

IUPAC Recommendation

Jeremy G. Frey, Richard M. Hartshorn* and Leah R. McEwen

Specification of International Chemical Identifier (InChI) QR codes for linking labels on containers of chemical samples to digital resources (IUPAC Recommendations 2021)

<https://doi.org/10.1515/pac-2021-0604>

Received June 5, 2021; accepted October 20, 2022

Abstract: This article discusses the ways of linking physical objects to digital information relevant to chemical entities, specifically those that can be described by the use of the IUPAC International Chemical Identifier (InChI). It makes recommendations on the form of the computer readable components of labels provided for chemicals and materials that are used on product/sample containers and on the associated documentation that is used when transporting these containers (either internally or during export/import). The focus is on specification of the content of the 2D Quick Response bar codes required to describe the molecular content of the containers and link to digital resources to supplement that provided on a physical label. The necessary technical and (possible) business infrastructure necessary to support the use of the InChI and InChIKey for rapid recall of relevant information is considered here and suggestions are made.

Keywords: Chemoinformatics; InChI; InChIKey; International Chemical Identifier; IUPAC; QR code.

CONTENTS

1	Introduction	1196
1.1	Use Cases: Labels, Shipping, and Regulation	1196
1.2	What is on a label: the label content	1196
1.3	QR Codes	1198
2	Discussion	1198
2.1	International Chemical Identifier: Linking to Generally Available Digital Information on Chemicals, with Examples	1198
2.2	The InChIKey	1200
2.3	Organization-Specific QR InChI/InChIKey Use – Implementation in Commercial/User Applications	1201
2.4	QR Code Version and Appearance	1203
2.5	The Possible Supporting Infrastructure – The case/need for a General InChIKey Resolver	1203
3	InChI QR Code Specification	1203
3.1	Specification	1203

Article note: This manuscript was prepared in the framework of IUPAC project 2015-019-2-800 “Identifying International Chemical Identifier (InChI) Enhancements – QR codes and Industry Applications”; Sponsoring bodies: IUPAC Chemical Nomenclature and Structure Representation Division.

***Corresponding author: Richard M. Hartshorn**, School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand, e-mail: richard.hartshorn@canterbury.ac.nz

Jeremy G. Frey, School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

Leah R. McEwen, Physical Sciences Library, Cornell University, 293 Clark Hall, Ithaca, NY 14853-2501, USA

4 Future work for IUPAC and the InChI Trust	1204
5 Conclusions	1205
Membership of the sponsoring body	1205
References	1205

1 Introduction

The International Chemical Identifier (InChI) for a molecule, and its fixed-length derivative, the InChIKey, are textual representations that are derived from the chemical structure of that molecule using methods [1] that have been approved by the International Union of Pure and Applied Chemistry (IUPAC) [2] and are overseen by the InChI Trust [3]. The InChI can be decoded to produce a structure for the molecule, while the fixed length of the InChIKey makes it particularly well-suited for searching for entries for and references to that molecule within databases and other resources [4–6]. Both identifiers have considerable potential for use in linking digital information about the molecule, or particular samples of it, in ways that will facilitate ready access to health and safety information, or allow more detailed and specific tracking of both real samples and data obtained from them.

1.1 Use Cases: Labels, Shipping, and Regulation

Labels for chemicals are used on bottles, containers, and other packaging, to provide identification of the content to enable users to find out what is contained within, for inventory management purposes, and for legal and safety reasons. The latter two relate to the variety of rules and regulations that have to be followed in the production, purchase, shipping, and use of chemicals. Rapid access to more extensive information than can possibly be contained on a label, such as that in a Safety Data Sheet (SDS) [7], would be very useful in safety critical situations, custom checks, and related applications. Furthermore, a digital link to an SDS provides a mechanism whereby the latest version of the SDS can be accessed even if the sample is rather old (as shown in Fig. 1). While SDS documents are often shipped with bottles of chemicals, the label that identifies the bottle is more likely to remain with the bottle than the SDS sheet and be still available sometime after the materials have been packed/unpacked and moved around within an organization or operational site. Any such label also has the potential to be used as part of inventory management systems within an organization, particularly if it is both unique and digitally readable.

