrestrial contamination. The nature of the macromolecules would need to be assessed in the next steps to determine whether they arise from terrestrial contamination or martian biology and if these macromolecules are suggestive of extant or preserved extinct biology. A negative test for macromolecular patterns with a positive test for organic patterns suggests that, if biology is present, it is a metabolic hypercycle (Eigen and Schuster, 1997) or uses macromolecules that are resistant to analysis or that the life died and its macromolecules degraded before analysis. Best current knowledge of the samples, martian environment, and the environments the samples have experienced from collection to analysis must be used collectively to assess whether the molecular patterns observed could have originated from degraded biological macromolecules.

Failure to detect organic compounds, molecular patterns, or macromolecules is not considered sufficient to declare a sample safe. Among other reasons for a negative detection (*e.g.*, strong binding to the mineral matrix), the sensitivity of the available techniques could miss the equivalent organic molecules of hundreds to thousands of terrestrial cells (see

Section 3.3.2). As a consequence, a negative detection in Step 4-6 must be followed up with an amplification step (*i.e.*, Steps 7 and 8). Step 7 is important for two reasons—to detect any remnant terrestrial biological contamination in the samples and to detect evidence of martian life that is similar to terrestrial biology. It is expected that this step could lead to a number of positive events that are likely associated with terrestrial contamination. However, until any evidence for life can be clearly associated with terrestrial contamination, the conservative assumption (positive hypothesis) is that it could be martian biology. A negative detection in Step 7 would demonstrate that the samples are free from terrestrial biological contamination, within the detection limits of the analytical techniques. Even so, the potential for martian life to be present still cannot be excluded because this step is highly biased toward life as we know it. The only definitive gate is actually Step 8. If there is no evidence for the presence of martian life in the samples and there are no open, uncertain, or ambiguous issues remaining that could associate sample characteristics to martian biology, then the sample of a sample tube would be deemed safe within the pre-defined level of assurance.

In the case that potential evidence of *extant* martian life is detected, a Hold and Critical Review (HCR) must be initiated to evaluate the *status quo* before proceeding. This approach is similar to having a spacecraft enter a safe mode: until it is understood what triggered the safe mode and it has been concluded that it is safe to proceed, normal spacecraft operations would be suspended. Details of the HCR must be described in the Sample Safety Assessment Protocol (SSAP). The Critical Review must include a comprehensive and holistic evaluation of all relevant data acquired, the analytical techniques and specific instruments and equipment used, the methods and procedures used to control the safety of Earth (*e.g.*, containment design and operations, sterilization procedures and criteria), and the overall risk assessment. Only then could it be decided as to whether the Hold would apply to investigations on subsamples from the one sample tube being analyzed, on samples in other sample tubes, and/or on samples already released from containment. Further investigations that are responsive to the data and the understanding at that time would likely be required to assess whether and how a hazard analysis could be executed. While not directly a concern for the safety of a specific sample, finding evidence of *extinct* martian life must also lead to an HCR. In such a case, the overall risk posture reflected in the level of assurance must be reviewed. Establishing the initial level of assurance typically follows a conservative approach. However, there is a significant difference between the *a priori* assumption that there is life on Mars and having evidence that life emerged on Mars. The need for this is further illustrated by samples from Earth that simultaneously contain evidence of both *extinct* and *extant* life (*e.g.*, surface exposed rock on Earth that contains evidence of ancient fossils and viable microbial inhabitants).

The HCR approach would have to be reflected in agreements that cover the release of samples from the SRF and their subsequent use. An important aspect in terms of managing expectations is to acknowledge that an HCR might be a re-occurring event due to possible terrestrial biological contamination. Comprehensive contamination knowledge (CK) could expedite the HCR. The HCR and any

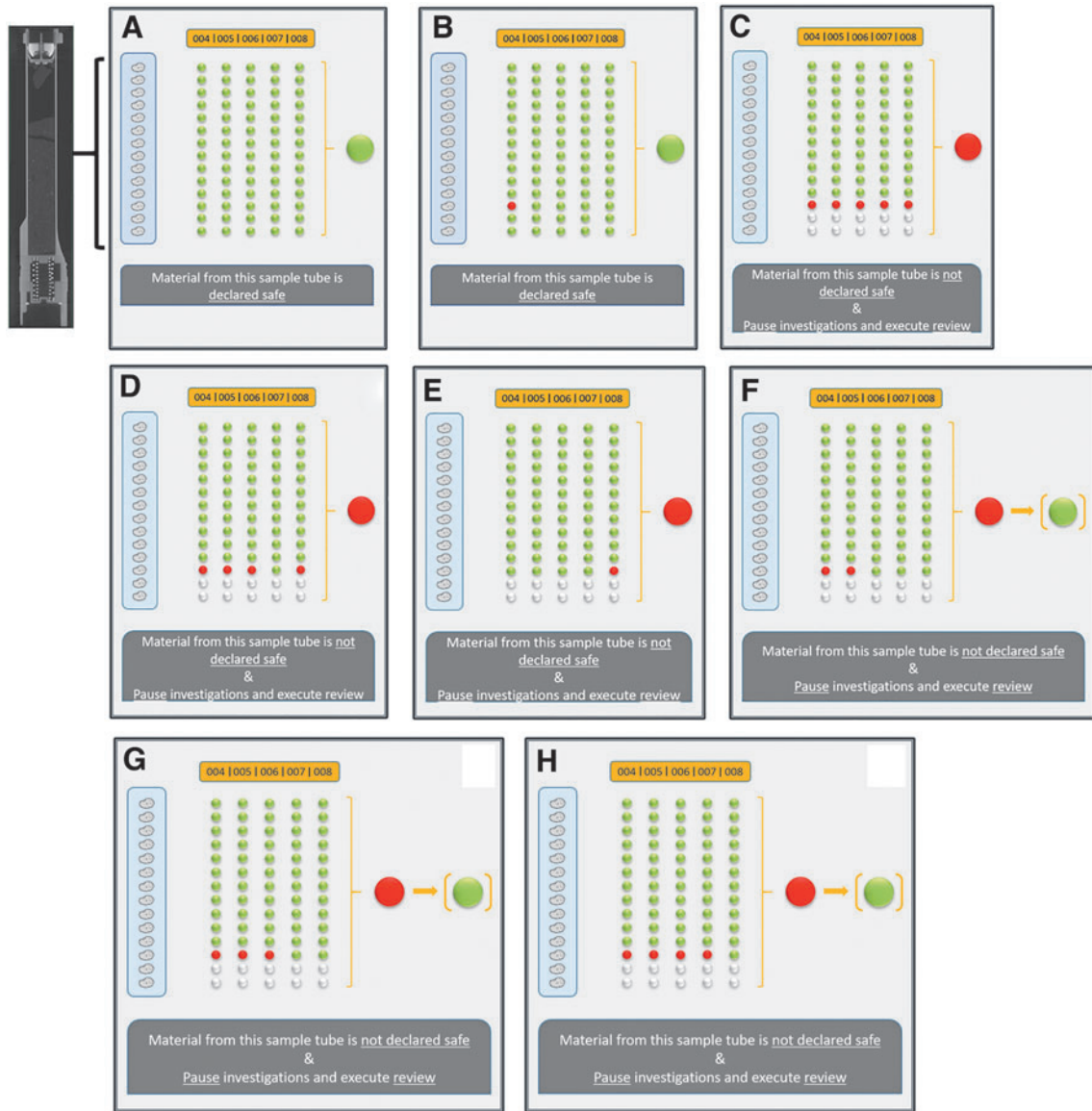


FIG. 5. Eight generic cases for one sample tube are shown in this figure. This example is based on a situation where 14 negative results running the test-sequence on every single subsample are required to achieve a pre-defined level of assurance. Cases A and C are straightforward. Case B represents a situation where we have either terrestrial contamination or evidence for abiotic martian organics. Cases D and E represent a situation where we have evidence of martian life that is quite different from terrestrial life. Cases F, G, and H represent a situation where we either have terrestrial contamination, evidence of martian life that is quite similar to terrestrial life or a combination of both.

decision associated with it must be performed by an independent team that has decision authority for executing the SSAP (see additional details in Section 4.1).

Figure 5 describes some of the possible outcomes of going through the test sequence. The element of terrestrial biological contamination is specifically highlighted in several cases.

4. Implementation of the SSAP

The implementation of the SSAP is focused on the safety assessment of each individual sample tube. As already pointed out in Section 3.3.1, the most effective approach is to start with the sample tube(s) that have the highest pre-test proba-

bility with respect to finding martian life. In the case in which a dedicated dust sample is returned to Earth, this might be a good starting point. The result of such an assessment can inform the pre-test probability for other sample tubes.

