



Komputerowo wspomagana synteza organiczna związków biologicznie aktywnych

mgr inż. Tomasz Klucznik

Promotor

prof. dr Bartosz Grzybowski

Praca przedstawiona Radzie Naukowej
Instytutu Chemii Organicznej
Polskiej Akademii Nauk
celem uzyskania stopnia
doktora nauk chemicznych

Biblioteka Instytutu Chemii Organicznej PAN

Org.-B.417/20



8000000341314

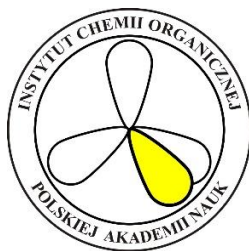
A-21-6, K-c-121, K-c-130

Warszawa, 2019



B. Org. 417/20

Część doświadczalna niniejszej pracy powstała
w Instytucie Chemii Organicznej Polskiej Akademii Nauk
w latach 2013-2019



Badania były finansowane z grantu SYMFONIA 2014/12/W/ST5/00592
przyznanego przez Narodowe Centrum Nauki



Wyniki mojej pracy zostały częściowo zawarte w publikacjach:

1. S. Szymkuć, E. P. Gajewska, **T. Klucznik**, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski, *Angew. Chem. Int. Edit.*, **2016**, 55(20), 5904-5937.
2. **T. Klucznik**, B. Mikulak-Klucznik, M. P. McCormack, H. Lima, S. Szymkuć, M. Bhowmick, K. Molga, Y. Zhou, L. Rickershauser, E. P. Gajewska, A. Touthkine, P. Dittwald, M. P. Startek, G. J. Kirkovits, R. Roszak, A. Adamski, B. Sieredzińska, M. Mrsksich, S. L. J. Trice, B. A. Grzybowski, *Chem*, **2018**, 4(3), 522-532.
3. B. A. Grzybowski, S. Szymkuć, E. P. Gajewska, K. Molga, P. Dittwald, A. Wołos, **T. Klucznik**, *Chem*, **2018**, 4(3), 390-398.

Publikacje nieobjęte niniejszą rozprawą:

1. W. Jaworski, S. Szymkuć, B. Mikulak-Klucznik, K. Piecuch, **T. Klucznik**, M. Kaźmierowski, J. Rydzewski, A. Gambin, B. A. Grzybowski, *Nature Commun.*, **2019**, 10(1), 1434: 1-11.

Wyniki mojej pracy zaprezentowane w formie referatów na konferencjach:

1. Cambridge-Warsaw Young Scientists Meeting 2014
14.09-16.09.2014 Cambridge (Anglia)
Chematica – a machine that thinks like a chemist
T. Klucznik, E. Gajewska, B. A. Grzybowski
2. XLIII Ogólnopolska Szkoła Chemii
7.11-11.11.2015, Pokrzywna
Chematica – komputerowo planowana synteza organiczna
T. Klucznik, S. Szymkuć, E. Gajewska, K. Molga, P. Dittwald, M. Bajczyk, B. Mikulak, B. Sieredzińska, B. A. Grzybowski
Nagroda za najlepsze wystąpienie doktoranta
3. XLIV Ogólnopolska Szkoła Chemii
30.04-4.05.2016, Koszuty
Chematica – komputerowo planowana synteza totalna
T. Klucznik, S. Szymkuć, E. Gajewska, K. Molga, P. Dittwald, M. Bajczyk, B. Mikulak, B. Sieredzińska, B. A. Grzybowski
4. XLVI Ogólnopolska Szkoła Chemii
30.04-4.05.2017, Jachranka
Komputerowo wspomagana synteza organiczna. Studium praktyczne
T. Klucznik, B. Mikulak, S. Szymkuć, E. Gajewska, K. Molga, P. Dittwald, B. A. Grzybowski

5. XLVIII Ogólnopolska Szkoła Chemii
28.04-2.05.2018, Karczowiska
Komputerowo wspomagana synteza organiczna związków biologicznie aktywnych.
T. Klucznik, B. Mikulak-Klucznik, A. Adamski, B. Sieredzińska,
B. A. Grzybowski

Wyniki mojej pracy zaprezentowane w formie posterów na konferencjach:

1. IUPAC–2015, 45th World Chemistry Congress
9.08-14.08.2015 Busan, (Korea Południowa)
Retrosynthesis of complex molecules using Chematica
T. Klucznik, S. A. Szymkuć, E. P. Gajewska, K. Molga,
P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski
2. International Conference of Organic Synthesis, Balticum Organicum Syntheticum
1.07-4.07.2018, Tallin (Estonia)
Computer-planned organic synthesis of medicinally relevant targets
T. Klucznik, B. Mikulak-Klucznik, A. Adamski, B. Sieredzińska,
B. A. Grzybowski
3. 21st JCF Frühjahrssymposium
20.03-23.03.2019 Brema (Niemcy)
Computer-planned organic synthesis of medicinally relevant targets
T. Klucznik, B. Mikulak-Klucznik, A. Adamski, B. Sieredzińska,
B. A. Grzybowski
4. International Symposium on Synthesis and Catalysis
3.09-6.09.2019 Évora (Portugalia)
Computer-planned organic synthesis and its experimental verification
T. Klucznik, B. Mikulak-Klucznik, A. Adamski, B. Sieredzińska,
B. A. Grzybowski

Oraz wykładu na zaproszenie:

Chematica – komputerowo wspomagana synteza organiczna

Wykład na zaproszenie, w ramach cyklu: *Mam ochotę na chemię*, Warszawa, 30.03.2016

T. Klucznik

*Dziękuję mojemu promotorowi
Panu profesorowi Bartoszowi Grzybowskiemu
za lata opieki merytorycznej nad moją pracą,
wszystkie udzielone mi lekcje i wskazówki
oraz za danie okazji zmierzenia się
z wyjątkowym naukowym wyzwaniem*

*Dziękuję wszystkim członkom zespołu XI za lata wspólnej,
zwieńczonej sukcesem pracy*

*Szczególne podziękowania składam
żonie Barbarze Mikulak-Klucznik
za wszystkie wspólnie pokonane zagadki chemiczne
oraz nieustające wsparcie*

Lista używanych skrótów

9-BBN	9-Borabicyklo(3.3.1)nonan
Å	Ångstrom, jednostka długości równa 10^{-10} m
Ac	grupa acylowa
B ⁻	zasada
BINAP	2,2'-bis(difenylofosfino)-1,1'-binaftyl
Bu	grupa butylowa
cod	1,5-cyklooktadien
DACH	1,2-diaminocykloheksan
DCC	<i>N,N'</i> -dicykloheksylokarbodiimid
DCM	dichlorometan
DIBALH	wodorek diizobutyloglinu
DIEA	<i>N,N</i> -diizopropylloetyloamina
DIPEA	<i>N,N</i> -diizopropylloetyloamina
DMAP	4-dimatyloaminopirydyna
DMCyA	<i>N,N</i> -dimetyloaminocykloheksan
DMF	dimetyloformamid
DMSO	dimetylosulfotlenek
DOI	<i>ang. digital object identifier</i> , (cyfrowy identyfikator dokumentu elektronicznego)
dppp	1,3-bis(difenylofosfino)propan
Et	grupa etylowa
ESI-MS	spektrometria mas wykorzystująca elektrosprej
eq.	ekwiwalent
godz.	godzina
HPLC	wysokosprawna chromatografia cieczowa
HRMS	wysokorozdzielcza spektrometria mas
kat.	katalizator, ilość katalityczna
LDA	diizopropylloamidek litu
Me	grupa metylowa
min.	minuta

MS	<i>ang. molecular sieves</i> (sita molekularne)
Ms	grupa metylosulfonowa
nas.	nasycony
NMR	magnetyczny rezonans jądrowy
NOE	jądrowy efekt Overhausera
PCC	chlorochromian pirydyny
Pd G2	katalizator palladowy drugiej generacji
PE	eter naftowy
Ph	grupa fenylowa
Pr	grupa propylowa
pTsOH	kwask para-toluenosulfonowy
py	pirydyna
RAMP	(<i>R</i>)-(+)-1-amino-2-(metoksymetyleno)pirolidyna
RaNi	nikiel Raney'a
R _f	współczynnik opóźnienia w chromatografii
ROESY	spektroskopia NOE w rotującym układzie współrzędnych
roztw.	roztwór
SAMP	(<i>S</i>)-(-)-1-amino-2-(metoksymetyleno)pirolidyna
SPhos	2-dicykloheksylofosfino-2',6'-dimetoksybifenyl
TBAF	fluorek tetrabutylamoniowy
TBS	grupa <i>tert</i> -butylodimetylosililowa
THF	tetrahydrofuran
TLC	chromatografia cienkowarstwowa
t. pok.	temperatura pokojowa
t. wrz.	temperatura wrzenia
USPTO	<i>ang. United States Patent and Trademark Office</i> (Urząd Patentów i Znaków Towarowych Stanów Zjednoczonych)
XPhos	2-dicykloheksylofosfino-2',4',6'-triizopropyllobifenyl

