

Веб-ресурси вільного доступу для хіміків-біоорганіків

Інтернет — величезне джерело практично будь-якої інформації. Існують сотні мільйонів веб-сторінок, частина яких цікава для хіміків-органіків, біоорганіків, біохіміків і спеціалістів суміжних галузей науки. Мета цієї статті — дати загальні уявлення про веб-ресурси, корисні для біоорганіків, про деякі важливі «відправні точки» хімічного Інтернету й можливості пошуку потрібної інформації. Акцент у статті робиться на ресурси вільного доступу, тобто безплатні. Слід відзначити, що за оцінками понад 80 % інформації, доступної в світовому Інтернеті, є англійською (а в галузі науки й технології цей відсоток ще вищий), відповідно переважна більшість веб-ресурсів для спеціалістів у галузі хімічних і біологічних наук англійською.

Веб-директорії. Пошук інформації проводять за допомогою двох основних інструментів — директорій і пошукових систем. Директорії — списки веб-сайтів, у тій чи іншій мірі класифіковані й індексовані, часто з коротким описом відповідних ресурсів. Існують і спеціалізовані хімічні директорії, часто з дуже великим обсягом інформації. Особливо корисними є портали Chemie.DE (www.chemie.de), ChemDex (www.chemdex.org), Organic Chemistry Resources Worldwide (www.organicworldwide.net), Rolf Claessen's Chemistry Index (www.claessen.net/chemistry), Network Science (www.netsci.org/Resources/Web/index.html). Вони містять велику кількість посилань на сайти наукових установ і фірм, журналів, професійних товариств і фондів, хімічні бази даних, комп'ютерні програми для хіміків тощо. Багаті каталоги лінків на хімічні ресурси викладені на серверах хімічних факультетів Ліверпульського (www.liv.ac.uk/Chemistry/Links/links.html) та Кембриджського університетів (www.ch.cam.ac.uk/c2k). Факультет біоорганічної хімії Уппсальського університету (www.boc.uu.se/boc14www/www_links/Links_general.html) надає, окрім іншого, велику добірку лінків на ресурси, пов'язані з ЯМР і молекулярним моделюванням. Велика колекція лінків на ЯМР-ресурси розміщена на сайті Інституту органічної хімії РАН (www.nmr.ioc.ac.ru/flinks.htm).

Портал ChemWeb (www.chemweb.com) — своєрідний онлайн-клуб для міжнародного хімічного співтовариства. Існує багато його мовних версій, у т.ч. й українська. Там можна знайти колекцію журналів, баз даних, тижневик «The Alchemist» (новини, огляди, коментарі), є можливість пошуку інформації про структури. Корисним ресурсом є Organic Chemistry Portal (www.organic-chemistry.org) з великою, добре структурованою колекцією лінків на хімічні ресурси будь-якого напрямку, багатьма даними про хімічні трансформації (від іменних реакцій до захисних груп у синтезі), пошуком синтетичних методів і фірм, які поставляють конкретні реагенти й синтони та ін.

Серед українських порталів, що мають колекції посилань на спеціалізовані хімічні й біохімічні ресурси, відзначимо BioChemWeb (www.biochemweb.org.ua) і BioUA (www.bioua.org.ua/index.php). Велику колекцію лінків (більшість із них російськомовні), базу даних властивостей хімічних сполук, а також багато довідкових матеріалів знайдемо на сайті Chemister (www.chemister.da.ru/Links/links.htm).