In estimating the amount of sample material needed to inform the safety assessment, it must be taken into consideration that the sample safety assessment and many mission science objectives are complementary, overlapping, and apply similar methods. None of the samples used to inform the sample safety assessment should be considered wasted, as scientifically useful data will be generated and will inform each step. The amount of sample material required is naturally inversely proportional to the amount of biological material present and strongly depends upon the extraction

processes and analytical techniques available at the time. Considering collectively the current capabilities for measurements on natural samples, the number of subsamples per sample tube required to reach a certain level of assurance that no martian life is in a sample (Section 3.2), and the quality control necessary to achieve confident results (Section 4.1), it is estimated that hundreds of milligrams to a few grams of sample mass per sample tube are required to inform the safety assessment. It is expected that targeted developments in extraction processes and advances in analytical techniques would further reduce the required amount of sample material to be processed and inform the sample safety assessment.

In addition to the four elements of the SSAF (Fig. 1), there are a number of implementation constraints that are part of the SSAF and are described in the following section.

4.1. Quality control

The consequence of an incorrect safety assessment could range from reduced sample access to harmful impacts on Earth's systems. To increase confidence in the results of the assessment, it is essential to have decision-critical investigations in the test sequence performed independently by more than one team. The detailed implementation of this approach will depend on the nature of the investigation and the associated measurements. There are three possible implementation approaches:

- Investigations where a single measurement conducted (*e.g.*, an XCT scan of the sample) is deemed to be determinative, and two different teams independently analyze and interpret the data.
- Investigations where two independent measurements utilize a single technique, for which significant expertise and experience are required to obtain reliable results (*e.g.*, two GC teams analyze aliquots of an extract).
- Investigations where two independent measurements made with complementary techniques are utilized to increase the predictive value of the results (two different techniques are used, *e.g.*, spectroscopy and spectrometry, to analyze aliquots of an extract).

The use of complementary techniques increases the probability that a given result is true if the datasets agree and that they will trigger further investigations if disparate. Although use of the same technique more than once compensates for intra-technique variability, it is more critical to address the measurement accuracy in the safety assessment context (since precision is already accounted for in the selection and validation of the chosen techniques). The execution of the test sequence must follow an approach typically used in science and engineering when assessing public safety or environmental impacts, namely deploy two independent teams to perform the measurements or data analysis (see three cases above), with a third independent team responsible for decision making. This approach must be considered in the planning of opportunities for science teams that will cover the objective-driven science investigations on the samples, some of which will inform the safety assessment, and in the planning and operation of the associated infrastructure (*e.g.*, SRF).

All analytical methods used for the sample safety assessment must be documented and independently reviewed

in advance. Any variations that occur as a single incident or result in a change of the test must be assessed and their consequences recorded at the time. ISO 17025 (ISO 17025, 2017) and equivalent standards are the mark of a laboratory with good quality systems, record keeping, and general operation, including appropriate staff training. For laboratories that test samples from Mars, these standards are a sound foundation on which to build to ensure reliable results. To allow scientific inquiry to follow a thread that is informed by successive findings that may not have been foreseen, while at the same time maintain a high standard of quality control, record keeping, data integrity, and data security, it is essential to apply methods of Good Laboratory Practice (GLP) (*e.g.*, OECD, 1998) and Hazard Analysis at Critical Control Points (HACCP). These are routine methods used in the clinical setting and in industrial process quality control. HACCP was developed for the food industry, including to assure the safety of food products for the U.S. space program, but can be adapted for almost any complex operation in which safety risks and potential risks to a product are concerned (*e.g.*, Hulebak and Schlosser, 2002). An HACCP-like process is a dynamic way to predict problems in advance and put in place risk-mitigating steps before any experiment or process is performed. In a general application, these risks may be anything from instrument failure and external contamination of a key sample or product to a human mistake, and can include factors that affect safety, quality or scientific output, and integrity. In any process there are steps where something could go wrong, and HACCP-like analysis concentrates on these points. Inevitably “stuff happens,” and the lessons from these events are used to update the HACCP-like assessment and mitigations in a continuous fashion. Traceable records that maintain a log of any alteration in procedures or risk assessment performed and by whom are kept throughout. The HACCP-like process can be applied theoretically while mapping the process. For the SSAF, this must be supplemented through full scale sample safety assessment simulations with analogue materials and implementation of the final test sequence. HAZOP (Hazard and Operability) is an example implementation of the basic HACCP process for general industrial use and is explained in detail in the IEC 61882 standard (IEC 61882, 2016). Figure 6 shows the basic steps involved in setting up a HACCP-like system that can be adapted to fit the workflows around handling samples from Mars. It shares some common features (*e.g.*, documentation) with the ISO 17025 standard, and the two processes can be combined. Aspects of the HACCP-like process can be developed alongside the set up and calibration of the instrumentation that, under ISO 17025, must be performed at the actual site of use before any genuine part of the test sequence is undertaken.

Humans are a key factor in creating errors, most of which arise from a lack of training, lack of experience in a particular situation, poor management practices that place workers in a situation they cannot control, which can garner fear of raising concerns or simple lack of coordination within the team. A related cause is an ergonomic one, where poorly laid out displays, inaccessible controls, poor seating or repetitive manual tasks predispose operators to errors and disasters. Aircraft flight crews and space crews are trained to work together, especially if something goes wrong, and are

1	2	3 [†]	4	5	6 [†]	7 [†]
Conduct a Hazard Analysis	Determine Critical Control Points (CCP)	Establish Critical Limits	Establish Monitoring Procedures	Establish Corrective Actions	Establish Verification Procedures	Establish Documentation Procedures
List all steps in the SSAP and identify the steps in which something could go wrong.	Points where a potential error could be mitigated.	Maximum and minimum values within which a parameter must fall for a test to be valid.	Monitor critical limits at each CCP.	Plan to prevent each potential hazard or problem from occurring or reoccurring at a future date, acting proactive and reactive.	Ensure the HACCP plan is working as it should.	Documentation must be detailed and comprehensive. It must also comply with local government and international legislation & guidance.

FIG. 6. The example described here aligns with the ISO 17025 process in many areas[†], so that the same documentation can be shared between the two systems. The difference is that HACCP-like analysis covers risk in addition to defining a set of documented processes and allows that risk to be mitigated by forward planning and regular review. This approach is described for application on the detailed SSAP, derived from the SSAP. The term *hazard* in the context of the HACCP-like quality control measure proposed describes hazards to the SSAP process and not the potential biological hazard of material from Mars.

usually deployed in teams where such skills and leadership are essential. Human factors are a critical part of the SSAP and must take into account deliberate wrongdoing as part of a framework for detecting and mitigating an adverse event.

The sample safety assessment, must always consider what will ultimately be done with a sample. Information gathered even after the samples are sent for further analysis (*i.e.*, after they are declared safe) could indicate the need to update the safety assessment for all or specific samples. Therefore, if the sample handling and analysis protocols are modified from those in the submitted program of work, the sample safety assessment protocol will also need to be reviewed to ensure it is still valid for the new circumstances.

4.2. Analogue test program

To make an informed judgement for subsampling and to optimize the capture rate requires high-resolution physical and chemical information about the samples—covered by Steps 1 and 3 of the test sequence. This information is a necessary, though not sufficient, part of the process of rendering a robust and credible informed subsampling decision. The second and essential part of the process requires correlation of this information with knowledge base information obtained from subsampling terrestrial samples. To establish such a knowledge base requires an analogue test program tailored to include the types of materials expected from Mars that are analyzed with the types of instruments to obtain the kinds of measurements that are planned to be used to establish the physical and chemical sample information from returned martian samples.

Steps 1 and 3 are only part of the test sequence. The overall test sequence is a series of individual tests of dif-

ferent types, each with its respective sensitivities and specificities. In terms of the Bayesian analysis, what matters is the overall sensitivity and specificity of the test sequence. This overall sensitivity and specificity might be derived from combining the sensitivity and specificity of the individual measurements of the test sequence. Although it is important to establish the sensitivity and specificity of the individual measurements to optimize the use of samples and maximize the incremental flow of information, it is not likely that a pure mathematical combination of these values would correctly reflect the sensitivity and specificity of the overall test sequence. By exercising the test sequence on terrestrial samples that represent the expected material from Mars using an analogue test program, it is possible to optimize details of the test sequence (*i.e.*, by guiding selection of appropriate measurement techniques and instrument ensembles as well as the order of applying them on samples). General considerations for optimization of the test sequence are:

- The selection of instruments that will generate the highest true-to-false positive ratios;
- Ordering of the sequence of analyses to start with instruments with a higher sensitivity before moving to instruments with higher specificity.