1.2 What is on a label: the label content

In this article, we discuss and make recommendations on the possible form of computer-readable components of labels provided for containers of chemicals. It is not the purpose of this article to recommend what items or information should be included on a label of chemicals when they are shipped. Much of the information on labels

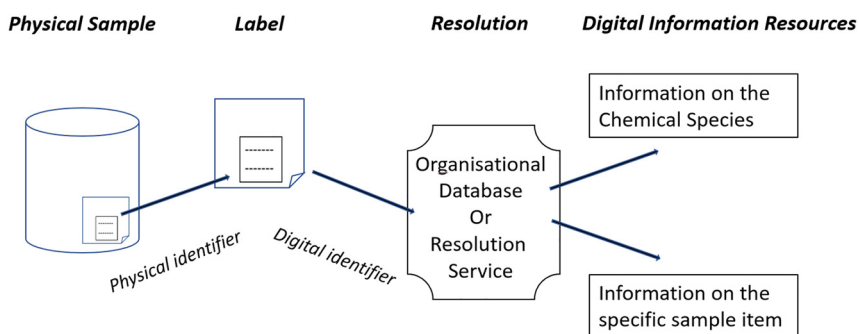


Fig. 1: Linking physical samples to digital resources through a digitally read label and an organizational database or resolution service.

is regulated, for example, by the requirements of the Globally Harmonized System (GHS) [8] for classification and labeling of chemicals, as implemented in different jurisdictions and other packaging and shipping regulations. Our aim is to enhance the linking of physical containers to relevant digital information about the contents of those containers and to provide additional tools for use in regulatory frameworks (e.g., REACH H – European Union regulation on Registration, Evaluation, Autorization and Restriction of Chemicals [9]). The regulations for labeling are, in fact, very strict and leave little room for optional additional information on a typical bottle label. While the documents shipped with the containers have more room for additional information, once the containers have been unpacked, the link between the paperwork and the container is very easily lost.

In a similar manner, we also sidestep the issue of exactly what chemical information should be listed on the container, and we assume that the regulation or custom of the particular chemical sector (organics, inorganics, biologicals, etc.) would provide a pathway to determine what the primary chemical species to be identified explicitly on the container should be. This could be the major component, or the major component that is not a solvent, an active pharmaceutical agent, or some combination of these. In each sector, regulations and practice determine which components of the mixture are highlighted on the labels. While this may not always be transparent to users from other sectors, it is not something we seek to solve here. We do note, however, that access to more extensive information on contents and history of an individual container is something that a digital link can enable, something many suppliers and manufacturers have already recognized.

In this article, we do, though, seek to make a recommendation about how the specific species chosen to be named on a label can be supplemented by a machine-readable digital link to digital information sources using the (IUPAC) International Chemical Identifier (InChI) and its related InChIKey [4–6, 10, 11].

The InChI is a character string that contains information on the composition of a compound, the connectivity of the atoms that make it up, and a range of additional structural information. The lengths of InChI strings vary and can be very long for complicated molecules, but they can be used by appropriate software to draw the structure of the molecule encoded in the string. Unfortunately, for many applications, such as searching of databases and websites, such long and variable length strings can be difficult or impossible to handle. A shortened version, the InChIKey, was developed for such applications, but the trade-off for more convenient dimensions is that the InChIKey can no longer be used to produce the structure directly. Instead, it is necessary to provide a database or resolver that links the InChIKey to the original InChI string. Much more detail on the InChI can be found on the InChI Trust website [3] and articles describing its use [10, 12–15].

It is our aim that the recommendations will apply to chemicals for which it is appropriate to assign a single name on the label and is a compound which can be represented by an InChI. At the time of writing, InChIs are available for organic compounds only, although work is underway to extend the InChI to coordination compounds and organometallic species [16], large biomolecules [17], and polymers [18]. Other areas of InChI development include improvements in ways that the InChI deals with tautomers [19] and stereo-configurations [20].

In addition to the InChI for a specific (currently organic) defined chemical structure, the International Union of Pure and Applied Chemistry (IUPAC) and the InChI Trust are working on versions of the International Chemical Identifier (InChI) that can cope with and describe the kinds of mixtures that are often found in real chemical samples [21]. We anticipate that the recommendations in this document will also be relevant to, and be able to make use of, the InChI application for mixtures (the MInChI and the related MInChIKey) [22].

In this article, we propose the way the InChI (or more probably, the InChIKey) should be used within a two-dimensional bar type code, the Quick Response (QR) code, more commonly known as a QR code [23], to provide a pathway to identify the molecule and facilitate accessing digital resources for the molecule and its properties. More specifically, we recommend a standard that can be used by chemical suppliers and users to produce QR codes that link to digital resources that are maintained by their own (or other established) organizations.