Пошукові системи. Директорії дають змогу крок за кроком знаходити потрібну, проте насамперед відносно загальну інформацію, більш спеціальні дані можна знайти за допомогою пошукових систем. Пошукова система (search engine), по суті, є індексом веб-сайтів, що дозволяє користувачам здійснювати пошук за певними параметрами. Безумовно, якість і повнота пошуку інформації залежить від повноти охоплення веб-ресурсів і технології пошуку, закладеної в системі. Дуже важливою є коректна побудова пошукового запиту користувачем, що вимагає певної практики. Пошукові системи загально-спрямування, наприклад, усім відомі Google, Alta Vista, Excite, AllTheWeb, Lycos, Рамблер, Яндекс, МЕТА та ін., часто можуть бути корисними і для хіміків-професіоналів. Існують і загальнонаукові пошукові системи, наприклад, SciSeek (www.sciseek.com), Search4Science (www.search4science.com), SiteSeer (www.citeseer.nj.nec.com), а також спеціалізовані хімічні, про які йтиметься далі. Списки наукових пошукових систем викладені на сайтах SearchAbility (www.searchability.com/academic.htm) і Лейденського університету (www.leidenuniv.nl/ub/biv/specials.htm).

Різниця між директоріями й пошуковими системами сьогодні не така очевидна, як на ранніх етапах розвитку Інтернету. Зараз більшість пошукових систем мають власні директорії і навпаки. Безумовно, жодна пошукова система не охоплює весь Веб, тому для підвищення ефективності пошуку корисно

провести його за допомогою кількох пошуковиків, чи то в індивідуальному порядку, чи то із застосуванням так званого мета-пошуку — у системах, що здійснюють пошук одночасно через кілька окремих пошукових систем, наприклад, Dogpile (www.dogpile.com), Go2Net (www.go2net.com), MetaCrawler (www.metacrawler.com) та SavvySearch (www.savvysearch.com). Можна використати і спеціальні пошукові програми, які проводять одночасний пошук через велику кількість (часто кілька десятків) пошуковиків. Найвідоміші програми такого типу — Copernic і Search+. Як онлайніві мета-пошукові системи, так і відповідні програми об'єднують результати, отримані від індивідуальних пошуковиків.

Наведені вище пошукові системи, як уже було сказано, мають загальний характер і спрямовані на охоплення максимуму інформації в усіх сферах знання. Кориснішими для науковців є спеціалізовані пошукові системи й директорії, які охоплюють окремі сегменти науки й технології, проте дають змогу знайти більш адекватну інформацію. Зупинимось детальніше на Інтернет-ресурсах з академічним контентом, необхідним для досліджень у галузі хіміко-біологічних наук. Насамперед нас цікавлять можливості пошуку наукової літератури.

Пошук хімічної літератури в Інтернеті. Практично всі більш-менш серйозні наукові журнали сьогодні доступні через Інтернет, у т.ч. й повнотекстові. Існують і онлайніві журнали, які не мають друкованої версії. Корисні ресурси для знаходження наукових публікацій — загальнонаукові пошукові системи, про які йшлося вище. Можна використовувати портали видавців академічних публікацій. Найбільші такі ресурси — Science Direct (www.sciencedirect.com) і портал журналів Американського хімічного товариства (www.pubs.acs.org). Багато профільних журналів доступні на сайтах інших найбільших видавців — Springer (www.springer.com), Wiley Interscience (www3.interscience.wiley.com), Oxford University Press (www.oup.co.uk) та ін. Онлайніві версії журналів російського видавництва «Наука» («Биоорганическая химия», «Биохимия», «Журнал органической химии» та ін.) можна знайти за адресою: www.maik.rssi.ru/win/online/index.htm. Основні профільні українські видання, що мають повнотекстові онлайніві версії, — Ukrainica Bioorganica Acta (www.bioorganica.org.ua), «Біополімери та клітина» (www.biopolymers.org.ua), «Український біохімічний журнал» (www.ubj.biochemistry.org.ua). Значно ефективнішим, як правило, є пошук літератури у спеціальних бібліографічних пошукових системах.

Однією з провідних і найкращих спеціалізованих систем пошуку наукової бібліографічної інформації є Scirus (www.scirus.com), що належить видавництву Elsevier. Scirus здійснює пошук насамперед серед журналів цього видавництва, з-поміж яких дуже багато хімічних і біологічних видань (у т.ч. Tetrahedron, Bioorganic and Medicinal Chemistry та багато інших), і разом із тим проводить пошук за заданими параметрами та у Вебі загалом, у т.ч. в системах Medline та BioMedNet, а також надає патентну інформацію.