To illustrate the utility of this approach to optimize the test sequence, a common issue for almost all investigations that target the organic content of samples is the separation of organic molecules, cell debris, or whole life forms from a mineral matrix. The best sensitivity and specificity of the analytical techniques used in the test sequence can only be employed if the targets of interest (*i.e.*, molecules, cell

debris, cells) are presented in a useful form. A poor extraction efficiency increases the chance for false-negatives, independent of however good the analytical techniques are. Some mineral matrices are well known for their ability to retain organic compounds, owing to their surface properties and structure, for example, clay minerals. Other matrices are known for their propensity to attract organic compounds as a function of their chemical properties and structure, for example, macromolecular organic matter and carbonaceous materials. Mineral matrices that retain organic compounds are actually used on Earth in analytical and industrial processes as fluid filters to remove organic compounds, for example, clay and activated carbon gas filters for analytical chemistry. For organic compound extraction chemistry, the maxim is that “like dissolves like,” that is, to isolate a compound it must be matched with a solvent of similar properties. However, due to the different polarities of organic molecules, for example, amino acids are relatively polar, while hydrocarbons are relatively non-polar, no single solvent system is able to extract all organic molecules in a sample (Mitra, 2004). The type of matrix (rock) that the organic molecules are trapped in also affects the solubility of the analyte (Mitra, 2004). It is possible to use mixtures of less and more polar solvents to extract organic compounds of different polarity or change the solubility of the analyte using ultrasonication of the solvent or to use supercritical fluids (Mitra, 2004). In general, extraction protocols need to consider the full range of polarities presented by the potential target materials. Streamlining the extraction protocols, in particular consideration of the solvent strength for sharing an extract for multiple different analyses, would be beneficial to limit the use of sample material and support the independent analysis approach required in the frame of the test sequence. In this context, it is important to be aware that some solvents might interfere with other types of analysis (*e.g.*, phenols used for certain omics investigations might interfere with other organic analyses) or exhibit inhibitory effects (*e.g.*, denaturing). The complete extraction of organic compounds from highly retentive matrices may be unachievable, though the most efficient levels and knowledge of the extraction efficiencies are essential for the sample safety assessment.

It is also necessary to exercise all four elements of the safety assessment (not only the test sequence) in end-to-end tests that utilize analogue samples. These end-to-end tests must include blind testing as an integral and essential part of the quality control measures. The added value of blind testing, however, can only be realized if the blind tests are well prepared and properly executed (*e.g.*, Ginsburg, 1997; Casertano *et al.*, 2008; Evans, 2014; van Driel *et al.*, 2019). Such end-to-end tests can be used to optimize the sample flow and help to estimate the resources needed to perform the safety assessment. In addition, end-to-end tests serve to educate and train the personnel and test the various elements of the infrastructure, equipment, and instrumentation necessary to conduct the sample safety assessment.

In summary, there is a need for a tailored analogue test program that covers the following components to transition from the SSAF to an SSAP:

1. Assess and improve the capture rate and the associated subsampling strategy.
2. Optimize the selection of the instrument ensemble to be used for the test sequence and estimate the overall sensitivity and specificity, including the efficiency required to extract evidence of life from the host materials.
3. Exercise all four elements of the safety assessment, including blind testing, to optimize processes, test equipment and infrastructure, and train personnel and science teams.

The selection of the analogue samples to be tested needs to be based on the specific environment (*e.g.*, Cockell *et al.* 2019) and the information obtained during sample collection on Mars. The analogue materials could include synthetically made samples, natural terrestrial analogs, and meteorites. Due to the special role of clays (see Section 3.2), the assumptions about random and targeted subsampling must be verified as part of the analogue test program. The analogue samples must include both negative and positive controls. These could include sterilized and/or organic-free analogue samples (negative controls) and samples doped with microbes and/or organic molecules or well-characterized natural terrestrial analogues known to contain life and/or organic molecules (positive controls).

4.3. Terrestrial biological contamination

Martian meteorites have been shown to be colonized by terrestrial organisms (Toporski and Steele, 2007). In the same way, terrestrial biological contamination of martian samples returned to Earth by the MSR Campaign would reduce the specificity of the overall safety assessment test sequence (see Section 3.1.1). It might also lead to a re-occurring Hold and Critical Review (HCR) of activities on the samples until the root cause of a detection can be clearly identified as terrestrial biological contamination (see Fig. 5). The contamination baseline for returned martian samples must be established from the CK obtained during the assembly of the various spacecraft that will fly as part of the MSR Campaign, along with blanks and witness samples returned with the martian samples. Of particular importance in this regard are the M2020 Witness Tube Assemblies (WTA), which are opened and sealed during different mission phases including pre-launch, launch, cruise, and Mars Entry Descent and Landing (EDL), and M2020 surface operations, as well as the M2020 drillable blank which can provide CK of the M2020 drilling operation. CK samples should also be collected during the construction of the SRF to establish a complete archive of potential contaminants, including biological contaminants that may come into contact with martian samples during sample analysis. Minimizing terrestrial biological contamination in the samples and a higher CK would reduce uncertainty in the scientific interpretation of the data and ease handling and treatment of the samples.

To differentiate between martian or terrestrial origin, the field of omics will play an essential role. The use of transcriptomics, proteomics, and metagenomics can provide a predictive comparison of material at protein, mRNA, and DNA levels, respectively. The biological CK samples (*e.g.*, fallout coupons, spare hardware, microbial DNA and isolates collected from the assembly and test phases during pre-launch, etc.) can be used as a reference library of pre-flight conditions that can be directly compared to any signals from the potential

biological material. The direct comparison may allow for assessments in expression profiles, unique or modified proteins, and changes in the DNA that occur during spaceflight. These advanced molecular techniques are commonly used to study both complex microbial communities and the evaluation of environmental stressors, such as the space environment and catabolism of pollutants in bioremediation (*e.g.*, Biljani *et al.*, 2021; Chandran *et al.*, 2020; and Kumar, 2020).

4.4. Life detection and machine learning

The sample safety assessment defined by the SSAF depends upon the simultaneous interpretation of numerous variables and criteria as well as proper statistical treatment of large datasets that exclude false negatives and false positives. In several ways, this challenge is similar to that of biogenicity tests for putative traces of fossil life in the Earth's rock record. Classical tests of biogenicity in deep time involve the evaluation of multiple biosignature characteristics and context- and contamination-related criteria that need to be satisfied to substantiate a claim (*e.g.*, Buick, 1990; Schopf *et al.*, 2010; Brasier and Wacey, 2012; Neveu *et al.*, 2018). The number and combinations of these characteristics and criteria, however, are subject to debate, since it is easy to include false positives or exclude false negatives. In reality, there are no clear yes/no answers in biogenicity tests since all biosignatures have a certain probability that life created them and a certain improbability that an abiotic process created them (Des Marais *et al.*, 2008). The qualitative nature of many individual biosignatures (*e.g.*, morphological characteristics) add further ambiguity, as they are often not standardized and depend on the interpretation and experience of individual observers. To overcome this inherent uncertainty in life detection and during efforts to exclude a biological origin, several recent studies have expressed the specific need for standardized criteria and a more quantitative approach in data treatment (Chan *et al.*, 2019; Neveu *et al.*, 2018; Rouillard *et al.*, 2020, 2021). The use of multiple well-defined and quantifiable variables for life detection could greatly benefit from recent advances in statistical methods and machine learning methods to find commonalities in large datasets. While standard statistical methods passively draw inferences from a dataset, machine learning methods create mathematical models based on training data and use this "experience" to find predictive patterns in new datasets (Jordan and Mitchell, 2015; Hastie *et al.*, 2017). Typical tasks carried out by machine learning include classification, regression, ranking, clustering, and pattern recognition. A so-called supervised machine learning algorithm builds a mathematical model based on a training dataset that contains both input and desired output. The program thus responds to feedback. In contrast, an unsupervised machine learning algorithm builds a mathematical model without any desired output. It finds patterns in a dataset, and then attempts to find similar ones in newly supplied datasets. With the ongoing increase in computing power and the development of artificial neural networks, these collective methods are rapidly improving. Machine learning is currently transforming the field of medical diagnostics (*e.g.*, Aggarwal *et al.*, 2021), and is widely applied for face recognition in forensic applications (*e.g.*, Phillips *et al.*, 2018), general

image recognition (*e.g.*, Krizhevsky *et al.*, 2017), use of large data sets of space missions (*e.g.*, Kronberg *et al.*, 2021), and speech recognition (*e.g.*, Hinton *et al.*, 2012). Some machine learning methods have already found applications in the detection and classification of life. For instance, convolutional neural network classification models were developed and trained to perform visual palynological identification and taxonomic classification of fossil pollen (Romero *et al.*, 2020). For the purpose of the SSAF, there are three categories to be considered:

1. No life
2. Life as we know it
3. Life as we don't know it

For the first two categories, a supervised machine learning algorithm can potentially be created if appropriate training datasets can be supplied and unambiguous desired outputs can be formulated. Training datasets for "no life" would consist of the morphologic, compositional, and isotopic characteristics of structures, such as crystals, sand grains, and precipitates, for which an explicit geologic formation process exists. Training datasets for "life as we know it" constitute established biosignatures, such as DNA sequences, certain organic molecules, and certain complex morphologies. Many examples of known biosignatures are presented in various overview papers, most of which focus on the recognition of fossil life (Buick, 1990; Neveu *et al.*, 2018; Schopf *et al.*, 2010; Brasier and Wacey, 2012; Rouillard *et al.*, 2021a). There is, however, a growing literature on "false biosignatures", *i.e.*, physiochemical processes that lead to the formation of minerals or molecules with life-like features (Cosmidis and Templeton, 2016; Garcia-Ruiz *et al.*, 2003; Jordan *et al.*, 2019; Kotopoulou *et al.*, 2021; Rouillard *et al.*, 2018, 2021b; McMahon *et al.*, 2021). It may thus prove difficult to define clearly the exact difference between category 1 "no life" and category 2 "life as we know it". An important future scientific challenge lies in properly defining this difference, and subsequently creating widely accepted training datasets for categories 1 and 2 that can be used for developing a machine learning protocol. This supervised machine learning protocol, however, would not work for category 3 "life as we don't know it," since training datasets, as well as a desired output, are fundamentally missing from Earth-based analogue samples. It may be possible, though, to identify this third category by exclusion of the first two categories. This effectively involves searching for levels of complexity that are incompatible with "life as we know it" (category 2) and with the absence of life (category 1). Thus, a complete machine learning protocol would start with a supervised algorithm 1 to find "no life." If it fails to find "no life," then it is possible that there is some form of life there. Supervised algorithm 2 would then be applied to find "life as we know it." If both algorithms fail, then the sample must fall into category 3 "life as we don't know it." An unsupervised machine learning algorithm can potentially be applied to find as of yet unidentified patterns that may be assigned as provisional biosignatures, which can then be searched for in other samples.

