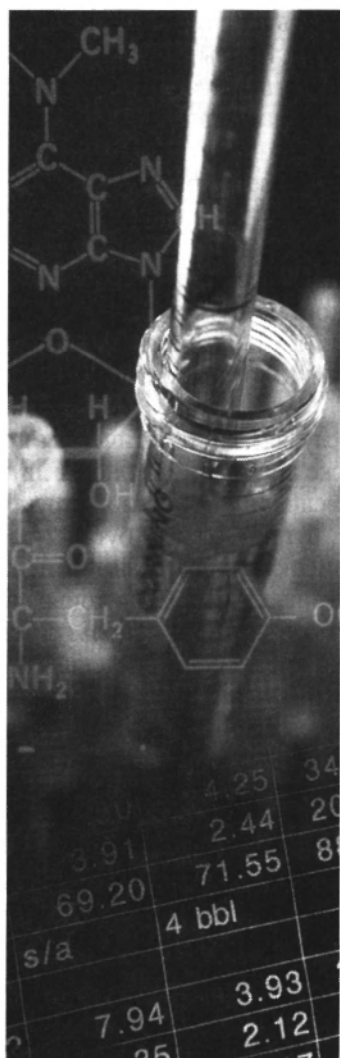


Д. М. Жилин
СШ № 192, Москва

ХИМИЧЕСКАЯ НОМЕНКЛАТУРА

в цифровую эпоху



Китайская мудрость гласит: «Неназванное не существует». В соответствии с этой мудростью химики до недавних пор стремились дать название каждому веществу, с которым они работали. Как известно, сначала они давали тривиальные названия (например, яблочная кислота). Однако к концу XIX в. число синтезированных и выделенных из окружающей среды веществ превысило число слов в языке. Кроме того, запомнить такое многообразие названий было нереально. К тому моменту общим требованием стало установление структуры всех вновь полученных соединений. Поэтому после Женевской конференции 1892 г. началась работа над созданием систематической номенклатуры и окончательно сформировалось требование однозначного соответствия структуры и названия вещества [1]. Как же решается эта проблема в настоящее время?

СИСТЕМАТИЧЕСКАЯ НОМЕНКЛАТУРА: ПАЦИЕНТ СКОРЕЕ МЁРТВ

Изначально в этом была призвана помочь так называемая систематическая номенклатура, правила которой до сих пор излагаются в школьных учебниках. Для органических соединений с 1993 г. в мире действует номенклатура ИЮПАК. Её правила представляют собой девять (!) многостраничных томов [2, 3], к которым в 1999 г. были опубликованы исправления [4]. Очевидно, что запомнить такой объёмный свод правил ничуть не легче, чем все тривиальные названия, которые существовали к началу XX в. А значит, все эти девять томов оказываются абсолютно бесполезными в сколько-нибудь сложных ситуациях. Один случай из научной практики автора это блестяще подтвердил. В какой-то момент встала задача выяснить структуру красителя под тривиальным названием Stains-all. Его номенклатурное название звучало так: 1-Ethyl-2-[3-(1-

ethylnaphtho[1,2-d]thiazolin-2-ethylidene)-2-methylpropenyl]naphtho[1,2-d]thiazolium bromide. Автор обошёл несколько самых сильных специалистов в химической номенклатуре на химическом факультете МГУ, и ни один из них не смог даже приближённо нарисовать формулу этого соединения (рис. 1).

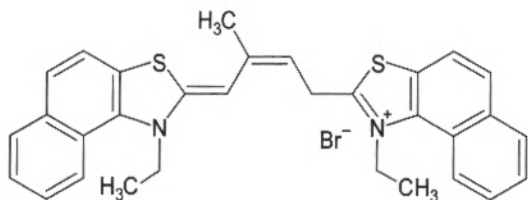


Рис. 1. Краситель Stains-all, номенклатурное название 1-этил-2-[3-(1-этилнафто[1,2-d]тиазолин-2-этилиден)-2-метилпропенил]нафто[1,2-d]тиазолия бромид

И действительно, кто из вас, дорогие читатели, с ходу назовёт по номенклатуре ИЮПАК такое распространённое соединение, как метиловый оранжевый (рис. 2), или узнает в названии «2-[(4-гидроксифенил)(4-оксоциклогекса-2,5-диен-1-илиден)метил]бензойная кислота» обычный фенолфталеин (рис. 3)?