Подібною до Scirus є бібліографічна пошукова система Scopus (www.scopus.com). Google Scholar (www.scholar.google.com) — ще один чудовий пошуковий ресурс, близький за можливостями до Scirus.

Американський центр National Center for Biotechnology Information, що входить у систему Національних Інститутів Здоров'я (НИН), створив ряд баз даних вільного доступу високого рівня. Надзвичайно популярною у фахівців у галузі наук про життя є система PubMed-Medline (www.ncbi.nlm.nih.gov/sites/entrez). Це бібліографічна база даних, що дає змогу здійснювати ефективний пошук наукової літератури за ключовими словами. На момент написання статті ця база охоплювала близько 15 млн посилань. Medline індексує велику кількість журналів, у т.ч. з біоорганічної та медичної хімії.

ChemRefer (www.chemrefer.com) — система повнотекстового пошуку хімічної і фармацевтичної літератури. Доступ до цієї пошукової системи, як і до попередніх, є вільним. Широкі можливості пошуку хімічної літератури має німецький хіміко-інформаційний центр FIZ CHEMIE Berlin (www.fiz-chemie.de/index.html.en), де пошук ведеться в базах ChemInform, Infotherm, ChemGuide і MedPharmGuide.

Безумовно, слід мати на увазі, що в більшості випадків веб-ресурси (наукові статті, інформація з баз даних тощо), знайдені науковими пошуковими системами, є комерційними й недоступними в повному обсязі. Сьогодні це є однією з проблем Інтернету в наших умовах, оскільки достатньо легко знайти великі обсяги потрібних даних, які, однак, не перебувають у вільному доступі. Хоча абстрак-

ти статей практично завжди відкриті, доступ до повнотекстових публікацій найчастіше є платним. Цікаво відзначити, що видавці деяких журналів (наприклад, *Nucleic Acids Research* та ін.) пішли іншим шляхом — їх повнотекстові публікації знаходяться у вільному доступі, проте авторам статті доводиться сплачувати вартість її публікації (суми звичайно понад \$1000).

Багато інформації можна відшукати і в наукових журналах із відкритим доступом (Open Access). Велику кількість таких видань розміщено на порталі BioMed Central (www.biomedcentral.com). Підрозділом PubMed є PubMed Central — архів наукової біомедичної літератури, що забезпечує вільний доступ до повнотекстових статей (www.pubmedcentral.nih.gov). На жаль, більшість журналів, доступних на цих порталах, мають біологічний профіль. Недавно запущений і аналогічний хімічний сервіс Chemistry Central (www.chemistrycentral.com).

Патенти в Інтернеті. Патентна інформація стає все більш актуальною для хіміків-синтетиків, причому повнотекстові патенти в цілому значно доступніші за статті в академічних журналах. Так, патенти США вільно доступні за адресою: www.uspto.gov/patft/index.html. Патенти у форматі HTML з можливістю пошуку доступні з 1976 р. Починаючи з 1790 р., вони представлені у вигляді посторінково відсканованих документів (TIF-файли).

Інші основні бази патентів також безплатні. Серед них esp@cenet (www.ep.espacenet.com, асоційований з Європейським патентним офісом), РСТ (міжнародна система Patent Cooperation Treaty, www.wipo.int/pctdb/en), база Британського патентного офіса (www.ipo.gov.uk/patent.htm). Усюди можливий пошук як за номером патенту, так і за ключовими словами. Німецькі патенти доступні на порталі DEPATISNET (www.depatiset.dpma.de). Система пошуку добре організована, крім німецьких, тут можна знайти патенти і багатьох інших країн, часто повнотекстові. В усіх цих базах патенти подаються у форматі PDF. Інші великі бази патентної інформації (хоча не обов'язково безплатної!) — Delphion (www.delphion.com), Derwent (www.derwent.com), Micropatent (www.micropat.com) і Direct Patents (www.direct-patent.n).