In general, it is not envisioned at this point that such work will be entirely dependent on these forms of artificial intelligence. At this time, an experienced human observer is

superior to a set of algorithms. However, the use of these machine learning methods in the treatment of large datasets could assist in finding patterns and focus attention on specific features of astrobiological interest. For instance, in a large set of close-up images of martian sediments, it may be useful to have a supervised machine learning program with a category 1 algorithm for “no life” that has been trained in grouping crystal types, grain sizes, and distinguishing common sedimentary patterns, followed by a category 2 algorithm for “life as we know it” that checks for a list of known biosignatures. A human observer can then discard any “non-life” data and focus entirely on samples identified with “life as we know it” or define subsets of data that can be studied for “life as we don’t know it.” The most important application of machine learning for the sample safety assessment may be data reduction and generation of data sub-sets for subsequent study by the science and safety assessment teams.

5. Conclusions

The SSAF would be incomplete without pointing out the importance of transparent and professional risk communication (*e.g.*, ESF, 2012). The information presented and the risk perception of the various stakeholders will evolve over time and might be influenced by events that have nothing to do with space exploration. It is therefore crucial to re-evaluate assumptions and strategies described in this SSAF on a regular basis, and to communicate the results of this re-evaluation process in a timely manner in order to build trust and preserve the sovereignty of information. A robust quality control program (see Section 4.1) is a fundamental prerequisite to achieve this aim.

The following is a summary that captures the major elements of the COSPAR Sample Safety Assessment Framework (SSAF):

The Sample Safety Assessment Framework (SSAF) Safety Approach

1. The **objective** of the sample safety assessment is to evaluate whether there is martian life present in samples intentionally returned from Mars that could pose a hazard for Earth’s systems.
2. Conducting a comprehensive safety assessment with the required rigor to predict harmful or harmless consequences for Earth is not feasible. Therefore, the **scope** of the SSAF is limited to evaluating whether the presence of martian life can be excluded in the samples. Any possible hazard is only considered in the sense that if there is no martian life, there is no extraterrestrial biological hazard in the samples. Occupational hazards are not covered by the SSAF.
3. The SSAF is starting from the **positive hypothesis** that there is martian life in the samples. Testing this hypothesis, *i.e.*, excluding the presence of martian life, is complementary to the scientific objective to search for martian life. Science and sample safety assessments use the same scientific methodologies though the purpose and associated burden of proof is reversed. Therefore, the SSAF is not a life detection framework.
4. The structure of the sample safety assessment is composed of **four elements**:

- (a) **Bayesian statistics**, a standard approach for addressing complex statistical issues and decision making.
 - (b) **Subsampling strategy**, tailored to the specific sample, to optimize the probability that the subsamples contain evidence of martian life, if there is martian life somewhere in the sample.
 - (c) **Test sequence**, a set of investigations in a specific logical order.
 - (d) **Decision criteria**, associated with a pre-defined level of assurance.
5. Key **assumptions** in the test sequence are that potential martian life is based on carbon chemistry and soluble or dispersible in polar or non-polar solvents.
 - (a) Organic molecules are defined as a group of covalently bonded molecules that contain carbon and at least one more element.
 - (b) Macromolecules are defined as organic compounds greater than 2500 Daltons.
 6. The **investigations** that are part of the SSAF must be able to detect evidence of self-replicating biological entities (*e.g.*, cell-like), biological entities that are replicated by other life (*e.g.*, virus-like), and biologically active molecules (*e.g.*, prion-like, gene transfer agent (GTA)-like molecules).
 - (a) The SSAF must include two or more orthogonal **agnostic life detection** investigations, with amplification steps.
 - (b) Investigations that lead to safety-critical decisions must be carried out by two **independent teams**, after which decisions are made by a third independent decision-making group.
 - (c) The conduct of tests that are part of the safety assessment must comply with ISO 17025, or equivalent **quality standards**, and apply GLP and HACCP methods to demonstrate the required competence and quality control.
 7. The test sequence, using a stepwise approach from more chemistry-based investigations (*e.g.*, organic molecules, molecular patterns and macromolecules) to more biologically based investigations (*e.g.*, life as we know it, life as we don’t know it), must cover both **common and unique** features of the samples.
 8. The **level of assurance** needed to declare a sample safe must be specified by the appropriate regulatory authority and incorporated into the SSAF.
 - (a) If evidence of extinct or extant martian life is detected, a **Hold and Critical Review (HCR)** must be established to evaluate the relevant data and the risk management measures before deciding on the next steps.
 - (b) No samples can be released from containment during the HCR and a procedure must be developed for samples already released from containment.

- (c) The absence of detecting organic compounds, molecular patterns, and macromolecules is not sufficient to declare a sample safe (*i.e.*, devoid of martian life).
 - (d) The positive hypothesis (*i.e.*, there is martian life in the samples) can be rejected if there is no evidence for the presence of martian life in the samples and there are no open, uncertain or ambiguous issues remaining that could associate sample characteristics to martian biology. In such cases, the tested sample of a sample tube would be safe within the pre-defined level of assurance.
9. The sample safety assessment is not a one-time exercise but rather a **dynamic process** that must respond to the results of various investigations. It must be updated if a subsequent investigation on any of the martian samples invalidates the original sample safety assessment or any of the assumptions used.

Execution

- 10. Every **sample-tube** is considered a separate sample.
- 11. Bayesian statistics together with the subsampling strategy and the sensitivity and specificity of the overall test sequence, allow estimating the **number of subsamples** necessary to reach a pre-defined level of assurance that a sample tube is safe.
- 12. The sample with the highest pre-test probability to contain martian life provides the best and most economic (time and material) **starting point** for executing the sample safety assessment.
- 13. A **targeted subsampling strategy** must be used to optimize the number of subsamples from one sample tube that need to be tested with the safety assessment. Three elements are required to develop such a strategy:
 - (a) Information about the **3-dimensional morphological characteristics** of the external and internal structures of each sample at a micrometer-level spatial resolution, while still in the sealed sample tube, is the required basis for planning and executing the sample safety assessment in general, and the subsampling strategy in particular.
 - (b) Information about the **chemistry and mineralogy** associated with the 3-D structure to refine the targeted subsampling strategy.
 - (c) An analogue test program to correlate the specific martian sample information to a relevant terrestrial sample knowledge base.
- 14. Depending on the type of **fine-grained minerals**, targeted subsampling (*e.g.*, for localized fine-grained alteration products or localized features, such as fractures in lithified fine-grained rocks) and random subsampling (for unconsolidated fine-grained sediments) are appropriate approaches.

- 15. Random sampling can be applied to **dust samples** though it is unlikely that any serendipitous dust would be of a sufficient quantity that it can be declared safe based on a safety assessment, except for a dedicated dust sample.

Development needs

- 16. An **analogue test program** is necessary to:
 - (a) Assess and improve the capture rate and the associated subsampling strategy.
 - (b) Optimize the selection of the instrument-ensemble to be used for the test sequence and estimate the overall sensitivity and specificity of the test sequence, including efficiency to extract evidence of life from host materials.
 - (c) Exercise all elements of the sample safety assessment, including blind testing, to optimize processes, equipment and infrastructure, train personnel and science teams, and build confidence.
- 17. **Contamination Knowledge (CK)** covering all flight (Mars 2020, MSR program) and ground (SRF) elements is critically important to reduce uncertainty in the interpretation of the data and, as a consequence, avoid unnecessary rigor in handling and treating the samples.
- 18. Use of **machine learning** to support the sample safety assessment, in particular for data reduction and pattern recognition in large and diverse datasets, has the potential to improve the quality of the results and accelerate the process.
- 19. Once the MSR science investigations are selected, the appropriate regulatory authorities are in place, and any open development needs with respect to the overall sample safety assessment are addressed, this SSAF must be critically reviewed by the relevant stakeholders. The latest applicable version of the SSAF would be the basis for developing a detailed **Sample Safety Assessment Protocol (SSAP)**.
- 20. A transparent risk communication throughout the development and execution of the SSAF and subsequent SSAP is essential to preserve the sovereignty of information.

