

Томачинський Сергій Миколайович,

Судовий експерт лабораторії права промислової власності Науково-дослідного центру судової експертизи з питань інтелектуальної власності Міністерства юстиції України, 01133, м. Київ, бульвар Лесі Українки, буд. 26, оф. 501, +38 050-683-65-61, e-mail: tomachyn@gmail.com, <https://orcid.org/0000-0002-7865-0615>

МІЖНАРОДНИЙ ТЕКСТОВИЙ ХІМІЧНИЙ ІДЕНТИФІКАТОР INChI ЯК ПЕРСПЕКТИВНИЙ ІНСТРУМЕНТ СУДОВОЇ ЕКСПЕРТИЗИ В ГАЛУЗІ ІНТЕЛЕКТУАЛЬНОЇ ВЛАСНОСТІ

Анотація. Розвиток інформаційних технологій в галузі хімічної інформатики пропонує різні продукти, які можуть зробити судову експертизу в галузі інтелектуальної власності, яка пов'язана з хімічними сполуками, продуктивнішою та більш переконливою для учасників судового провадження, які не мають спеціальних знань в галузі хімії, фармацевтики та суміжних з ними галузях. Одним з таких корисних інструментів є міжнародний текстовий хімічний ідентифікатор InChI та його скорочене похідне InChIKey.

На прикладі, взятому з судової практики, показано переваги застосування міжнародного текстового хімічного ідентифікатора InChI під час проведення судової експертизи або дослідження, наприклад, для порівняння та ідентифікації хімічних сполук. InChI та InChIKey також є корисними для пошуку документів, необхідних для пояснень в суді та більш наочного обґрунтування висновку судового експерта. В статті докладно розкрито алгоритм генерування InChI та InChIKey. Також, наведено посилання на безоплатні інформаційні ресурси, за допомогою яких можна завантажити програмне забезпечення або прямо на сайті застосовувати цей алгоритм для вирішення різних задач, які зустрічаються під час проведення судової експертизи або експертних досліджень в галузі інтелектуальної власності. Показано переваги InChI, зокрема, очевидність для учасників судового провадження, які не мають спеціальних знань в галузі хімії, результатів порівняння двох хімічних формул, що зображені в різних стилях хімічних формул. На відміну від досить складної процедури порівняння двовимірних хімічних формул таких тривимірних об'єктів, як молекули, після генерування відповідних InChI задача трансформується у порівняння двох рядків символів, що легко виконати за допомогою пошукової функції будь-якого текстового редактора, наприклад, Word або навіть простим порівнянням. Також коротко описано застосування InChIKey для ефективного пошуку науково-технічних публікацій.

Матеріал, викладений в статті, дасть змогу навіть читачам без хімічної освіти зробити перші кроки в опануванні такого сучасного інструменту, як InChI. Міжнародний текстовий хімічний ідентифікатор InChI є корисним інструментом з вели-

кими перспективами для спеціалістів різних галузей знань та, зокрема, для судових експертів в галузі інтелектуальної власності.

Ключові слова: судова експертиза, хімічні формули, хімічний ідентифікатор, InChI, InChIKey.

Tomachynskyi Serhii Mykolayovych,

Judicial expert of the Laboratory of Industrial Property Rights, Research Center for Forensic Examination on Intellectual Property of the Ministry of Justice of Ukraine, 01133, Kyiv, L. Ukrainka Boulevard, 26, of. 501, tel. +38 050-683-65-61, e-mail: tomachyn@gmail.com, <https://orcid.org/0000-0002-7865-0615>

INCHI, THE INTERNATIONAL TEXT CHEMICAL IDENTIFIER, AS A PROSPECTIVE TOOL FOR FORENSIC EXAMINATION ON INTELLECTUAL PROPERTY

Abstract. In the field of chemical informatics, the development of information technologies offers various products that can make forensic experts in the field of intellectual property related to chemical compounds productive and more convincing for court participants, who do not have special knowledge in the field of chemistry, pharmaceuticals, and related industries. International Text Chemical Identifier InChI and its abbreviated derivative InChIKey is one such useful tool.