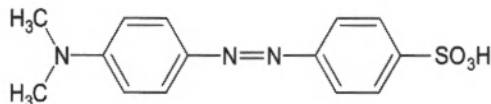


Рис. 2. Метиловый оранжевый, номенклатурное название 4-[(4-(диметиламино)фенил)диазенил]бензолсульфовая кислота

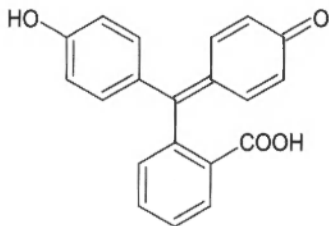


Рис. 3. Фенолфталеин. Номенклатурное название 2-[(4-гидроксифенил)(4-оксоциклогекса-2,5-диен-1-илиден)метил]бензойная кислота

Но сложность не единственная проблема номенклатуры ИЮПАК. Дело в том, что она сформирована на основе английского языка. Чтобы привести её в соответствие с нормами

других языков (в частности, русского), должны потрудиться национальные комиссии по химической номенклатуре. А теперь внимание: в России таковой комиссии нет. Более того, нет никаких нормативных актов, так или иначе утверждающих правила русскоязычной химической номенклатуры. И ещё более того: нет даже официального перевода действующих рекомендаций. Поэтому *никакой законодательной базы под русскоязычными номенклатурными названиями нет в принципе!* И все задания ЕГЭ, использующие номенклатурные названия, строго говоря, представляют собой правовую самостоятельность.

Кроме того, правила номенклатуры органических соединений, принятые в 1993 г., заметно изменились по сравнению с принятыми ранее, в 1979 г. Привыкшие к номенклатуре 1979 г. перестраиваются с трудом, а у тех, кто читал учебники, вышедшие в разные годы, вообще возникла в голове путаница. И если органическая номенклатура 1993 г. как-то приживается (так как принципиально не особо отличается от номенклатуры 1979 г.), то неорганическая номенклатура 2005 г. [5] оказалась полностью отторгнута научным сообществом. Никто так и не стал называть сульфит аммония триоксидосульфатом (2-) азания.

В каких же случаях систематическая номенклатура всё же полезна? Наверное, в двух. Во-первых, для веществ с несложной структурой (например, бутан-1-ол или бутан-2-ол), поскольку их номенклатурные названия ненамного сложнее тривиальных (если тривиальные вообще есть) и активно используются. Во-вторых, нужно иметь общие представления о систематической номенклатуре, чтобы из нескольких изомеров выбрать нужный (такая задача иногда возникает при поиске по молекулярным формулам). И это, пожалуй, всё.

ОРГАНИЗАЦИЯ ХИМИЧЕСКОЙ ИНФОРМАЦИИ В ЦИФРОВУЮ ЭПОХУ

Как же идентифицировать более сложные вещества? Способов несколько, и все они так или иначе, связаны с развитием информаци-

онных технологий. Но сначала следует понять, для чего вообще нужно знать названия и формулы веществ.

Рискнём высказать крамольную мысль: во многих случаях знать химическую формулу вещества не нужно вообще. Любого химика (если только он не собирается синтезировать данное вещество или использовать его для синтеза другого вещества), а тем более не химика, в первую очередь интересуют *свойства* вещества. Иногда их можно предсказать, зная его структурную формулу, а иногда нельзя. Действительно, многие ли, увидев формулу метилового оранжевого, способны сказать, что это индикатор, который меняет красный цвет на жёлтый в интервале pH 3,1–4,4 [6]? Такие свойства устанавливаются только экспериментально. Вопрос в том, как найти результаты соответствующих экспериментов. И здесь необходимо сделать отступление о том, как вообще структурирована химическая информация, накопленная человечеством.