Оптимізація пошуку. Для того, щоб швидше й ефективніше отримати потрібну інформацію, в Інтернеті існують шляхи оптимізації пошуку. Насамперед необхідно максимально коректно задавати умови пошуку для мінімізації кількості отриманих посилань (часто їх сотні тисяч і мільйони). Одним із найпростіших методів є пошук фрази (як правило, її треба взяти в лапки) замість набору ключових слів. Так, на запит *nucleoside synthesis* пошуковик Google видає понад 1,6 млн посилань, тоді як на запит у форматі «*nucleoside synthesis*» — лише близько 19 000, і в отриманому списку вже відсутні посилання на синтез, не пов'язаний конкретно з нуклеозидами. Інший шлях — використання логічних операторів. У найпростішому випадку включення чи виключення специфічного терміна із запиту досягається додаванням префікса + чи - перед терміном. Наприклад, запит *drugs +antiinflammatory -aspirin* запустить пошук протизапальних препаратів, що відмінні від аспірину. Більш гнучкі методи пошуку використовують інтуїтивно зрозумілі логічні оператори AND, OR, NOT (деякі пошукові системи дають змогу використовувати й інші оператори). Ці оператори в комбінації з дужками дозволяють задавати достатньо складні умови пошуку. Правила організації запиту можуть відрізнитись у різних пошукових системах, тому в кожному конкретному випадку варто ознайомитись із ними перед тим, як почати застосовувати складні варіанти пошуку.

Пошукові системи часто дають змогу обмежити пошук певними додатковими параметрами. Наприклад, пошук тільки в заголовках веб-сторінок, пошук документів заданого формату (наприклад, лише PDF-файлів) і навіть обмеження доменів, у яких проводиться пошук. Щоб уникнути нераціонального витрачання часу на пошук спеціалізованої інформації, варто починати його на добре відомих порталах, присвячених хіміко-біологічним наукам.

Пошук хімічних структур у базах даних. Більшість баз даних хімічного характеру (властивості молекул, методи синтезу, відповідна література й патенти тощо) є комерційними. Найвідоміші з них — Beilstein, SciFinder, STN. Безумовно, бази вільного доступу навряд чи здатні конкурувати з ними в повному обсязі й не містять такої багатой інформації. Однак прекрасних ресурсів цього класу з вільним доступом в Інтернеті теж багато. Серед безплатних баз відмітимо Chemfinder (www.chemfinder.camsoft.com), де існує можливість пошуку за назвою, CAS-індексом чи за структурною фор-

мулою (використовується відповідний безплатний плагін ChemDraw Net). Серед виданої інформації — фізико-хімічні властивості молекули, лінки на сайти з додатковою інформацією про молекулу, а також на фірми-продавці того чи іншого реагенту.

NIST Chemistry WebBook (www.webbook.nist.gov/chemistry) дозволяє вести пошук фізико-хімічних властивостей молекул (за назвою, структурою, CAS-індексом, молекулярною формулою тощо). При цьому надаються термодинамічні характеристики, а в багатьох випадках навіть УФ-, ІЧ- і мас-спектри сполук.

ChemBank (www.chembank.broad.harvard.edu) — публічна база хіміко-біологічних даних, що фінансується Національним Інститутом Раку США. Вона дає змогу здійснювати пошук органічних молекул за субструктурами, назвою, функцією, дескрипторами і навіть за прізвищем хіміка, який уперше синтезував сполуку. Надаються дані щодо біологічної активності молекул. ChemBank корисний як для хіміків, що синтезують нові сполуки, так і для біологів, які шукають невеликі молекули, що здатні впливати на певні біологічні процеси.

Подібною до ChemBank є ChEBI (Chemical Entities of Biological Interest, www.ebi.ac.uk/chebi) — база вільного доступу, яка фінансується Європейською Комісією.