There are a number of activities that need to feed into the SSAF and some consequences of the SSAF that would need to be reflected in future sample science plans (see Table 4).

The most important near-term **Research and Development (R&D)** activities to enable the preparation and execution of the SSAF are:

- 1. Establishing an analogue test program to inform and improve the capture rate, extraction efficiency, sensitivity and specificity of the overall test sequence, and exercise the entire sample safety assessment before it is used on samples returned from Mars.
- 2. Maturing agnostic life detection techniques.

Targeted investments in developing tailored machine learning capabilities to support the data reduction and data cross-correlations is considered beneficial to optimize the time economy once the samples start to enter the curation and science analysis stage. The development of such machine learning tools would need to be integrated in the analogue test program.

The only impact identified for the **Mars 2020 mission and the MSR program** is the need to provide Contamination Knowledge (CK) from all relevant mission phases (ground and flight) and mission elements with a potential to introduce terrestrial contamination to the samples during nominal and off-nominal events. This is considered critical for the interpretation of the data used for the safety assessment and any subsequent decisions. This CK is directly linked to the achievable specificity of the test sequence and to rectify events that would lead to a Hold & Critical Review. Therefore, the CK is an important element and driver in

the schedule of sample analysis and consequences that could lead to an unnecessary rigor in handling and treating the samples.

To optimize the use of precious martian samples and remain aligned with the stated goal to use the scientific investigations of competitively selected science teams to inform the sample safety assessment, a number of elements need to be considered for planning the future **selection of science teams to cover the objective-driven science for MSR**:

- Investigations described in the test sequence.
- The need for independent analyses for certain investigations (*i.e.*, more than one science team working on certain investigations).
- Optimizing the overall sensitivity and specificity of the test sequence (*i.e.*, consideration of using a complementary instrument-assemble with known sensitivities and specificities).

TABLE 4. CONSEQUENCES OF THE SSAF FOR VARIOUS ELEMENTS OF THE MSR CAMPAIGN

SSAF #	Impact on Safety Approach	Impact on Mars 2020 and MSR Program	Impact on Science Calls	Impact on Execution	R&D required
1	X				
2	X				
3	X				
4	X		X	X	X
5	X				
6	X		X	X	X
7	X			X	X
8	X		X	X	
9	X		X	X	
10	X		X	X	
11	X		X	X	X
12	X			X	X
13	X		X	X	X
14	X			X	X
15	X			X	X
16	X		X	X	X
17		X		X	
18				X	X
19	X		X	X	
20	X	X		X	

If all elements cannot be satisfied in the course of the science team selection, then directed investigations to fill in the gaps would need to be considered by the MSR Campaign Partners.

For all practical purposes, the sample properties that need to be measured to inform the SSAF fall under the sterilization-sensitive and time-critical categories as defined by MSPG2 (Velbel *et al.*, 2021; Tosca *et al.*, 2021). This means that most of these investigations would need to be conducted within biological containment, *i.e.*, a **Sample Receiving Facility** (SRF).

The Sample Safety Assessment Framework (SSAF) has been established with sufficient detail to allow for proper planning for a Sample Receiving Facility (SRF) and preparations for the scientific analysis of the samples. At the same time, the SSAF avoids being overly prescriptive. The SSAF is using an iterative approach to risk, combining multiple types of data and analyses, to derive an evidence-based safety assessment. As long as martian life is based on carbon chemistry, the SSAF and the subsequent SSAP would be able to identify it. The one parameter that must be set by the appropriate regulatory authorities is the **level of assurance** to exclude the presence of martian life. This would be the stopping threshold, *i.e.*, level of confidence in the statement “there is no martian life in the sample”. Setting such a level is important to avoid open-ended discussions and to better estimate the efforts and resources necessary to conduct the sample safety assessment.

Once the MSR science investigations are selected, the appropriate regulatory authorities are in place, and any open development needs with respect to the overall safety assessment are addressed, this SSAF must be critically reviewed by the relevant stakeholders. The resulting updated version of the SSAF would be the basis for developing a detailed Sample Safety Assessment Protocol (SSAP). COSPAR would provide an appropriate international forum to review the SSAF and develop the SSAP.

The SSAF is developed specifically for assessing samples from Mars in the context of the currently planned NASA-ESA MSR Campaign (Meyer *et al.*, 2021) though it can actually be used for any Mars Sample Return mission concept, with only minor tailoring. This minor tailoring would be required for the following aspects of the SSAF:

- Representing the specificity of sample type, acquisition, and packaging, reflected in point 10 of the SSAF.
- Representing the necessary CK of the applicable flight and ground elements, reflected in point 17 of the SSAF.

In addition, the SSAF is considered a sound basis for other COSPAR Planetary Protection Category V, restricted Earth return, mission concepts beyond Mars.

Acknowledgments

The COSPAR Sample Safety Assessment Framework (SSAF) has been developed by the COSPAR Sample Safety Assessment Protocol Working Group during the COVID-19 pandemic. GK would like to thank all members of the Working Group for their commitment and dedication during this challenging time and in particular, those that were involved in public health response activities throughout the drafting of this document.

We would like to thank Hayley Jones, Hugo Pedder, Chin Yang Shapland (all University of Bristol, UK), and Linda Sharples (London School of Hygiene & Tropical Medicine) for their comments on the statistical section of this paper.

We would like to thank the US Centers for Disease Control and Prevention (CDC) team—David Lowe, Melissa Brykailo Pearce, Samuel Edwin, Matthew J. Arduino, Brandon Hatcher and Jennifer McQuiston—for their review and comments during and at the end of preparing this document.

We would also like to thank Svetla Tsoleva (European Centre for Disease Prevention and Control—ECDC), Mika Salminen (Finnish Institute for Health and Welfare), Dave W. Beaty and Brandi L. Carrier (both Jet Propulsion Laboratory, California Inst. of Technology), the COSPAR Panel on Planetary Protection leadership, Athena Coustenis and Niklas Hedman, and Pericles D. Stabekis (Independent Consultant) for reviewing and commenting on the final version of this document.

We thank COSPAR, NASA, ESA, JAXA, DLR, the Swedish National Space Agency and UK Research and Innovation (UKRI) for financial support.

Unfortunately, Tullis C. Onstott (Princeton University) passed away during the preparation of this document. He was a very active member in the Working Group and we all are grateful for the opportunity to have worked with him.

References

- Aggarwal R, Sounderajah V, Martin G, *et al.* (2021) Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *npj Digit Med* 4, doi:10.1038/s41746-021-00438-z.
- Allwood A, Beaty D, Deborah B, *et al.* (2013) Conference summary: life detection in extraterrestrial samples. *Astrobiology* 13:203–216.
- Avila-Herrera A, Thissen J, Urbaniak C, *et al.* (2020) Crew-member microbiome may influence microbial composition of ISS habitable surfaces. *PLoS One* 15, doi:10.1371/journal.pone.0231838.
- Banci L and Bertini I (2013) Metallomics and the cell: some definitions and general comments. In *Metallomics and the Cell*, Metal Ions in Life Sciences Vol. 12, edited by L Banci. Springer, Dordrecht, doi:10.1007/978-94-007-5561-1_1.
- Barnes JJ, McCubbin FM, Santos AR, *et al.* (2020) Multiple early-formed water reservoirs in the interior of Mars. *Nat Geosci* 13:260–264.
- Berry D, Mader E, Lee TK, *et al.* (2015) Tracking heavy water (D₂O) incorporation for identifying and sorting active microbial cells. *Proc Natl Acad Sci USA* 112:E194–E203.
- Bhartia R, Salas EC, Hug WF, *et al.* (2010) Label-free bacterial imaging with deep-UV-laser-induced native fluorescence. *Appl Environ Microbiol* 76:7231–7237.
- Bijlani S, Stephens E, Singh NK, *et al.* (2021) Advances in space microbiology. *IScience* 24, doi:10.1016/j.isci.2021.102395.
- Bodzon-Kulakowska A and Suder P (2016) Imaging mass spectrometry: instrumentation, applications, and combination with other visualization techniques. *Mass Spectrom Rev* 35:147–169.
- Boschker HTS and Middelburg JJ (2002) Stable isotopes and biomarkers in microbial ecology. *FEMS Microbiol Ecol* 40: 85–95.
- Bradley JP, Harvey RP, McSween HY, *et al.* (1997) No ‘nanofossils’ in martian meteorite. *Nature* 390:454–455.
- Brady AL, Gibbons E, Sehlke A, *et al.* (2020) Microbial community distribution in variously altered basalts: insights