Как известно, учёные самыми разными способами исследуют вещества. Результаты своих исследований они публикуют в научных журналах в виде оригинальных статей. Раньше все журналы были бумажными, сейчас почти все они имеют электронную версию и так или иначе доступны в Интернете. Более того, даже старые выпуски многих журналов отсканированы и переведены в электронный вид. Например, Американское химическое общество выложило электронные версии всех своих журналов с 1879 г. Это принципиально облегчает компьютерный поиск нужных статей и их получение из Интернета.

Оригинальных статей по химии публикуется огромное количество (в 2007 г. более миллиона [7]). Чтобы облегчить их поиск, в 1907 г. в США была организована служба химического реферирования (Chemical Abstracts Service, сокращённо CAS, www.cas.org). Её сотрудники писали краткие рефераты всех выходивших статей по химии и издавали их в реферативных сборниках (Chemical Abstracts). К каждому реферату прилагался список ключевых слов, а к сборнику, соответственно, указатель статей по ключевым

словам. Со временем все уважающие себя журналы стали требовать от авторов самостоятельно писать рефераты и указывать ключевые слова. Рефераты отсылали в CAS, которая включала их в сборники. С 1996 г. рефераты стали выходить на электронных носителях, что принципиально облегчило поиск в них. В России издаётся аналог Chemical Abstracts — реферативный журнал «Химия», однако его охват уже и он менее оперативен.

Со временем появился целый ряд альтернативных баз данных оригинальных статей, доступ к которым оказался дешевле и проще, а поиск и получение оригинальных статей — удобнее. Наиболее мощная из них в настоящий момент — ISI Web of Knowledge (www.isiknowledge.com). Кроме того, свои базы данных есть у всех крупных научных издательств, которых не так уж и много. На русском языке есть частично открытая научная электронная библиотека www.elibrary.ru.

Искать сведения о конкретных веществах в оригинальных статьях можно, но утомительно. Раньше их сводили в справочники и энциклопедии, сейчас — в базы данных о веществах. Наиболее мощная открытая база данных такого рода — PubChem (<http://pubchem.ncbi.nlm.nih.gov>). Она содержит важнейшие сведения о 30 млн. веществ, а также ссылки на оригинальные статьи, посвящённые этим веществам (в том числе и доступные в Интернете). В своё время CAS, почуввав конкуренцию, предприняла на PubChem ряд юридических атак (часто на грани фола), однако все эти атаки были отбиты.

Помимо PubChem, существует Википедия, она же свободная сетевая энциклопедия. Её русскоязычная версия находится по адресу ru.wikipedia.org, англоязычная — en.wikipedia.org. Это огромная свободная энциклопедия с простым механизмом поиска в поисковой строке, гиперссылками и указанием смежных тем. Она содержит информацию не только о веществах, доступна всем желающим для чтения и всем зарегистрированным пользователям — для правки. Из-за того, что править статьи в ней может кто угодно, они не всегда оказываются корректными. Однако с недав-

них пор англоязычная Википедия стала сотрудничать с CAS, что придало надёжности содержащимся в ней сведениям о веществах. Англоязычная Википедия гораздо полнее русскоязычной как по числу статей, так и по их объёму.

Очень надёжны базы данных производителей реактивов, в первую очередь Sigma Aldrich (www.sigmaaldrich.com/catalog). В них описаны важнейшие свойства веществ, которыми торгует фирма, и часто даются ссылки на оригинальные статьи, где эти вещества описаны.