Відносно недавно запущена база PubChem (www.pubchem.ncbi.nlm.nih.gov) — аналог PubMed, спеціалізований на біохімічній інформації. Тут існує пошук молекул за назвою, структурою та ін., у результаті можна отримати дані про біологічну активність сполук.

ChemSpider (www.chemspider.com) — постійно зростаюча онлайн база хімічних структур. На момент написання статті вона містила 16,5 млн індивідуальних структур. Фактично це пошукова машина структур та асоційованої з ними інформації. У результаті пошуку отримуємо, окрім іншого, набір розрахованих (чи експериментальних) параметрів типу logP, корисних для досліджень у галузі drug design. База eMolecules (www.emolecules.com) дозволяє знайти фізико-хімічні дані молекул за структурою, а також інформацію про виробників цих сполук (у базі понад 7 млн структур).

На порталі ChemSpy (www.chemspy.com/index.html) можна проводити пошук інформації в багатьох базах даних одночасно. Система ChemIDplus Національної бібліотеки медицини США (www.chem.sis.nlm.nih.gov/chemidplus) також здійснює пошук даних про речовину за структурою, назвою, CAS та іншими параметрами в кількох десятках баз.

Окремо варто відзначити базу Organic Syntheses (www.orgsyn.org), в якій можна отримати методику синтезу й використання в синтезі сполук як за ключовими словами, так і за структурою (із застосуванням плагіну ChemDraw).

Колекція посилань на бази даних спектрів органічних сполук (ЯМР, ІЧ, УФ та ін.) знаходиться за адресою: www.chemexpress.fatal.ru/Navigator/Spectrumfinder.html. Серед них відмітимо японську базу Spectral Database for Organic Compounds SDBS (www.aist.go.jp/RIODB/SDBS/cgi-bin/cre_index.cgi).

Великий список хімічних і біологічних баз різного профілю викладено на порталі ChemBioGrid (www.chembiogrid.org/related/resources/databases.html).

Пошук виробників реагентів. З метою пошуку фірм, що виробляють певні реагенти й синтони, у т.ч. й рідкісні, чи надають сервіси у галузі хімічного синтезу, рекомендуємо, крім згаданих вище порталів, відвідати сайти ChemExper (www.chemexper.com), ChemSources (www.chemsources.com), Chem.com (www.chem.com), ChemConnect (www.wce.chemconnect.com), ChemNavigator (www.chemnavigator.com), Chemolink (www.chemolink.com), ChemIndustry (www.chemindustry.com).

Програмне забезпечення. Програмні пакети для наукових досліджень часто надто дорогі. На щастя, в Інтернеті доступна велика кількість якісного безплатного софту практично всіх видів, що необхідні хімікові-біоорганіку. Головні з них, безумовно, програми для візуалізації хімічної інформації.

Потужні багатофункціональні хімічні редактори ChemDraw (фірми Cambridge Soft) та ChemWindow (Softshell, зараз перейшла до Bio-Rad) є комерційними. Однак прекрасна програма векторної хімічної графіки ISIS/Draw (остання версія 2.5) є безплатною і може бути завантажена після простої реєстрації з сайту MDL Information Systems (www.mdli.com/download/index.html). Цей хімічний редактор входить і в комерційний пакет ISIS/Base. На цьому ж сайті доступна програма Chime — плагін для

веб-браузерів (підтримуються не всі), що дає змогу створювати й переглядати веб-сторінки, які містять 3-мірні хімічні структури. Структурами можна інтерактивно маніпулювати безпосередньо на сторінці в реальному часі з браузера. Іншою надзвичайно цінною безплатною програмою від MDL є Autonom — програма, що дозволяє давати коректні назви хімічним структурам згідно з номенклатурою IUPAC. Програма є плагіном до ISIS/Draw, існує і версія для ChemDraw.