The advantages of using the InChI international textual chemical identifier are shown here, using an example taken from judicial practice, when conducting a forensic examination or research, for example, for comparison and identification of chemical compounds. InChI and InChIKey are also useful for searching for documents, which are necessary for explanations in court and for more visual substantiation of the conclusion of a forensic expert. This article details the algorithm for generating InChI and InChIKey. There are also links to free information resources with which you can download software or use this method directly on the site to solve various problems that occur during forensic examinations or expert studies in the field of intellectual property. In a particular case, the evidence for trial participants who do not have special knowledge in chemistry, the results of comparing two chemical formulas that are depicted in different styles of chemical formulas, show the advantages of InChI. In contrast to the rather complicated procedure for the equation of two-dimensional chemical formulas of such three-dimensional objects as molecules, after generating the corresponding InChI, the task is transformed into equations of two lines of characters, it is easy to perform using the search function of any text editor, for example, Word, or even a simple comparison. It also briefly describes the use of InChIKey for the efficient search for scientific and technical publications. The material presented in the article will allow even readers without a chemical education to take the first steps in mastering such a modern tool as InChI. International Text Chemical Identifier InChI is a useful tool with great prospects for specialists in various fields of knowledge and, in particular, for forensic experts in the field of intellectual property.

Keywords: forensic examination, chemical formulas, chemical identifier, InChI, InChIKey.

Постановка проблеми. Судова експертиза, що стосується хімічних сполук, досить часто призначається в судових провадженнях. Висновок судового експерта, крім надання правильної відповіді на поставлені питання, має бути ясным, переконливим і зрозумілим не тільки для фахівців в галузях наук та техніки, близьких до хімії. Розвиток інформаційних технологій пропонує різні продукти, які можуть зробити судову експертизу в галузі інтелектуальної власності, яка пов'язана з хімічними сполуками, продуктивнішою та більш переконливою для учасників судового провадження, які не мають спеціальних знань. Одним з таких корисних інструментів є міжнародний текстовий хімічний ідентифікатор InChI та його скорочене похідне InChIKey.

Аналіз дослідження і публікацій. Міжнародний текстовий хімічний ідентифікатор InChI є порівняно новим продуктом – він введений в практику в 2005 році і більша частина публікацій по ньому стосується застосувань в спеціалізованих хімічних, біологічних та фармацевтичних базах даних, хімічного пошуку та хімічної інформатики. Про застосування InChI в правовій сфері згадано в роботі [1], але без наведення конкретних прикладів чи посилань.

Мета статті – на прикладі, взятому з судової практики, показати переваги застосування міжнародного текстового хімічного ідентифікатора InChI під час проведення судової експертизи або дослідження, зокрема, для порівняння та ідентифікації хімічних сполук, для пошуку документів, необхідних для пояснень в суді, або для більш наочного обґрунтування висновку експерта. Метою цієї статті було не наведення великої

кількості прикладів, а надання докладно розписаного алгоритму генерування InChI та InChIKey та їх застосування. Це, разом з наведеними посиланнями, дасть змогу навіть читачам без хімічної освіти зробити перші кроки в опануванні такого сучасного інструменту, як InChI, який буде корисний спеціалістам різних галузей знань та, зокрема, судовим експертам в галузі інтелектуальної власності.

Виклад основного матеріалу. Досить часто, в процесі досліджень, що стосуються лікарських засобів, гербіцидів, пестицидів та інших хімічних речовин, перед судовим експертом в галузі інтелектуальної власності виникає потреба в порівнянні формул хімічних речовин. Одна з цих формул, як правило, міститься в формулі винаходу, а інша – в інструкції, результатах хімічного аналізу або інших документах до продукту, що перебуває в господарському обороті. При цьому, хімічні формули одних і тих самих речовин можуть бути зображені по-різному. Як правило, у фахівців не виникає труднощів у порівнянні таких формул. При цьому, викласти переконливо для нефахівців докази ідентичності сполук, зображених формулами, побудованими за різними правилами, вдається не завжди. З одного боку, включення до висновку експертизи викладення суті хімічних явищ, які є причиною різних позначень одних і тих же сполук, призводить до збільшення обсягу самого висновку і не завжди такого стислого викладення достатньо, щоб для учасників судового провадження, які, не мають спеціальних знань в галузі хімії, проблема стала ясною. З іншого боку, занадто лаконічне і зрозуміле тільки фахівцям викладення матеріалів у висновку судового експерта, може призвести до того, що, згідно статті 113 ЦПК [2], у випад-