Кроме того, развивается открытая база данных ChemSpider (www.chemspider.com), сеть токсикологических данных <http://toxnet.nlm.nih.gov>, база данных по токсическим веществам американского Агентства по охране окружающей среды actor.epa.gov/toxrefdb и некоторые другие. Даже CAS в 2009 г. снизошла до того, что открыла широкой общественности часть своей базы, содержащую сведения о 7900 соединениях, наиболее интересных общественности (<http://www.commonchemistry.org>).

К сожалению, все эти базы данных (кроме Википедии) исключительно англоязычные. Помимо этих баз, можно искать сведения о веществах при помощи любой стандартной поисковой системы, например www.google.com.

Таким образом, задача называния вещества заменяется задачей поиска сведений о нём в сетевых базах данных. Использовать систематические названия для этого далеко не всегда удобно по причинам, изложенным выше. Зато можно назвать сразу три пути, позволяющие решать эту задачу без привлечения систематической номенклатуры. Это использование тривиальных названий, цифровая идентификация и непосредственный поиск по структурным формулам.

РЕНЕССАНС ТРИВИАЛЬНЫХ НАЗВАНИЙ

Тривиальное название вещества, как правило, гораздо проще, чем номенклатурное, а потому легче запоминается. С появлением сетевых баз данных тривиальные названия

переживают второе рождение, так как вероятность ошибки при их вводе меньше, чем у систематических названий, а компьютерным базам (в отличие от человека) всё равно, сколько тривиальных названий хранить.

Казалось бы, по состоянию на 3 января 2011 г. описано более 56 млн. соединений, и каждые 2–3 секунды добавляется новое [8], поэтому проблема подбора тривиальных названий остаётся. Однако подавляющее большинство соединений описаны в одной-двух статьях, а с не менее подавляющим большинством остальных веществ работают только небольшие группы учёных. Тривиальные названия нужны веществам, с которыми работают многие химики, а таких веществ всего несколько тысяч. Запомнить столько названий уже можно. Но даже этого не нужно: если тривиальное название есть в базе данных, то можно найти не только формулу этого вещества, но и описание его свойств, что и требуется. Так, формулу и свойства красителя Stains-all автор в итоге нашёл в каталоге Sigma Aldrich по тривиальному названию.

Живучесть тривиальных названий особенно ярко проявляется в учебниках по биохимии. Ни в одной схеме вы не найдёте названий «2-гидроксипропан-1,2,3-трикарбонная кислота» и «2-гидроксипропан-1,2,3-трикарбонная кислота» — только «лимонная» и «яблочная». Более того, появляются новые тривиальные названия. Например, в 2006 г. российские учёные синтезировали соединение $(C_2S)_8$, формула которого напоминала цветок подсолнечника (рис. 4) [9]. Недолго думая, авторы назвали его «сульflower» (гибрид англ. *sulfur* — сера и *sunflower* — подсолнечник). Именно под таким названием это соединение теперь упоминается в статьях.

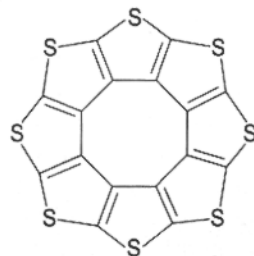


Рис. 4. Сульфловвер

ИДЕНТИФИКАЦИОННЫЕ НОМЕРА

Специалисты по компьютерным базам данных прекрасно знают, что самый простой способ идентифицировать какой-либо объект — это присвоить ему уникальный идентификационный номер. Поиск объекта по идентификационному номеру — наиболее простая задача поиска. И действительно, коль скоро существует 50 млн. соединений, достаточно их пронумеровать, и каждый номер потребует всего 26 бит информации. Даже если соединений будет миллиард и потребуются ввести в номер дополнительные цифры для повышения надёжности, то потребуются не более 40 бит. Для сравнения: номенклатурное название Stains-all занимает более 800 бит. Таким образом, идентификационный номер занимает гораздо меньше места. Хотя он не содержит никакой информации о структуре, её можно найти в базе данных по идентификационному номеру.