Our recommendation concerns primarily the use of 2D QR bar codes [20] in a way similar, but in a more networked and connected manner, to the industry-wide use of 1D bar codes. Key aspects of our recommendations are also applicable to other systems that link physical and digital objects, such as RFID (Radio Frequency Identification) tags. The issues are likely to be centered around how much information can be held by the coded systems and what is needed to resolve any identifiers used (see later in Section 2.3 for more explanation on this key point for actual implementation of the recommendations).

1.3 QR Codes

QR codes are ubiquitous in modern society, not the least because of their widespread use in marketing campaigns and other situations where an organization wishes to generate traffic for a particular website. Quick Response codes, coupled with appropriate code reader apps, have proven their ability to rapidly provide the URL (Uniform Resource Locator) for a website and possibly also other useful information that will allow a user to interact with material on that website, or perhaps to engage with augmented reality systems. This ability will be harnessed in this InChI application, as we seek to provide users with a way to rapidly access information about the chemical/material in the QR code-labeled container, usually from a website. That website may be that of the chemical supplier, the receiving organization, or perhaps a third party that maintains databases of chemical information.

QR codes were originally used in the car industry as a technology to track parts, particularly in cases where rapid reading is required, wherefrom comes the name QR, for Quick Response. This original application was closely tied to inventory control, and one of the key use cases for the InChI-related QR application is in inventory control and tracking within organizations that wish to have electronic records of their chemical assets.

The QR code, like any bar code system, encodes textual information in a graphical way that is rapidly and easily read by a digital image sensor. The squares at the corners of the QR code are used to orient and scale the code, which can then be converted to binary numbers, or any derived set of characters, which are then used in the way determined by the reader software (often to take the user to a URL). Clearly, the number of characters that can be included within a QR code depends on its size, type (there are numerous versions available), and standards for the application that determine the level of error correction that might be required.

It is, therefore, very important to consider the nature of the information to be included in the InChI QR code before making any recommendations on the size and version of code that should be used (see below).

2 Discussion

2.1 International Chemical Identifier: Linking to Generally Available Digital Information on Chemicals, with Examples

To set the scene, we consider the generalized case as an example of what we strive to achieve in linking the label to information resources. The QR code ideally would contain the standard InChI string, for example, for bottles labeled as aspirin, the InChI string is shown in Fig. 2, and the related QR code would enable an appropriate app to interpret this as the chemical structure for aspirin or link out to web resources describing aspirin. Figure 2 also shows the general form for the QR InChI that we recommend with the acronym IUPAC above and InChI below the QR square. The acronyms are added as a way to identify the purpose of the QR code. The size and resolution of the QR code image will depend on the amount of information that is being carried by the code and the level of redundancy (and therefore reliability in reading) that is required.



InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)

Fig. 2: QR code for the InChI for Aspirin. If viewed with a general QR code reader on a smart phone, the text shown here should be reported.



Fig. 3: ChemSpider InChI link on the Aspirin web page and a more general search via Google. A QR code reader would obtain this string and then offer to follow the URL which in this case will initiate the Google search.

To enable a general QR code reader to make best use of this InChI QR code, the actual message could be like that used by the Royal Society of Chemistry ChemSpider resource [24] in its InChI link given on the molecule's web page <http://www.chemspider.com/Chemical-Structure.2157.html>, or a more general Google search which is given by the search URL shown in Fig. 3. Such an approach is recommended to reduce the possibility of software misinterpreting parts of the string (for example, treating sequences of digits as phone numbers or something similar).

Alternative searches such as Wikipedia would be similarly useful. A specific QR-InChI app for a smart phone would be able to recognize the InChI string and offer the user a variety of searches and even render the structure direct from the InChI. The latter capability would make such a labeling convention superior to many other systems that have been used to identify compounds using some kind of numerical code.