- into astrobiology sample site selection. *Planet Space Sci* 194, doi:10.1016/j.pss.2020.105107.
- Brandes N and Linial M (2019) Giant viruses—big surprises. *Viruses* 11, doi:10.3390/v11050404.
- Branton D, Deamer DW, Marziali A, *et al.* (2008) The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26: 1146–1163.
- Brasier MD and Wacey D (2012) Fossils and astrobiology: new protocols for cell evolution in deep time. *Int J Astrobiol* 11: 217–228.
- Brasier MD, Green OR, Jephcoat AP, *et al.* (2002) Questioning the evidence for Earth's oldest fossils. *Nature* 416:76–81.
- Braun RM, Beyder A, Xu J, *et al.* (1999) Spatially resolved detection of attomole quantities of organic molecules localized in picoliter vials using time-of-flight secondary ion mass spectrometry. *Anal Chem* 71:3318–3324.
- Braun S, Morono Y, Becker KW, *et al.* (2016) Cellular content of biomolecules in sub-seafloor microbial communities. *Geochim Cosmochim Acta* 188:330–351.
- Buick R (1990) Microfossil recognition in Archean rocks: an appraisal of spheruloids and filaments from a 3500 m.y. old chert-barite unit at North Pole, Western Australia. *Palaios* 5: 441–459.
- Carrier BL, Beaty DW, Hutzler A, *et al.* (2022) Science and curation considerations for the design of a Mars Sample Return (MSR) Sample Receiving Facility (SRF). *Astrobiology* 22, doi:10.1089/ast.2021.0110.
- Carr CE, Mojarro A, Hachey J, *et al.* (2017) Towards *in situ* sequencing for life detection. In *2017 IEEE Aerospace Conference*. IEEE, Piscataway, NJ, doi:10.1109/AERO.2017.794389.
- Casertano S, Lattanzi MG, Sozzetti A, *et al.* (2008) Double-blind test program for astrometric planet detection with Gaia. *Astron Astrophys* 482:699–729.
- Chan MA, Hinman NW, Potter-McIntyre SL, *et al.* (2019) Deciphering biosignatures in planetary contexts. *Astrobiology* 19:1075–1102.
- Chan QHS, Stroud R, Martins Z, *et al.* (2020) Concerns of organic contamination for sample return space missions. *Space Sci Rev* 216, doi:10.1007/s11214-020-00678-7.
- Chandran H, Meena M, and Sharma K (2020) Microbial biodiversity and bioremediation assessment through omics approaches. *Front Environ Chem* 1, doi:10.3389/fenvc.2020.570326.
- Chen LH, Kenyon GL, Curtin F, *et al.* (1992) 4-Oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer. *J Biol Chem* 267:17716–17721.
- Claus G and Nagy B (1961) A microbiological examination of some carbonaceous chondrites. *Nature* 192:594–596.
- Cleaves HJ, Crapster-Pregont E, Jonsson CM, *et al.* (2011) The adsorption of short single-stranded DNA oligomers to mineral surfaces. *Chemosphere* 83:1560–1567.
- Cockell CS, McMahon S, Lim DSS, *et al.* (2019) Sample collection and return from Mars: optimising sample collection based on the microbial ecology of terrestrial volcanic environments. *Space Science Review* 215, doi:10.1007/s11214-019-0609-7.
- Cosmidis J and Templeton AS (2016) Self-assembly of biomorphic carbon/sulfur microstructures in sulfidic environments. *Nat Commun* 7, doi:10.1038/ncomms12812.
- COSPAR (2021) *COSPAR Policy on Planetary Protection*. Prepared by the COSPAR Panel on Planetary Protection and approved by the COSPAR Bureau on 3 June 2021. Available online at https://cosparhq.cnes.fr/assets/uploads/2021/07/PPPPolicy_2021_3-June.pdf
- Craven E, Winters M, Smith AL, *et al.* (2021) Biological safety in the context of backward planetary protection and Mars Sample Return: conclusions from the Sterilization Working Group. *Int J Astrobiol* 20:1–28.
- Des Marais DJ, Nuth JA III, Allamandola LJ, *et al.* (2008) The NASA Astrobiology Roadmap. *Astrobiology* 8:715–730.
- DeVincenzi DL, Race MS, and Klein HP (1998) Planetary protection, sample return missions and Mars exploration: history, status, and future needs. *J Geophys Res* 103:28577–28585.
- Dworkin JP, Deamer DW, Sandford SA, *et al.* (2001) Self-assembling amphiphilic molecules: synthesis in simulated interstellar/precometary ices. *Proc Natl Acad Sci USA* 98:815–819.
- Eigen M and Schuster P (1977) The hypercycle: a principle of natural self-organization. *Naturwissenschaften* 64:541–565.
- Emerson JB, Adams RI, Román CMB, *et al.* (2017) Schrödinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome* 5, doi: 10.1186/s40168-017-0285-3.
- European Science Foundation (2012) *Mars Sample Return Backward Contamination—Strategic Advice and Requirements*. European Science Foundation, Strasbourg.
- Evans AA (2014) On the importance of blind testing in archeological science: the example from lithic functional studies. *Journal of Archeological Science* 48:5–14.
- Farley KA, Williford KH, Stack KM, *et al.* (2020) Mars 2020 mission overview. *Space Sci Rev* 216, doi:10.1007/s11214-020-00762-y.
- Fitch FW and Anders E (1963) Observations on the Nature of the “organized elements” in carbonaceous chondrites. *Ann NY Acad Sci* 108:495–513.
- Franchi IA, Wright IP, Sexton AS, *et al.* (1999) The oxygen-isotopic composition of Earth and Mars. *Meteorit Planet Sci* 34:657–661.
- Frank KL, Rogers DR, Olins HC, *et al.* (2013) Characterizing the distribution and rates of microbial sulfate reduction at Middle Valley hydrothermal vents. *ISME J* 7:1391–1401.
- Franz HB, McAdam AC, Ming DW, *et al.* (2017) Large sulfur isotope fractionations in martian sediments at Gale crater. *Nat Geosci* 10:658–662.
- Füri E and Marty B (2015) Nitrogen isotope variations in the Solar System. *Nat Geosci* 8:515–522.
- Garalde DR, Snell EA, Jachimowicz D, *et al.* (2018) Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods* 15:201–206.
- García-Ruiz JM, Hyde ST, Carerup AM, *et al.* (2003) Self-assembled silica-carbonate structures and detection of ancient microfossils. *Science* 302:1194–1197.
- García-Ruiz JM, Melero-García E, and Hyde ST (2009) Morphogenesis of self-assembled nanocrystalline materials of barium carbonate and silica. *Science* 323:362–365.
- Gazit E (2007) Self-assembly of short aromatic peptides into amyloid fibrils and related nanostructures. *Prion* 1:32–35.
- Ginsburg RN (1997) Perspective on the blind test. *Marine Micropaleontology* 29:65–103.
- Glavin DP, Burton AS, Elsila JE, *et al.* (2019) The search for chiral asymmetry as a potential biosignature in our solar system. *Chem Rev* 120:4660–4689.
- Gorbushina AA (2007) Life on the rocks. *Environ Microbiol* 9: 1613–1631.
- Graham HV, Meadows VS, co-chairs (2021) *Report of the 2021 Standards of Evidence in Life Detection Workshop*. Available online at <https://www.nfold.org/standards-of-evidence-workshop>
- Green J, Hoehler T, Neveu M, *et al.* (2021) Call for a framework for reporting evidence for life beyond Earth. *Nature* 598:575–579.