Эти соображения в 1957 г. сподвигли CAS на то, чтобы дать регистрационные номера всем веществам, когда-либо упоминавшимся в рефератах. Эти номера (обозначаются CAS RN) даются абсолютно бессистемно, но к каждому веществу номер «приклеен». Так, уже знакомый нам Stains-all имеет CAS RN 7423-31-6, метиловый оранжевый — 547-58-0, а фенолфталеин — 77-09-8. PubChem тоже даёт имеющимся в её базе веществам идентификационные номера. Они обозначаются PCID, или PCCID, или PubChemID. Stains-all имеет PCID = 16219338, а фенолфталеин — 4764.

Регистрационный номер CAS в обязательном порядке указывают практически во всех статьях, посвящённых соответствующему веществу. Во многих статьях также указывается номер PubChem. Также эти номера поддерживаются всеми серьёзными базами данных. Исключения составляют базы CAS и Американского химического общества, которые до сих пор не могут смириться с существованием PubChem, поэтому игнорируют PCID. С другой стороны, бесплатно узнать CAS RN для произвольного вещества невозможно, а PCID — можно на сайте PubChem.

ПОИСК ИНФОРМАЦИИ ПО ХИМИЧЕСКОЙ ФОРМУЛЕ

Когда CAS вводила свои регистрационные номера, компьютеры были большими и дорогими. Их мощностей хватало только на поиск информации по идентификационному номеру, но не по графической информации о нём. Между тем человеку гораздо удобнее идентифицировать объект по графической информации. И если идентифицировать людей по лицам компьютер до сих пор не может, то задача идентификации веществ по структурным формулам полностью решена. Для этого используют *четыре* способа:

- 1) непосредственное введение структурной формулы в базу;
- 2) распознавание нарисованной структуры;
- 3) выражение структуры цепочкой текстовых символов;
- 4) автоматическое генерирование систематических названий по формулам.

Самый новый, но при этом самый удобный для пользователя способ — использование интерфейса компьютерных баз данных, позволяющего ввести структурную формулу и искать информацию по ней. Этот интерфейс реализован, в частности, службой PubChem. Введя структурную формулу при помощи весьма простого инструментария (или загрузив её изображение в формате *.jpg), пользователь получает PCID для этого соединения, основные сведения о нём и множество ссылок на оригинальные работы.

Гораздо раньше (в конце 1980-х гг.), когда компьютеры ещё не воспринимали графическую информацию, появился способ отображения структуры цепочкой текстовых символов. Он называется SMILES, от «Simplified Molecular Input Line Entry Specification» — упрощённая спецификация строчного ввода молекул (www.opensmiles.org). Например, формула метилового оранжевого в SMILES выглядит так: OS(=O)(=O)c2ccc(/N=N/c1ccc(cc1)N(C)C)cc2, формула сульфолера — s1c2sc3sc4sc5sc9sc8sc7sc1c6c2c3c4c5c9c8c67, а формула Stains-all —

[Br-].CCN4c5c6cccc6ccc5S\C4=C/C(/C)=C\Cc3sc2ccc1cccc1c2[n+]3CC. В общем случае SMILES отображает структуру в гораздо более сжатом виде, чем номенклатурное название, и правила SMILES гораздо легче. При некотором навыке даже удаётся записывать и понимать SMILES без компьютера.

SMILES поддерживается Википедией и PubChem. Последняя позволяет генерировать SMILES по структурной формуле, которую можно ввести прямо в интерфейсе базы данных. Также существует программа онлайн-генерации SMILES по структуре и наоборот (cactus.nci.nih.gov/services/translate). Такая же возможность есть в программе ChemSketch, предназначенной для рисования структурных формул¹. Чтобы сгенерировать цепочку SMILES для нарисованной структуры, структуру выделяют и используют пункт меню Tools:Generate:SMILES Notation. Чтобы нарисовать структуру по известной цепочке SMILES, используют меню Tools:Generate:Structure from SMILES.