However, this approach is ultimately problematic, as the number of characters that can be represented by a typical QR code is limited (while still depending on the particular version of the QR code being used, see below), and the standard InChI can be a very long string. The average length of InChI strings calculated for a real collection of approximately 10 million records from PubChem is 146 characters [25]; however, for the molecules in PubChem that have an InChI, >99 % have an InChI that is less than 300 characters in length. Figure 4 shows a histogram of InChI lengths from PubChem. It follows that the number of compounds that

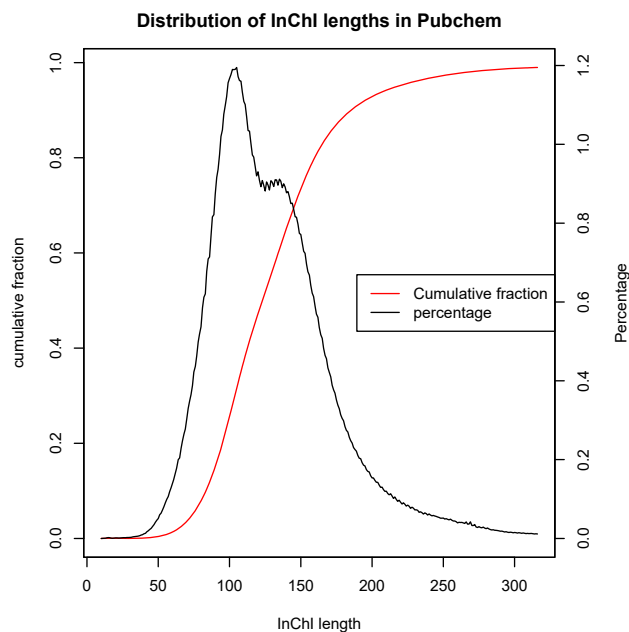


Fig. 4: The black curve is the frequency distribution (smoothed histogram) as a percentage of the length of the InChI of the 96 million molecules from PubChem in 2018 and the red curve is the cumulative fraction.

InChIKey=BSYNRYMUTXBXSQ-UHFFFAOYSA-N

Fig. 5: InChIKey for Aspirin.

could be covered by a small QR-InChI would be limited. However, for small molecules and those where only the first few layers of the InChI are necessary, it is possible and, indeed, perhaps desirable for the QR-InChI to contain the actual InChI, as this requires no external lookup to be interpreted.

2.2 The InChIKey

A potential solution to the large and variable length of the InChI lies in the use of the InChIKey. Other applications of InChI, for example, in databases, have made use of the InChIKey. The InChIKey itself is a hash of the InChI, so it always has the same fixed length (27 characters) and format, which easily fit within the QR code restrictions. The standard InChIKey for aspirin is shown in Fig. 5.

Thirty-six characters are therefore needed for the string required in the InChIKey QR code, and for the complete URL of the forms discussed below, another 35–45 characters will be required, depending on the string needed for the base URL for the search site or company site. For the basic system, we suggest that a QR code containing 100 characters should be sufficient – the table in the QR code website [26] indicates the range of options for the QR code to hold this amount of information.

While QR codes are relatively resilient to misreading, we would still recommend at least a medium level of data correction. However, given the potential problems arising from an incorrect link, higher levels of error correction might be appropriate for chemical labeling. However, the higher-level codes do require a larger physical printed code and/or higher-resolution printing and imaging, so there will inevitably be a need for compromise here, especially for small containers. Of course, this will not be a problem for the associated shipping documentation, and, in such circumstances, it may be necessary and desirable to have larger QR codes, containing more information and/or the full InChI in the documentation while using a smaller QR code on a small container.

As suggested above, the actual string stored in the QR code could be the relevant search (for the exact structure), but as most QR Code readers will automatically offer the opportunity to search using one of the common search engines, the QR Code could merely contain the InChI or InChIKey string (in most cases, the form of the InChIKey is so distinctive that it will lead to a satisfactory search, even without it being identified as an InChIKey). This is not encouraged, and we recommend there be a clear identifier of the nature of the following string, as shown in Fig. 6. We also note that the stereochemical information contained in the second part for the InChIKey will always be UHFFFAOYSA for a molecule without stereochemical complexity (e.g., 2-acetoxybenzoic



InChIKey=BSYNRYMUTXBXSQ-UHFFFAOYSA-N

Fig. 6: QR code with the InChIKey=BSYNRYMUTXBXSQ-UHFFFAOYSA-N.



Fig. 7: QR InChIKey for caffeine in ChemSpider <http://www.chemspider.com/inchikey=RYYVLZVUVIJVGH-UHFFFAOYSA-N>.

acid is stereochemically “flat” if the hydrogen atoms of the methyl group are not considered). The InChI community may, in future, consider an alternative convention to shorten the InChIKey in these cases.