- Greenland S (2021) Introduction to Bayesian statistics. In *Modern Epidemiology*, 4th ed., edited by TL Lash, TJ Vanderweele, S Haneuse, *et al.* Wolters Kluwer, Philadelphia, PA, pp 563–580.
- Haltigin T, Lange C, Mugnolo R, *et al.*, and the iMARS Working Group (2016) iMARS Phase 2: a draft mission architecture and science management plan for the return of samples from Mars. *Astrobiology* 18, doi:10.1089/ast.2018.29027.mars.
- Hastie T, Tibshirani R, and Friedman J (2017) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York.
- Heeren RM (2015) Getting the picture: the coming of age of imaging MS. *Int J Mass Spectrom* 377:672–680.
- Hinton G, Deng L, Yu D, *et al.* (2012) Deep neural networks for acoustic modelling in speech recognition. *IEEE Signal Processing Magazine* 29:82–97.
- Horneck G and Rabbow E (2007) Mutagenesis by outer space parameters other than cosmic rays. *Adv Space Res* 40:445–454.
- Hulebak KL and Schlosser W (2002) Hazard Analysis and Critical Control Point (HACCP) history and conceptual overview. *Risk Analysis* 22:547–552.
- Hunink MGM, Weinstein MC, Wittenberg E, *et al.* (2014) *Decision Making in Health and Medicine*. Cambridge University Press, Cambridge, UK.
- IEC 61882 (2016) *Hazard and Operability Studies (HAZOP Studies)—Application Guide*. International Electrotechnical Commission, Geneva, Switzerland.
- ISO 17025 (2017) *General Requirements for the Competence of Testing and Calibration Laboratories*. International Organization for Standardization, Geneva, Switzerland.
- Jardillier L, Zubkov MV, Pearman J, *et al.* (2010) Significant CO₂ fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* 4:1180–1192.
- Jayasena SD (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin Chem* 45:1628–1650.
- Jennings RdM, Moran JJ, Jay ZJ, *et al.* (2017) Integration of metagenomic and stable carbon isotope evidence reveals the extent and mechanisms of carbon dioxide fixation in high-temperature microbial communities. *Front Microbiol* 8, doi: 10.3389/fmicb.2017.00088.
- Johnson SS, Anslyn E, Graham HV, *et al.* (2018) Fingerprinting nonterran biosignatures. *Astrobiology* 18:915–922.
- Joyce GF (2002) The antiquity of RNA-based evolution. *Nature* 418, doi:10.1038/418214a.
- Jordan MI and Mitchell TM (2015) Machine learning: trends, perspectives, and prospects. *Science* 349:255–260.
- Jordan SF, Ramm H, Zheludev IN, *et al.* (2019) Promotion of protocell self-assembly from mixed amphiphiles at the origin of life. *Nat Ecol Evol* 3:1705–1714.
- Kempes CP and Krakauer DC (2021) The multiple paths to multiple life. *J Mol Evol* 89:415–426.
- Kminek G, Conley C, Allen CC, *et al.* (2014) Report of the workshop for life detection in samples from Mars. *Life Sci Space Res* 2:1–5.
- Kminek G, Meyer MA, Beaty DW, *et al.* (2022) Mars Sample Return (MSR): planning for returned sample science. *Astrobiology* 22, doi:10.1089/ast.2021.0198.
- Koskinen K, Rettberg P, Pukall R, *et al.* (2017) Microbial biodiversity assessment of the European Space Agency's ExoMars 2016 mission. *Microbiome* 5:143–143.
- Kotopoulou E, Lopez-Haro M, Calvino Gamez JJ, *et al.* (2021) Nanoscale anatomy of iron-silica self-organized membranes: implications for prebiotic chemistry. *Angewandte Chemie* 133:1416–1422.
- Krizhevsky A, Sutskever I, and Hinton G (2017) ImageNet Classification with deep convolutional neural networks. *Communications of the ACM* 60:84–90.
- Kronberg EA, Hannan T, Huthmacher J, *et al.* (2021) Prediction of soft proton intensities in the near-Earth space using machine learning. *Astrophys J* 921, doi:10.3847/1538-4357/ac1b30.
- Kumar AM, Ravindran B, Sarsaiya S, *et al.* (2020) Metagenomics for taxonomy profiling: tools and approaches. *Bioengineered* 11:356–374.
- La Duc MT, Venkateswaran K, and Conley CA (2014) A genetic inventory of spacecraft and associated surfaces. *Astrobiology* 14:15–23.
- Litchman E (2010) Invisible invaders: non-pathogenic invasive microbes in aquatic and terrestrial ecosystems. *Ecol Lett* 13: 1560–1572.
- Lloyd KG, Steen AD, Ladau J, *et al.* (2018) Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. *mSystems* 3, doi:10.1128/mSystems.00055-18.
- Madigan MT, Bender KS, Buckley DH, *et al.* (2019) *Brock Biology of Microorganisms*, Global Edition, 15th ed. Pearson, New York.
- Marshall CP, Emry JR, and Olcott Marshall A (2011) Haematite pseudomicrofossils present in the 3.5-billion-year-old Apex Chert. *Nat Geosci* 4:240–243.
- Marshall SM, Mathis C, Carrick E, *et al.* (2021) Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nat Commun* 12, doi:10.1038/s41467-021-23258-x.
- McCubbin FM, Herd CDK, Yada T, *et al.* (2019) Advanced curation of astromaterials for planetary science. *Space Sci Rev* 215, doi:10.1007/s11214-019-0615-9.
- McKay DS, Gibson EK, Thomas-Keprta L, *et al.* (1996) Search for past life on Mars: possible relic biogenic activity in martian meteorite ALH84001. *Science* 27:924–930.
- McLennan SM, Sephton MA, Allen C, *et al.* (2011) Planning for Mars returned sample science: final report of the MSR End-to-End International Science Analysis Group (E2E-iSAG). *Astrobiology* 12:175–230.
- McMahon S, Ivarsson M, Wacey D, *et al.* (2021) Dubiofossils from a Mars-analogue subsurface palaeoenvironment: the limits of biogenicity criteria. *Geobiology* 19:473–488.
- Meyer MA, Kminek G, Beaty DW, *et al.* (2022) Final Report of the Mars Sample Return Science Planning Group 2 (MSPG2). *Astrobiology* 22, doi:10.1089/ast.2021.0121.
- Milam S, Dworkin JP, Elsila JE, *et al.* (2021) Volatile sample return in the Solar System. Planetary Science and Astrobiology Decadal Survey. *Bulletin of the American Astronomical Society* 53, doi:10.3847/25c2cfed.d94f26c7.
- Mills LS, Soule ME, and Doak DF (1993) The keystone-species concept in ecology and conservation. *BioScience* 43:219–224.
- Minich JJ, Zhu Q, Janssen S, *et al.* (2018) KatharoSeq enables high-throughput microbiome analysis from low-biomass samples. *mSystems* 3, doi:10.1128/mSystems.00218-17.
- Mitra S (2004) *Sample Preparation Techniques in Analytical Chemistry*, Vol. 237. John Wiley & Sons, New York.
- Moeller R, Reitz G, Nicholson WL, *et al.* (2012) Mutagenesis in bacterial spores exposed to space and simulated martian conditions: data from the EXPOSE-E spaceflight experiment PROTECT. *Astrobiology* 12:457–468.
- Moissl-Eichinger C, Auerbach AK, Probst AJ, *et al.* (2015) Quo vadis? Microbial profiling revealed strong effects of clean-room maintenance and routes of contamination in indoor environments. *Sci Rep* 5, doi:10.1038/srep09156.