Подібним до ISIS/Draw, а в багатьох аспектах і значно більш «просунутим», є безплатний хімічний редактор ChemSketch канадської фірми Advanced Chemistry Development (ACD Labs, www.acdlabs.com). Остання на час написання статті версія 10.0 включає, окрім іншого, калькулятор фізико-хімічних властивостей молекул, функцію створення анімованих GIF-файлів із 3D Viewer, експорт документів у формат PDF, розрахунок ^{13}C та ^1H ЯМР-спектрів, створення звітів і постерів. Фірма ACD Labs пропонує і ряд інших безплатних програм, наприклад, адони до ChemSketch, що дають змогу вести пошук у базах ChemSpider, eMolecules та PubChem, програма для обчислення $\log P$ та ін. Мають, найвідомішою серед них є ACD/Name — подібна до Autonom програма з функцією «structure-to-name». Існують версії для ChemSketch (входить у цей пакет), ISIS/Draw та ChemDraw (у вигляді плагінів). Безплатна версія ACD/Name має певні функціональні обмеження (наприклад, даються назви для структур, в які входить не більше 50 атомів або 3 циклів). Позбавлена обмежень комерційна версія ACD/Name Pro, яка до того ж, крім назв за правилами IUPAC, генерує і назви у варіанті номенклатури, прийнятому в Chemical Abstracts (Index Name).

Чудовий інтегрований програмний пакет із широкими можливостями KnowItAll Academic Edition фірми Bio-Rad (www.knowitall.com/academic/welcome.asp) доступний для безплатного завантаження після реєстрації для викладачів, студентів, аспірантів, науковців академічних інститутів. Крім звичайної функції «рисування формул», цей пакет дозволяє оперувати 3-мірними структурами й оптимізувати їх, працювати з ГЧ-, ЯМР- та раманівськими спектрами, розраховувати фрагментацію молекул у мас-спектрах, створювати презентації тощо.

RasMol — корисна безплатна програма для перегляду 3-мірних структур — від простих молекул до білків і нуклеїнових кислот. Вона може бути завантажена, наприклад, із сайту університету штату Масачусетс (www.umass.edu/microbio/rasmol) і входить у пакет ISIS/Draw. Подібну програму 3D Viewer включено до пакета ChemSketch. Аналогічною програмою візуалізації молекулярної графіки є Swiss PDB Viewer (www.expasy.ch/spdbv). Цей в'ювер розроблений в першу чергу для роботи з 3-мірними білковими структурами, наприклад, імпортованими з бази Protein Data Bank.

Загальновідомі програми для роботи з базами хімічних структур — ISIS/Base (MDL) та Chemfinder (Cambridge Soft) — є дорогими комерційними продуктами, тому ми не будемо зупинятися на них. На жаль, схоже, що якісних безплатних альтернатив немає.

Існують і безплатні програми з хімічної кінетики, кристалографії, спектроскопії та ін., які ми не в змозі охопити в цій публікації. Добірки програм, корисних для хіміків-біоорганіків, можна знайти на багатьох сайтах. Так, велика колекція безплатних хімічних програм на всі випадки життя знаходиться на сайті Tom's Free Chemistry Software (www.users.ugent.be/~tkuppens/chem.shtml). Хімічні програми для Linux знайдемо на сайті Linux4Chemistry (www.redbrick.dcu.ie/~noel/linux4chemistry). Більшість із них є програмами Open Source, безплатними або ж безплатними для використання в академічних дослідженнях.

У зв'язку з обмеженим обсягом публікації ми не могли зупинитися на багатьох інших ресурсах Інтернету для хіміків-біоорганіків, пов'язаних, наприклад, із молекулярним моделюванням та drug design, квантовою хімією, базами хімічних реакцій, хімічними форумами й конференціями та ін. Інтернет настільки багатогранний і масштабний, що охопити навіть частину його ресурсів у невеликому огляді вкрай складно. Тому закликаємо читачів шукати потрібні для них веб-ресурси самостійно. Практично в усіх випадках вдасться виявити ресурси вільного доступу, багато з яких не поступаються, а то й переважають за своїм рівнем комерційні аналоги.

І.Я. Дубей,
к.х.н., пров. науковий співробітник,
Інститут молекулярної біології і генетики НАН України