In the case of caffeine, or a similarly well-known molecule, a search on the InChIKey string, as illustrated in Fig. 7, would yield many websites and plenty of information, and on some of these there will be the structure of the molecule. In addition, a search of this kind should return detail on any commercially available material from a supplier website. In general, however, the InChIKey, as a hash, cannot be inverted, so given the InChIKey, we cannot simply calculate the InChI and thus access the molecular structure. This means that a QR-encoded InChIKey requires a resolver to connect back to the actual InChI and to information on the molecule that the InChIKey represents. Therefore, a QR InChI app would be desirable, one which has the capability of connecting to one or more of several resolvers capable of interpreting the InChIKey. A research or industrial organization should be able to use an interface with its inventory control software as a resolver to satisfy this need, particularly if production of a QR code label for a sample/container was coupled to create a matching entry in the resulting database. Given modern smart phone storage abilities, it may be possible to hold a cache copy of a suitable resolver to enable the app to work even without an internet connection.

It may be desirable for the InChIKey QR code to be framed as a search, even if it is intended for use by an organization’s QR code app. This should ensure that some information on the molecule is obtainable, through extraction of the InChIKey, even by someone not in possession of that QR code app.

2.3 Organization-Specific QR InChI/InChIKey Use – Implementation in Commercial/ User Applications

Key organizational users will likely prefer, or need, to have the QR InChI/InChIKey resolver for their own databases and digital resources and which we recommend. This can be achieved by modifying the general approach described above in key ways that depend on the nature of the organization and its needs. However, the standardization of the format of the InChI/InChIKey portion of the QR code is important, as this should allow wider use of a QR InChI/InChIKey outside the organization that created it.

First, the QR code generating software used by an organization such as a chemical manufacturer/supplier will still use the InChIKey for the compound, but their own catalog/database website will be included instead of the general search string. This would provide a link to safety data and other material data related to the compound that is provided by the manufacturer. This QR code would clearly be useful to, or indeed required by, user organizations and the people in them.

A second but similar approach could be used in a university or other organizations that have need to track chemicals that are prepared in them (for which they will print and apply their own QR code labels and establish their own database to which the QR code would point). Printing of the QR code label would be linked to creating an entry in the organizational database, which will effectively become a local resolver for the InChIKey URL for the samples. The QR code reader will then enable a user to enter that specific catalog/database and access the information that it contains (e.g., the full InChI, a structural diagram, quantities, and any other stored information).

Depending on the organizational use-case requirements, additional data (e.g., batch numbers, inventory data/bottle codes, lab notebook page numbers, or related user codes) will have been appended to the string that is encoded and used by the database controlling software to select relevant entries. Such organizations may also choose to provide mechanisms by which quantities in bottles, or removed for reactions, can be entered as part of chemical tracking processes and inventory control.

These latter mechanisms provide examples of alternative use cases that may require more details to be entered by users when generating or using QR codes. These may be the data that the organization require for linking to other relevant information, or organization-specific processes. This may involve some kind of QR code reader app that may also be used to make or modify entries in the database (that may be linked to others in various ways).

In the case of commercially sourced samples, research organizations may wish to retain and make use of both the commercial QR code and the self-generated one, so that there is the ability to access the supplier-derived information (e.g., SDS sheets) and also engage with tracking processes.

Overall, the idea is to make use of the supplier websites together with InChIKey identifiers in their catalogs wherever possible. An InChI QR app could follow the link to the supplier's web pages and appropriate data for the compound/sample and extract the InChI as well. If the InChIKey is readily resolvable *via* sites such as PubChem, ChemSpider, and Wikipedia, then alternative sources of information will be readily available as well.

A commercial QR code would contain a URL of the form:

```
http://www.supplier.com/catalog/?InChIKey=BSYNRYMUTXBXSQ, SupplierID = 123456789.
```

An organization QR code would contain a URL of the form:

```
http://www.organization/chemical/inventory?InChIKey=BSYNRYMUTXBXSQ, user_id = XXXYYY, book#, page#, quantity.
```

It would be the responsibility of the organization to ensure that the URL is resolvable. We note, however, that it is not a trivial responsibility to ensure that these URLs will always be resolvable.