ку, якщо суд визнає висновок експерта неповним або неясним, судом може бути призначена додаткова експертиза, а у випадку, якщо суд визнає висновок необґрунтованим або таким, що суперечить іншим матеріалам справи або викликає сумніви в його правильності, судом буде призначено повторну експертизу. Це призводить до небажаного затягування судового провадження і додаткових витрат для учасників процесу. Щоб уникнути такої ситуації, необхідно обрати критерій порівняння двох хімічних формул, однозначність застосування якого буде зрозуміла для судді та інших учасників судового провадження, які, як правило, не мають спеціальних знань в галузі хімії.

Свого часу було створено різні способи формалізації хімічних формул, наприклад, представлення їх у вигляді матриць суміжності, файлів в форматі MOL та SDF тощо. Однак, для можливості обробки їх комп'ютерами, набагато зручнішими є формати, де ідентифікатори хімічних сполук мають лінійний формат. Це, наприклад, формати Wiswesser, WLN, Sybyl, SLN, CAS Registry Number, SMILES та InChI. Для фахівців може бути цікавим огляд цих ідентифікаторів [3]. Найбільш зручним та універсальним зі згаданих ідентифікаторів є InChI (International Chemical Identifier). Цей ідентифікатор є порівняно новим продуктом – він був введений в практику в 2005 році, і його творці наділили його рядом властивостей, які роблять його особливо привабливим для фахівців різних галузей науки і техніки, які мають справу з хімічними речовинами. Серед багатьох переваг InChI слід відзначити такі:

1. Унікальність. Один і той же ідентифікатор завжди означає одну

і ту саму речовину. Це досягається за допомогою чітко визначеної процедури канонічної нумерації атомів, яка проводиться автоматично, без можливості користувача впливати на неї. При цьому, одна сполука може мати кілька InChI, які співпадають в основній частині, але відрізняються в додаткових. Такі InChI називають синонімічними. Докладно це явище та його причини ми розглянемо в наступній статті.

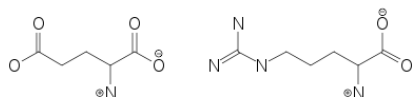
2. Відкритість. Система присвоєння InChI не є чиеюсь власністю. Програмне забезпечення до версії 1.04, є вільному доступі під ліцензією Open source. Наступні версії розповсюджуються під користувальницькою ліцензією IUPAC-InChI Trust License неприбуткової організації InChI Trust, яка є членом IUPAC.

3. Доступність. Будь-хто, за потреби, може генерувати InChI на основі хімічної формули, або збережених MOL або SDF файлів, або ідентифікаторів інших форматів. Якщо, наприклад, нова сполука вперше описана в патенті, або науковій статті, то потрібен досить тривалий час, поки цей документ буде реферовано службою Chemical Abstracts Service і сполукам, що описані в ній, буде присвоєно такий відомий ідентифікатор, як CAS Registry Number. При цьому, доступ до баз даних служби Chemical Abstracts Service є платним. В той же час, InChI можна безоплатно сгенерувати вже під час написання статті або патенту і включити в них цей ідентифікатор.

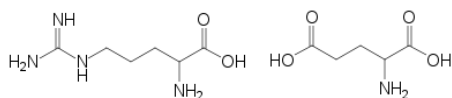
4. Зрозумілість. Людина навіть з невеликим досвідом роботи з InChI швидко звикає зчитувати значну кількість інформації безпосередньо з нього без допомоги комп'ютера.

5. Об'єктивність. Неможливо для певної хімічної сполуки сгенерувати інший стандартний InChI, ніж це робить призначене для цього програмне забезпечення.

Розглянемо як ці переваги допомагають у вирішенні практичних задач. Наприклад, необхідно дати відповідь на питання «Чи характеризують формули I та II різні сполуки чи одну і ту саму сполуку?» (Аналогічне питання ставилося перед судовими експертами при розгляді Справи № 910/2318/16 господарським судом міста Києва.)