Другую попытку отобразить структуру в виде текстовых символов предпринял IUPAC, разработав в 2005 г. Международный химический идентификатор (IUPAC International Chemical Identifier, InChI [10]). Руководство к нему, в отличие от номенклатуры, занимает «всего» 34 страницы, поэтому генерировать и воспринимать InChI «вручную» сложно (хотя и можно). Однако существует открытая программа (www.iupac.org/inchi/download/), специально разработанная для этой цели. Кроме того, переводить структуру в InChI и обратно можно в программе ChemSketch (Tools:Generate:InChI for Structure). В отличие от CAS RN пользование InChI и идентификация по нему бесплатны. Его поддерживает Википедия. Также цепочку InChI можно вводить в любой стандартный поисковик.

И наконец, генерацию названий по нарисованным структурным формулам можно считать данью уважения к уходящей система-

тической номенклатуре. Нужный инструмент обычно добавляют в программы для рисования структурных формул. Например, такая опция есть в программе ChemSketch. Чтобы сгенерировать название соединения (у свободной версии есть ряд ограничений по структурам, для которых это можно сделать), рисуют его структурную формулу, выделяют её и выбирают пункт меню Tools:Generate:Name for Structure. Появившееся название (на английском) можно скопировать в любую поисковую систему или базу данных. Однако название, которое генерируется этой программой, часто не соответствует названию, которое приводится в других базах данных, что лишней раз говорит о неэффективности систематической номенклатуры.

Вывод

Так что же должен знать и уметь школьник, чтобы в современных условиях найти информацию о том или ином веществе? Он, как и любой компьютерно грамотный человек, должен уметь пользоваться поисковыми системами и Википедией. Из специальных химических знаний и умений отметим следующие.

1. Знать систематическую номенклатуру на базовом уровне (включая суффиксы и префиксы, соответствующие основным функциональным группам) и уметь составлять простейшие названия (типа бутан-1-ол) по формуле и наоборот.
2. Знать о существовании CAS RN и PCID.
3. Знать о существовании баз PubChem и SigmaAldrich. Уметь вводить структурные формулы в базу PubChem.
4. Уметь рисовать формулы в программе ChemSketch, устанавливать их InChI и SMILES.

И ещё: для сколько-нибудь адекватного поиска химической информации нужно знать английский язык, ибо на нём существует гораздо больше информации, чем на русском. ■

ЛИТЕРАТУРА

1. Congres de nomenclature chimique, Geneve, 1892 // Bull. Soc. Chim. Paris. Ser. 3 7: XIII-XXIV, 1892. Доступно в Интернете по адресу <http://gallica.bnf.fr/ark:/12148/bpt6k2820064.image>.

¹ Программу ChemSketch можно скачать по адресу www.acdlabs.com. Она бесплатна для образовательных учреждений и частных лиц. Интерфейс исключительно английский, но интуитивно понятный для тех, кто хоть немного знает язык.

2. **Panico R., Powell W. H., Richer J.-C.** (1993). A guide to IUPAC nomenclature of organic compounds. Recommendations. Blackwell Science. Доступно в Интернете по адресу <http://www.acdlabs.com/iupac/nomenclature/>.

3. **Leigh G. J., Favre H. A., Metanomski W. V.** (1998). Principles of Chemical Nomenclature. A Guide to IUPAC Recommendations. Blackwell Science. Доступна по адресу http://old.iupac.org/publications/books/principles/principles_of_nomenclature.pdf.

4. **Favre Henri A., Hellwich Karl-Heinz, Moss G. P., Powell Warren H. and Traynham James G.** Corrections to A Guide to IUPAC Nomenclature of Organic Compounds. Recommendations, 1993. Pure Appl. Chem. Vol. 71, No. 7, pp. 1327–1330, 1999. Доступно в Интернете по адресу <http://www.chem.qmul.ac.uk/iupac/bibliog/errata.html>.