The following are examples of resolvers that currently exist, have a large coverage, and provide further examples of the kind of URL that we envisage. While the QR code could contain these links directly, they could realistically contain only one at most. Also, for some of these resolvers, the text "InChIKey=" does not occur and so it would not be as simple for an InChI app to recognize the InChI and provide additional search possibilities; this would not be ideal for recognition of the InChI/InChIKey. We reiterate that the most important aspect of the QR InChI is that the phrase InChIKey=xxxxxxxxxxxxxxxxxxxx is present in an easily identifiable form as this would allow for a QR code reader to direct the user to a variety of resources that can deal with the InChIKey.

Identifiers.org (a semantic web resource)

```
http://identifiers.org/inchikey
```

National Institutes of Health resources

```
http://cactus.nci.nih.gov/chemical/structure/RYYVLZVUVIJVGH-UHFFFAOYSA-N/names
```

And Royal Society of Chemistry, ChemSpider as mentioned above

```
http://www.chemspider.com/inchikey=RYYVLZVUVIJVGH-UHFFFAOYSA-N
```


2.4 QR Code Version and Appearance

The relatively compact version of the InChIKey identified in Fig. 5 means a QR code large enough to contain ca 60 characters will be sufficient, and this should not present much difficulty for the documentation but could be a problem on a small bottle. Additional information to carry supplier details etc. would need a larger QR code, but this would also in principle allow some of the text currently on a label to be removed and replaced by the information contained within the QR code. We note, however, that some of the text on a label is a legal requirement and space is at a premium. Governmental and industry agreement would be needed to alter any of the basic labeling requirements, but there may be some flexibility in how these requirements are met.

2.5 The Possible Supporting Infrastructure – The case/need for a General InChIKey Resolver

The principal use cases described above will be largely satisfied by a combination of databases/catalogs that are provided and maintained by chemical suppliers and those that may be set up by organizations as part of inventory control or data repository activities. However, there may be many smaller organizations that will have need for chemical tracking (perhaps legislatively required) but which lack the resources and infrastructure to maintain their own database. This need could be satisfied through provision of an InChIKey resolver that is linked to the production of a suitable QR code for an organization.

An InChIKey resolver is essentially a database that links an in-coming InChIKey with the full InChI (which can be returned to the interrogating device and used to generate the original structure) and any other data that the creator of the resolver may wish to provide (e.g., health and safety information, SDS files, inventory data, molecular properties, characterization information, and quality control data). Much of this is information that is in common with that provided in on-line catalogs that are maintained by chemical suppliers, and such catalogs could readily be adapted and used within or by an InChIKey resolver.

The challenge, therefore, is to provide a mechanism that allows the resolver database to be populated and grow as QR code labels are generated that will need to be linked to the InChI and related data.

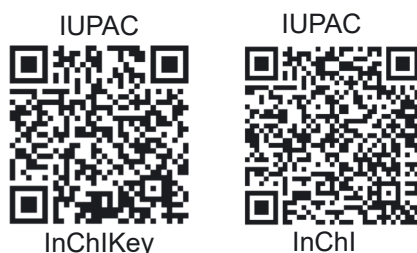
A general InChIKey resolver database would have an entry added to it by the QR code generating software – each time an InChI for a structure is supplied to the software, it will enter the InChI and InChIKey in the database (or perhaps multiple databases), create the general search string described above (using the InChIKey), and then produce and return the corresponding QR code that was requested by the user. The QR code reader would then trigger the search in the normal way.

It remains to be seen whether this can be practically provided, either by a not-for-profit organization, or as part of a commercial business that offers the service in return for a membership fee, or some other commercial consideration.

3 InChI QR Code Specification

3.1 Specification

1. The InChI QR code should have the IUPAC name above the QR Code and the InChI or InChIKey below the code, as in the following examples