I



II

Для того, щоб зрозуміло пояснити, чому був зроблений саме такий висновок експертизи або дослідження, судовий експерт може у відкритих джерелах (Інтернет, наукові або технічні видання) знайти документ, що містить обидві такі формули. Пошук по хімічних назвах в мережі Інтернет видає величезну кількість наукових документів та патентів. Перегляд всіх цих документів потребує багато часу і більша частина з них це «інформаційне сміття» – документи з низькою цінністю для вирішення поставленої задачі. Крім того, наукові публікації та патенти містять хімічні формули виконані в одному стилі. Формули різних стилів можуть міститися

в підручниках, але пошук по таким публікаціям вкрай трудомісткий, оскільки дуже мала частина підручників доступна для повнотекстового пошуку. До того ж, далеко не факт, що в прикладах будуть саме такі сполуки, щодо яких поставлені питання. Таким чином, такий варіант пошуку пояснень правильності висновку експерта є дуже трудомістким, не завжди результативним і не дуже переконливим для нефакхівців.

Тепер докладно розглянемо алгоритм вирішення цієї ж задачі з використанням InChI, який може повторити навіть людина без спеціалізованої хімічної освіти.

1. Знаходимо один з сайтів з генератором InChI, наприклад, сайт <https://cactus.nci.nih.gov/chemical/structure> (якщо є потреба в частому використанні InChI, можна завантажити генератор InChI на комп'ютер, наприклад, з сайту <https://www.inchi-trust.org/downloads/>)

2. На згаданому вище сайті, натискаємо клавішу «Structure» і в дописуємо простому графічному редакторі малюємо формулу I, та, по закінченню, натискаємо клавішу «Done». У вікні «Structure Identifier:» з'явиться наступний запис: NC(CCCNC(N)=N)C(O)=O. NC(CCC(O)=O)C(O)=O

Це представлення намальованої формули в лінійному форматі SMILES (Simplified Molecular Input Line Entry Specification). У вікні «convert to:» обираємо опцію «Standard InChI» і натискаємо клавішу «Submit». Система видає InChI та гіперпосилання на нього (при великій довжині InChI, його легше копіювати, відкривши гіперпосилання):

InChI=1S/C6H14N4O2.C5H9NO4/c7-4(5(11)12)2-1-3-10-6(8)9;6-3(5(9)10)1-2-4(7)8/h4H,1-3,7H2,(H,11,12)(H4,8,9,10);3H,1-2,6H2,(H,7,8)(H,9,10)

Аналогічні операції виконуємо і для формули II, одержуємо її SMILES:
NC(=N)NCCCC([NH3+])C([O-])=O.[NH3+]
C(CCC(O)=O)C([O-])=O

та генеруємо InChI для формули II: InChI=1S/C6H14N4O2.C5H9NO4/c7-4(5(11)12)2-1-3-10-6(8)9;6-3(5(9)10)1-2-4(7)8/h4H,1-3,7H2,(H,11,12)(H4,8,9,10);3H,1-2,6H2,(H,7,8)(H,9,10) InChI в обох випадках є ідентичними, що однозначно та беззаперечно доводить факт ідентичності хімічних сполук, що зображені формулами I та II. Важливою для судового експерта є та обставина, що це не пошук нових відомостей або документів в додаткових джерелах, а використання визнаного спеціалістами програмного забезпечення для трансформування інформації, що міститься в матеріалах справи, в інший, стандартизований формат за стандартною та відтворюваною процедурою з відкритим алгоритмом. При цьому, факт ідентичності обох InChI легко перевірити за допомогою, пошукової функції будь-якого текстового редактора, наприклад, Word, і навіть простим порівнянням, що є доступним і наочним навіть для учасників судового провадження, які не мають спеціальних знань в галузі хімії. Слід звернути увагу, що у форматі SMILES відображення формул I та II відрізняється, тобто, на відміну від InChI, властивість унікальності ідентифікатору цього формату не притаманна. Для додаткового незалежного підтвердження того факту, що знайдений нами InChI стосується обох вказаних формул, можна спробувати знайти відповідні наукові документи. InChI є не дуже зручним засобом для цілей пошуку в Інтернет. Велика довжина InChI для багатьох сполук та велика кількість символів різного формату

може призвести до помилок або до некоректної роботи пошукових серверів. Тому, для зручності пошуку, після генерування InChI для кожної формули, слід, не очищаючи вікно «Structure Identifier:», обрати у вікні «convert to:» опцію «Standard InChIKey». Для формули I система видасть таку відповідь: InChIKey=RVEWUBJVAHOGKA-UHFFFAOYSA-N та гіперпосилання на неї.