- Moore LD, Le T, and Fan G (2013) DNA methylation and its basic function. *Neuropsychopharmacology* 38:23–38.
- Mounicou S, Szpunar J, and Lobinski R (2009) Metallomics: the concept and methodology. *Chem Soc Rev* 38:1119–1138.
- Muscante AD, Czaja AD, Tuggle J, *et al.* (2018) Manganese oxides resembling microbial fabrics and their implications for recognizing inorganically preserved microfossils. *Astrobiology* 18:249–258.
- National Research Council (1997) *Mars Sample Return: Issues and Recommendations*. The National Academies Press, Washington, DC.
- National Research Council (2002) *The Quarantine and Certification of Martian Samples*. The National Academies Press, Washington, DC.
- National Research Council (2009) *Assessment of Planetary Protection Requirements for Mars Sample Return Missions*. The National Academies Press, Washington, DC.
- Neveu M, Hays LA, Voytek MA, *et al.* (2018) The ladder of life detection. *Astrobiology* 18:1375–1402.
- Nicholson WL, Moeller R, and Horneck G (2012) Transcriptomic responses of germinating *Bacillus subtilis* spores exposed to 1.5 years of space and simulated martian conditions on the EXPOSE-E experiment PROTECT. *Astrobiology* 12:469–486.
- Nie PF, Li XF, Ren YP, *et al.* (2015) Electricity generation from sulfide tailings using a double-chamber microbial fuel cell. *RSC Adv* 5, doi:10.1039/C5RA16459A.
- OECD Principles on Good Laboratory Practice (1998) *OECD Environmental Health and Safety Publications, Series on Principles of Good Laboratory Practice and Compliance Monitoring*, No 1, ENV/MC/CHEM(98)17. Organisation for Economic Co-operation and Development, Paris.
- Oehlschläger F and Eigen, M (1997) 30 years later—a new approach to Sol Spiegelman’s and Leslie Orgel’s *in vitro* evolutionary studies dedicated to Leslie Orgel on the occasion of his 70th birthday. *Orig Life Evol Biosph* 27:437–457.
- Onstott TC, Ehlmann BL, Sapers H, *et al.* (2019) Paleo-rock-hosted life on Earth and the search on Mars: a review and strategy for exploration. *Astrobiology* 19:1230–1262.
- Ouldali H, Sarthak K, Ensslen T, *et al.* (2020) Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat Biotechnol* 38:176–181.
- Parai R, Rodriguez J, Meshik A, *et al.* (2021) A gas extraction manifold for the Apollo 17 73001 core sample vacuum container [abstract 2548]. In *52nd Lunar and Planetary Science Conference*, LPI Contribution No. 2548. Lunar and Planetary Institute, Houston.
- Pearce N (2004) Public health and the precautionary principle. In *The Precautionary Principle: Protecting Public Health, the Environment and the Future of Our Children*, edited by M Martuzzi and JA Tickner. WHO Regional Office for Europe, Rome, pp 49–62.
- Pejchar L and Mooney HA (2009) Invasive species, ecosystem services and human well-being. *Trends Ecol Evol* 24:497–504.
- Petkowski JJ, Bains W, and Seager S (2020) On the potential of silicon as a building block of life. *Life* 10, doi:10.3390/life10060084.
- Phillips PJ, Yates AN, Hu Y, *et al.* (2018) Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proc Natl Acad Sci USA* 115:6171–6176.
- Pinheiro VB, Taylor AI, Cozens C, *et al.* (2012) Synthetic genetic polymers capable of heredity and evolution. *Science* 336:341–344.
- Rand AC, Jain M, Eizenga JM, *et al.* (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nat Methods* 14:411–413.
- Randolph SE and Rogers DJ (2010) The arrival, establishment and spread of exotic diseases: patterns and predictions. *Nat Rev Microbiol* 8:361–371.
- Regberg AB, Castro CL, Connolly HC, *et al.* (2020) Prokaryotic and fungal characterization of the facilities used to assemble, test, and launch the OSIRIS-REx spacecraft. *Front Microbiol*, 11, doi:10.3389/fmicb.2020.530661.
- Rinke C, Schwientek P, Sczyrba A, *et al.* (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437.
- Romero IC, Kong S, Fowlkes CC, *et al.* (2020) Improving the taxonomy of fossil pollen using convolutional neural networks and superresolution microscopy. *Proc Natl Acad Sci USA* 117:28496–28505.
- Rothman KJ and Lash TL (2021) Precision and study size. In *Modern Epidemiology*, 4th ed., edited by TL Lash, TJ Vanderweele, S Haneuse, *et al.* Wolters Kluwer, Philadelphia, pp 333–364.
- Rouillard J, Garcia-Ruiz J-M, Gong J, *et al.* (2018) A morphogram for silica-witherite biomorphs and its application to microfossil identification in the early Earth rock record. *Geobiology* 16:279–296.
- Rouillard J, Garcia-Ruiz J-M, Kah LC, *et al.* (2020) Identifying microbial life in rocks: insights from population morphometry. *Geobiology* 18:282–305.
- Rouillard J, van Zuilen MA, Pisapia C, *et al.* (2021) An alternative approach for assessing biogenicity. *Astrobiology* 21: 151–164.
- Rummel JD and Kminek G (2018) It’s time to develop a new “Draft Test Protocol” for a Mars Sample Return mission (or two...). *Astrobiology* 18:377–380.
- Rummel JD, Race MS, DeVincenzi DL, *et al.*, editors. (2002) *A Draft Test Protocol for Detecting Possible Biohazards in Martian Samples Returned to Earth*, NASA/CP-2002-211842. NASA Ames Research Center, Moffett Field, CA.
- Sabate R, Rousseau F, Schymkowitz J, *et al.* (2015) Amyloids or prions? That is the question. *Prion* 9:200–206.
- Saborio GP, Permanne B, Soto C, *et al.* (2001) Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* 411:810–813.
- Schaefer M, Kapoor U, and Jantsch MR (2017) Understanding RNA modifications: the promises and technological bottlenecks of the ‘epitranscriptome’. *Open Biology* 7, doi:10.1098/rsob.170077.
- Schopf JW (1993) Microfossils of the Early Archean Apex Chert: new evidence of the antiquity of life. *Science* 260:640–646.
- Schopf JW, Kudryavtsev AB, Sugitani K, *et al.* (2010) Precambrian microbe-like pseudofossils: a promising solution to the problem. *Precambrian Res* 179:191–205.
- Schopf JW, Kitajima K, Spicuzza MJ, *et al.* (2018) SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proc Natl Acad Sci USA* 115:53–58.
- Scott WG, Finch JT, and Klug A (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* 81:991–1002.
- Shaheen R, Niles PB, Chong K, *et al.* (2015) Carbonate formation events in ALH 84001 trace the evolution of the martian atmosphere. *Proc Natl Acad Sci USA* 112:336–341.
- Silveira JR, Raymond GJ, Hughson AG, *et al.* (2005) The most infectious prion protein particles. *Nature* 437:257–261.
- Simpson JT, Workman RE, Zuzarte PC, *et al.* (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 14:407–410.

- Sleiman D, Garcia PS, Lagune M, *et al.* (2021) A third purine biosynthetic pathway encoded by aminoadenin-based viral DNA genomes. *Science* 372:431–437.
- Summons RE, Sessions A, Allwood, A, *et al.* (2014) Planning considerations related to the organic contamination of martian samples and implications for the Mars 2020 rover. *Astrobiology* 14:969–1027.
- Suttle CA and Fuhrman JA (2010) MAVe: enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy. In *Manual of Aquatic Viral Ecology*, edited by SW Wilhelm, MG Weinbauer, and CA Suttle. ASLO, Waco, pp 145–153.
- Suzuki Y, Yamashita S, Kouduka M, *et al.* (2020) Deep microbial proliferation at the basalt interface in 33.5–104 million-year-old oceanic crust. *Commun Biol* 3, doi:10.1038/s42003-020-0860-1.
- Swindle TD, Atreya S, Busemann H, *et al.* (2022) Scientific value of including an atmospheric sample as part of Mars Sample Return (MSR). *Astrobiology* 22, doi:10.1089/ast.2021.0107.
- Tait KT, McCubbin FM, Smith CL, *et al.* (2022) Preliminary planning for Mars Sample Return (MSR) curation activities in a Sample Receiving Facility (SRF). *Astrobiology* 22, doi:10.1089/ast.2021.0105.
- Toporski J and Steele A (2007) Observations from a 4-year contamination study of a sample depth profile through martian meteorite Nakhla. *Astrobiology* 7:389–401.
- Tosca NJ, Beaty DW, Carrier BL, *et al.* (2022) Time-sensitive aspects of Mars Sample Return (MSR) science. *Astrobiology* 22, doi:10.1089/ast.2021.0115.
- Tykot, RH (2003) Stable isotopes and diet: you are what you eat. *International-School-of-Physics-Enrico-Fermi on Physics Methods in Archaeometry* 154:433–444.
- UN Space Treaty (1966) *Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and other Celestial Bodies*, UN General Assembly 21st Session, Resolution 2222 (XXI). United Nations, New York.
- Urey HC (1962) Lifelike forms in meteorites. *Science* 137:623–628.
- van der Putten WH, Klironomos JN, and Wardle DA (2007) Microbial ecology of biological invasions. *ISME J* 1:28–37.
- van Driel M, Ceylan S, Clinton JF, *et al.* (2019) Preparing for InSight: evaluation of the blind test for martian seismicity. *Seismological Research Letters* 90:1518–1534.
- Velbel MA, Beaty DW, Carrier BL, *et al.* (2022) Planning implications related to sterilization-sensitive science investigations associated with Mars Sample Return (MSR). *Astrobiology* 22, doi:10.1089/ast.2021.0113.
- Wacey D, Saunders M, Kong C, *et al.* (2015). 3.46 Ga Apex chert “microfossils” reinterpreted as mineral artefacts produced during phyllosilicate exfoliation. *Gondwana Res* 36:296–313.
- Watrous JD and Dorrestein PC (2011) Imaging mass spectrometry in microbiology. *Nat Rev Microbiol* 9:683–694.
- Woyke T, Doud FRD, and Schulz F (2017) The trajectory of microbial single-cell sequencing. *Nat Methods* 14:1045–1054.
- Ye BF, Zhao YJ, Cheng Y, *et al.* (2012) Colorimetric photonic hydrogel aptasensor for the screening of heavy metal ions. *Nanoscale* 4:5998–6003.
- Yu FB, Blainey CP, Schulz F, *et al.* (2017) Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples. *eLife* 6, doi:10.7554/eLife.26580.
- Zhou Y, Xu X, Wei Y, *et al.* (2021). A widespread pathway for substitution of adenine by diaminopurine in phage genomes. *Science* 372:512–516.

Address correspondence to:
 Gerhard Kminek
 European Space Agency
 Mars Exploration Group
 Noordwijk
 The Netherlands

E-mail: Gerhard.kminek@esa.int

Submitted 26 January 2022

Accepted 28 March 2022

Abbreviations Used

3-D	= three-dimensional
BC	= Basic Characterization
CK	= Contamination Knowledge
COSPAR	= Committee on Space Research
DNA	= deoxyribonucleic acid
ESA	= European Space Agency
ESn	= effective sensitivity
GC	= gas chromatography
GLP	= Good Laboratory Practice
HACCP	= Hazard Analysis at Critical Control Points
HCR	= Hold and Critical Review
MSPG2	= Mars Sample Return Science Planning Group 2
MSR	= Mars Sample Return
NASA	= National Aeronautics and Space Administration
NLR	= Negative Likelihood Ratio
NRC	= National Research Council
PCR	= polymerase chain reaction
PE	= Preliminary Examination
PLR	= Positive Likelihood Ratio
Pre-BC	= Pre-Basic Characterization
R&D	= Research & Development
RNA	= ribonucleic acid
Sn	= sensitivity
Sp	= specificity
SRF	= Sample Receiving Facility
SSAF	= Sample Safety Assessment Framework
SSAP	= Sample Safety Assessment Protocol
SSU rRNA	= small-subunit ribosomal RNA
XNAs	= xenonucleic acids