2. The InChI QR code will contain the InChI/InChIKey for the major or most significant component in the container. This should be the species that would normally be on the label and be consistent with the usual legal requirements for labeling a container, i.e., the compound name that would be expected to be on the bottle should be used. Note: Additional information such as a solvent, which would clearly be ideal and can be very significant, would not be part of the main label but could be included as an additional code. This recommendation should be reviewed when the mixtures InChI (MInChI) become formally available.
3. For small molecules and those with a compact InChI (for example, with less than about 150 characters), the InChI QR code may contain the actual InChI in the form of InChI='InChI text'.
4. In the general case, the InChIKey should be used to ensure that the whole string can be captured with a moderate-sized QR code with good accuracy. The QR code should then contain a string of the form InChIKey = 'InChIKey text'.
5. The option to specify a resolver for the InChIKey to ensure the InChIKey can be resolved is allowed, but the search string should explicitly contain the text InChIkey = 'InChIKey text' so that a QR code reader (or a specific InChI App) can recognize the presence of an InChIKey and provide alternative resolution and information if required. The direct use of a link to a specific entry on a website of database, e.g., the use of PubChem or ChemSpider or similar identifiers would not be allowed with a QR InChI code as this would bypass the InChI/InChIKey.
6. Providing additional information alongside the InChIKey to give, for example, batch numbers, purity, etc. would be allowed, subject to the information carrying capacity for the chosen QR code (assuming the use of the standard URL expression with standard keywords).
7. The QR InChI code should, as far as possible, be constructed in a way that always does something useful when read by a general QR code reader that would offer the ability to look up a URL or present the text in the QR code to a general search engine. The standard forms for the InChI/InChIKey expression within the QR code that are given in this paper should achieve this aim.

4 Future work for IUPAC and the InChI Trust

Based on the discussion above, and the specification in Section 3, we make the following suggestions to IUPAC and the InChI Trust regarding the frameworks within which this QR code specification can be used:

- (1) That an InChI resolver should be set up and run by IUPAC and the InChI Trust. Those running the resolver should work with the major (open and public) chemical database providers to ensure that the resolver contains as wide a coverage of the compounds mentioned in the chemical literature and suppliers as possible; we acknowledge that this again is far from trivial to support in the long term.
- (2) As the QR code is a formally un-enforced patent, IUPAC and the InChI Trust should provide an easy to use and free web service to create the QR codes for the InChI and InChIKey. This would most simply be added as part of the existing service to create an InChI/InChIKey and the InChI code.
- (3) That IUPAC and the InChI Trust should commission a QR InChI app (ideally for IOS and Android) that would read the InChI QR code and then enable the users to make the most of the InChI-enabled services that exist worldwide. This would significantly increase the worldwide exposure for the InChI.
- (4) A project should explore the most compact way to store the InChI/InChIKey pair representations to enable, if possible, the off-line use of an InChI resolver in an app to cover at least the compounds most likely to be seen by, for example, first responders to an incident. An InChIKey to GHS code mapping would be a very useful immediate information source.
- (5) Discussions should be opened, by IUPAC and the InChI Trust, with GS1 – the organization that is behind most of the bar codes used by industry and ensures worldwide compatibility – as this will be an important avenue for adoption. This is not needed for the details of the QR codes presented in this recommendation but is important for further use of the InChI as part of the process of linking more general bar codes (e.g., goods labels) to relevant chemical resources.

- (6) That there should be discussions with industry associations to understand in more detail how the QR codes we recommend can be used within their current labeling systems and requirements that need to be met by the chemical suppliers.

5 Conclusions

We have shown how the use of a QR bar code label can provide a convenient link between the physical objects (a chemical container, bottle, or package) and information about the main/active chemical species contained in that package. We have outlined the use cases that would make use of a new internationally agreed standard, principally through use of the InChIKey, that would allow for dissemination of relevant information (e.g., Material Safety Data Sheets, or other relevant health and safety information), and/or mechanisms for chemical tracking and data storage.

There is also a clear potential opportunity for provision of an authoritative InChIKey resolution service to facilitate obtaining the molecular structure and chemical information about the species. We have provided a clear recommendation for a standard that could be used in advance of a suitable international agreement, one that would make use of suppliers' digital catalogs and would allow suppliers to provide an InChIKey along with their own supplier ID, batch number, etc., without needing extra space on the label and still meeting all regulatory requirements.

Acknowledgments: The authors acknowledge IUPAC and the InChI Trust for funding this project and thank all of the volunteers who have contributed to the development of the InChI standards and participated in discussions and workshops related to development of the InChI QR code standard. This paper results from an IUPAC project: Identifying International Chemical Identifier (InChI) Enhancements – QR codes and Industry Applications. <https://iupac.org/project/2015-019-2-800>.