Генерування InChIKey для формули II дає ідентичний результат. Як ми бачимо, InChIKey складається тільки з літер англійської абетки та набагато коротше за InChI (InChIKey має фіксовану довжину з 27 символів для всіх молекул). Це скорочена, так звана «хешована» форма представлення InChI. Вона легко сприймається багатьма пошуковими інструментами Інтернету, як універсальними, як, наприклад, Google, так і спеціалізованими, наприклад, PubChem, Chem Spider або NIST. Слід зазначити, що вказані спеціалізовані бази даних є безоплатними, але при цьому вони мають можливість пошуку в них як за InChIKey, так і за InChI, а в БД PubChem та Chem Spider доступний пошук навіть за SMILES. За допомогою InChIKey навіть в пошуковій системі Google без надмірних зусиль можна знайти документи, що містять зображення формул I та II, і які відповідають одному і тому ж InChI. Це, наприклад, документ з бази даних SpectraBase, в якому міститься формула I, [4] і документ з тієї ж бази даних, в якому міститься формула II [5]. Як легко бачити, InChI в обох документах ідентичні, що беззаперечно свідчить, що згадані формули та інші фізико-хімічні дані, що містяться в обох документах, стосуються однієї і тої ж хімічної речовини. Окремо слід зазначити, що, внаслідок скорочення, InChIKey

набагато складніший для розуміння безпосередньо людиною і має дещо меншу унікальність, ніж InChI. Тому, після знайдення результатів з однаковими InChIKey, варто окремо перевіряти ідентичність InChI в них. Вкрай рідко, але теоретично можливі однакові InChIKey для різних сполук. При великій кількості переваг (зручність запису, селективність пошуку, незалежність від мови написання документу) у пошуку по InChIKey є особливість, яка іноді може бути розцінена як недолік – більшість документів, які знаходяться запити з його використанням, опубліковані після введення InChIKey в практику, тобто після 2005 року. Ретроспективне доповнення записів ідентифікаторами InChI та ключами InChIKey поки що проведена в спеціалізованих базах даних. Слід відмітити, що навіть в такому ресурсі, як Вікіпедія, особливо в її англійському сегменті, в більшості статей, які стосуються хімічних сполук, вже наведено як InChI так і InChIKey. Однак, з огляду на відкритість цього ресурсу для редагування користувачами, краще перевіряти правильність взятих з неї відомостей в спеціалізованих базах даних.

Також слід зазначити, що, за необхідності повторити всі дії при генеруванні InChI, наприклад, під час додаткової експертизи або, щоб продемонструвати цей процес при наданні пояснень в суді, немає потреби заново малювати формули сполук. Достатньо на сайті <https://cactus.nci.nih.gov/chemical/structure> у вікно «Structure Identifier:» скопіювати збережений раніше SMILES відповідної формули (наприклад, той, що відповідає формулі I з розглянутого прикладу, для прикладу, його можна просто скопіювати з електронної версії цієї статті) та натиснути клавішу

«Structure». Відкриється вікно графічного редактора з вже намальованими формулами. Залишається тільки, коли формули виходять за межі вікна редактора, на нижній панелі редактора зменшити масштаб (натиснувши кілька разів клавішу із зображенням мінусу в збільшувальному склі). Ми також можемо завантажити відповідно перші частини SMILES (до крапки) з наших прикладів – NC(CCCNC(N)=N)C(O)=O та NC(=N)NCCCC([NH3+])C([O-])=O, згенерувати InChI до кожної з них і порівняти. Розмістивши обидва результати один над одним, легко бачити, що вони повністю співпадають. В спеціалізованих базах даних легко встановити, що цей InChI відповідає амінокислоті аргінін. Факт повного співпадіння є очевидним і для людей без спеціальних знань в галузі хімії:

InChI=1S/C6H14N4O2/c7-4(5(11)12)2-1-3-10-6(8)9/h4H,1-3,7H2,(H,11,12)
(H4,8,9,10)

InChI=1S/C6H14N4O2/c7-4(5(11)12)2-1-3-10-6(8)9/h4H,1-3,7H2,(H,11,12)
(H4,8,9,10)

Аналогічно, можна відновити і перевірити формули других частин, які відповідають глутаміновій кислоті.