5. Nomenclature of Inorganic Chemistry. IUPAC Recommendations, 2005. RSC Publishing. Доступно в Ин-

тернете по адресу http://old.iupac.org/publications/books/rbook/Red_Book_2005.pdf.

6. **Лурье Ю. Ю.** Справочник по аналитической химии. — М.: Химия, 1989. — С. 197.

7. Chemical Abstracts Service Statistical Summary. 1907–2007. <http://www.cas.org/ASSETS/836E3804111B49BFA28B95BD1B40CD0F/casstats.pdf>.

8. www.cas.org.

9. Communication Sulflower: A New Form of Carbon Sulfide / Konstantin Yu. Chernichenko, Viktor V. Sumerin, Roman V. Shpanchenko, Elizabeth S. Balenkova, Valentine G. Nenajdenko // Angewandte Chemie International Edition. Vol. 45, Issue 44, pp. 7367–7370, 2006.

10. **Stein Stephen E., Heller Stephen R., Tchekhovskoi Dmitrii V.** The IUPAC Chemical Identifier — Technical Manual. http://sunet.dl.sourceforge.net/project/ninja/technical_manual/inchi-tech-manual-20060511/inchi_tech_man-20060511.pdf.

Ключевые слова: номенклатура, химическая информация, химические базы данных.

Key words: nomenclature, chemical information, chemical data bases.

ИССЛЕДОВАНИЯ, ОТКРЫТИЯ, ПРОГНОЗЫ

Наномембраны

Комбинация неорганических наночастиц и органических молекул позволяет изготавливать новые материалы с уникальными свойствами. Например, внедрение наночастиц в тонкие полимерные плёнки приводит к существенному улучшению их термохимических характеристик. Однако из-за того, что объёмная доля наночастиц в таких композитах обычно мала и они расположены друг относительно друга нерегулярным образом, длинные цепочки полимерной матрицы перепутываются и прочность композита падает.

И вот недавно американские исследователи создали нанокомпозит, в котором плотноупакованный квазидвумерный монослой образуют сферические наночастицы золота, прочно сцепленные друг с другом посредством молекул $\text{CH}_3(\text{CH}_2)_{11}\text{SH}$ (додекантиола). Расстояние между соседними наночастицами (1,7 нм) намного меньше их диаметра (6 нм), поэтому молекулы додекантиола, соединяющие наночастицы, не перепутываются и прочность монослоя значительно возрастает. Если такой монослой сформировать на подложке с предварительно проделанным круглым отверстием диаметром ~1 мкм, на отверстии образуется мембрана.

Измерения, выполненные с помощью атомно-силового микроскопа, показали, что такие мембраны хоть и тонкие, но очень прочные (модуль Юнга в среднем составляет около 6 ГПа). При этом они обладают высокой эластичностью (изгибаются под прямым углом вблизи краёв отверстия на длине ~10 нм, что равно нескольким диаметрам наночастиц) и упругостью, которую сохраняют даже при нагревании (механические повреждения после воздействия иглы микроскопа отсутствуют вплоть до $T \approx 400$ К).

Для экспериментаторов это стало неожиданно, поскольку теория предсказывает плавление сверхрешётки Au/додекантиол уже при $T \approx 350$ К. Термостойкость обусловлена, по-видимому, жёсткой фиксацией края мембраны, благодаря чему расстояние между наночастицами при нагревании почти не изменяется.

В отличие от большинства полимеров системы плотноупакованных наночастиц проводят электрический ток — за счёт туннельного механизма. Поскольку туннельное сопротивление экспоненциально зависит от расстояния между наночастицами, наномембраны можно использовать в качестве очень чувствительных электронных датчиков давления.

Природа. — 2008. — № 4. — С. 82