Membership of the sponsoring body

Membership of the IUPAC Chemical Nomenclature and Structure Representation Division for the period 2020–2021 was as follows:

President: Prof. Alan Hutton (South Africa); **Vice President:** Dr. Michelle Rogers (USA); **Secretary:** Prof. Risto Laitinen (Finland); **Titulars Members:** Prof. Mike Beckett (UK), Prof. Edwin Constable (Switzerland), Dr. Karl-Heinz Hellwich (Germany), Dr. Elizabeth Mansfield (USA), Prof. Ebbe Nordlander (Sweden), Prof. Amelia Rauter (Portugal), Prof. Jiri Vohlidal (Czech Republic); **Associate Members:** Prof. Neil Burford (Canada), Dr. Thomas Engel (Germany), Prof. Robin Macaluso (USA), Dr. Erik Szabo (Slovakia), Prof. Augusto Tomé (Portugal), Dr. Clare Tovee AM (UK); **National Representatives:** Dr. Ture Damhus (Denmark), Prof. Safiye Erdem (Turkey), Dr. Adeyinka Olubunmi Fasakin (Nigeria), Prof. Rafal Kruszynski (Poland), Prof. Ladda Meesuk (Thailand), Prof. Jozsef Nagy (Hungary), Dr. Maria Atanassova Petrova (Bulgaria), Prof. Dusan Sladić (Serbia), Dr. Molly Strausbaugh (USA), Prof. Guoqiang Yang (China/Beijing).

References

- [1] S. R. Heller, A. McNaught, I. Pletnev, S. Stein, D. Tchekhovskoi. *J. Cheminf.* **7**, 23 (2015), <https://doi.org/10.1186/s13321-015-0068-4>.
- [2] <https://iupac.org/>.
- [3] <https://www.inchi-trust.org/>.
- [4] S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi, I. Pletnev. *J. Cheminf.* **5**, 7 (2013), <https://doi.org/10.1186/1758-2946-5-7>.
- [5] S. M. Bachrach. *J. Cheminf.* **4**, 34 (2015).

- [6] I. Pletnev, A. Erin, A. McNaught, K. Blinov, D. Tchekhovskoi, S. Heller. *J. Cheminf.* **4**, 39 (2012), <https://doi.org/10.1186/1758-2946-4-39>.
- [7] <http://www.hse.gov.uk/chemical-classification/labelling-packaging/safety-data-sheets.htm>.
- [8] <http://www.hse.gov.uk/chemical-classification/legal/background-directives-ghs.htm>.
- [9] <https://echa.europa.eu/regulations/reach/understanding-reach>.
- [10] <https://www.inchi-trust.org/about-the-inchi-standard/>.
- [11] W. A. Warr. *J. Comput. Aided Mol. Des.* **29**, 681 (2015), <https://doi.org/10.1007/s10822-015-9854-3>.
- [12] M. J. Bryant, S. N. Black, H. Blade, R. Docherty, A. G. P. Maloney, S. C. Taylor. *J. Pharmaceut. Sci.* **108**, 1655 (2019), <https://doi.org/10.1016/j.xphs.2018.12.011>.
- [13] A. J. Williams, A. Yerin. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 150 (2013), <https://doi.org/10.1002/wcms.1118>.
- [14] C. Southan. *J. Cheminf.* **5**, 10 (2013), <https://doi.org/10.1186/1758-2946-5-10>.
- [15] A. Hersey, J. Chambers, L. Bellis, B. A. Patricia, A. Gaulton, J. P. Overington. *Technologies* **14**, 17 (2015), <https://doi.org/10.1016/j.ddtec.2015.01.005>.
- [16] InChI requirements for representation of organometallic and coordination compound structures, Task group Chair: C. Batchelor: <https://iupac.org/project/2009-040-2-800>.
- [17] Implementation of InChI for chemically Modified large biomolecules, Task group Chair: K. T. Taylor: <https://iupac.org/project/2013-010-1-800>.
- [18] InChI requirements for representation of polymers, Task group Chair: A. Yerin: <https://iupac.org/project/2009-042-1-800>.
- [19] Redesign of handling of tautomerism for InChI V2, Task group Chair: M. C. Nicklaus: <https://iupac.org/project/2012-023-2-800>.
- [20] Enhanced recognition and encoding of stereoconfiguration by InChI tools, Task group Chair: A. Yerin: <https://iupac.org/project/2019-017-2-800>.
- [21] InChI extension for mixture composition, Task group Chair: L. R. McEwen: <https://iupac.org/project/2015-025-4-800>.
- [22] <https://github.com/IUPAC/MInChI>.
- [23] <http://www.mobile-qr-codes.org/history-of-qr-codes.html>.
- [24] <http://chemspider.com>.
- [25] <http://www.inchi-trust.org/technical-faq/>.
- [26] <https://www.qrcode.com/en/about/version.html>.