Функція відновлення формул зі SMILES дозволяє швидко одержати зображення складних формул, малювання яких забирає багато часу, а також зберігати і потім відновлювати проміжні результати, поки малювання формули ще не завершено. При цьому, стилістика відображення формул (розподілення зарядів або таутомерна форма тощо) зберігаються. За допомогою такої ж операції, але з використанням InChI також можна відновити формули, але треба вводити ідентифікатор разом з символами «InChI=».

Також, можна відновити формули і за допомогою InChIKey. В цьому випадку, система сприймає значення ключа як з символами «InChIKey=», так і без них.

Висновки і перспективи подальших досліджень. Фахівцям з судової експертизи в галузі інтелектуальної власності, які мають справу з хімічними сполуками, слід звернути увагу на новий потужний інструмент пошуку і порівняння хімічних сполук, який є об'єктивним, доступним, зрозумілим і унікальним для кожної сполуки. Крім того, за необхідності надання пояснень в суді, вся процедура генерування InChI та їх порівняння може бути в стислий час повторена безпосередньо в суді або під час надання пояснень за допомогою сеансу віддаленого зв'язку, коли судовий експерт не може прибути на засідання суду внаслідок карантинних заходів, стану здоров'я чи з інших причин. В той час, як в деяких закордонних публікаціях вже є згадування про використання InChI для юридичних та регуляторних цілей [1], на даному етапі розвитку методів судової експертизи, застосування InChI є поки що додатковим інструментом до класичних підходів, але у нього є потенціал стати стандартизованою процедурою при вирішенні питань про ідентичність хімічних сполук.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Heller S. R. InChI, the IUPAC International Chemical Identifier / S.R. Heller, A. McNaught, I. Pletnev, S. Stein, D. Tchekhovskoi // *Journal of Cheminformatics*. – 2015.- v. 7:23. p. 1-34 – Режим доступу: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4486400/pdf/13321_2015_Article_68.pdf
2. Цивільний процесуальний кодекс України від 18 березня 2004 р. // Відомо-

сті Верховної Ради України (ВВР). – 2004, № 40-41, 42, Ст. 492. – Режим доступу: <https://zakon.rada.gov.ua/laws/show/1618-15>

3. Warr W.A. / Representation of chemical structures // *Wiley Interdisciplinary Reviews: Computational Molecular Science*. – 2011. – v. 1: p. 557–579.

4. SpectraBase. L-arginine, compound with L-glutamic acid. [Електронний ресурс]. – Режим доступу: <https://spectrabase.com/spectrum/Av1Qs8Ess0w>

5. SpectraBase. L-arginine, compound with L-glutamic acid. [Електронний ресурс]. – Режим доступу: <https://spectrabase.com/spectrum/8Ni3B5MGOhv>

REFERENCES:

1. Heller, S. R., McNaught, A., Pletnev, I., Stein, S., Tchekhovskoi, D. (2015). InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics*, 7, 23. doi: 10.1186/s13321-015-0068-4, PMID: PMC4486400. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4486400/> [in English].
2. Tsyvilnyi protsesualnyi kodeks Ukrainy [Code of Civil Procedure of Ukraine]. (2004). *Vidomosti Verkhovnoi Rady Ukrainy – Bulletin of the Verkhovna Rada of Ukraine – 40-41, 42*. Art. 492. Retrieved from <https://zakon.rada.gov.ua/laws/show/1618-15> [in Ukrainian].
3. Warr, W.A. (2011). Representation of chemical structures. *Wiley Interdiscip Rev Comput Mol Sci*, 1, 557-579 [in English].
4. L-arginine, compound with L-glutamic acid. ¹³C Nuclear Magnetic Resonance (NMR) Spectrum. (n.d.). *spectrabase.com*. Retrieved from <https://spectrabase.com/spectrum/Av1Qs8Ess0w> [in English].
5. L-arginine, compound with L-glutamic acid. ¹H Nuclear Magnetic Resonance (NMR) Chemical Shifts. (n.d.). *spectrabase.com*. Retrieved from <https://spectrabase.com/spectrum/8Ni3B5MGOhv> [in English].