Spis treści

Lista używanych skrótów	9
Spis treści	11
1. Założenia i cel pracy	15
1.1. Założenia	15
1.2. Cel pracy	15
2. Wstęp literaturowy	17
2.1. Historia programów komputerowych wspomagających planowanie wieloetapowych syntez chemicznych	17
2.1.1. Wstęp	17
2.1.2. OCSS & LHASA	18
2.1.2.1. Założenia i ogólna zasada działania	18
2.1.2.2. Strategie, kombinacje taktyczne i ekspercka baza danych	20
2.1.2.3. Percepcja struktury i kontekstu strukturalnego	25
2.1.2.4. Rezultaty i ich ocena	26
2.1.2.5. Podsumowanie	34
2.1.3. SYNCHEM i SYNCHEM2	36
2.1.3.1. Wprowadzenie	36
2.1.3.2. Rezultaty	37
2.1.3.3. Podsumowanie	41
2.1.4. SYNSUP	41
2.1.4.1. Wprowadzenie	41
2.1.4.2. Rezultaty	44
2.1.4.3. Podsumowanie	49
2.1.5. AIPHOS/KOSP	51
2.1.5.1. Wprowadzenie	51
2.1.5.2. Rezultaty	52
2.1.5.3. Podsumowanie	53
2.1.6. Inne programy	54
2.2. Współczesne programy komputerowe wspomagające planowanie wieloetapowych syntez chemicznych	55
2.2.1. Wstęp	55
2.2.2. Reaxys Synthesis planner & SciFinder SciPlanner	56
2.2.3. ARChem	56
2.2.4. ICSynth	57
2.2.5. Sieć neuronowa Seglera-Wallera	57
2.2.6. Podsumowanie	59

2.3. Chematica / Syntaurus / Synthia™	62
2.3.1. Wstęp	62
2.3.2. Założenia i ogólna zasada działania	62
2.3.3. Szczególne zasady działania	69
2.3.3.1. Aromatyczna substytucja elektrofilowa	69
2.3.3.2. Reakcja Dielsa-Aldera	71
2.3.3.3. Heurystyki konformacyjne	73
2.3.3.4. Dyskryminacja labilnych grup funkcyjnych	74
2.3.3.5. Ocena substratu do cyklizacji	74
2.3.3.6. Sekwencje strategiczne	75
2.3.4. Rezultaty	75
2.3.4.1. Synteza inhibitora białek BRD 7 i 9	76
2.3.4.2. Synteza α -hydroksyetizolamu	77
2.3.4.3. Synteza prototypowego leku na białaczkę	79
2.3.4.4. Synteza engelheptanoksydu C	79
2.3.5. Podsumowanie	81
3. Badania własne	82
3.1. Plan weryfikacji hipotezy badawczej	82
3.2. Tworzenie bazy reguł reakcji chemicznych	82
3.3. Weryfikacja syntetyczna planów programu Chematica	88
3.3.1. Synteza Dronedaronu (1)	88
3.3.1.1. Dyskusja planu z programu Chematica dla (1)	88
3.3.1.2. Adaptacja reakcji z planu, opis syntezy (1)	90
3.3.1.3. Podsumowanie	99
3.3.2. Synteza 5 β /6 β -hydroksylurazydonu (12,12')	100
3.3.2.1. Dyskusja planu z programu Chematica (12,12')	100
3.3.2.2. Adaptacja reakcji z planu, opis syntezy (12,12')	102
3.3.2.3. Podsumowanie	109
3.3.3. Synteza hydrazonu (23(R,R))	111
3.3.3.1. Dyskusja planu z programu Chematica (23(R,R))	111
3.3.3.2. Adaptacja reakcji z planu, opis syntezy (23(R,R))	113
3.3.3.3. Podsumowanie	119
3.4 Podsumowanie i wnioski	119
4. Część eksperymentalna	121
4.1. Stosowane metody syntetyczne i analityczne	121
4.2. Synteza Dronedaronu	122
4.2.1. Procedury syntetyczne zastosowane do uzyskania Dronedaronu (1)	122
4.2.2. Optymalizacja warunków karbonylatywnej annulacji (4)	127
4.3. Synteza 5 β /6 β -hydroksylurazydonu (12,12')	129
4.3.1. Procedury syntetyczne zastosowane do uzyskania	129

5 β /6 β -hydroksylurazydonu (12,12')	
4.4. Synteza hydrazonu (23(R,R)) w syntezie totalnej Takamonidyny (24)	134
4.4.1. Procedury syntetyczne zastosowane do uzyskania hydrazonu (23(R,R))	134
4.4.2. Określenie nadmiaru diastereoizomerycznego dla (23(R,R))	141
5. Bibliografia	142
Streszczenie rozprawy doktorskiej	153
Summary of doctoral thesis	154
Appendiks I. Widma NMR i chromatogramy	155
Appendiks II. Schematy pozostałych analiz retrosyntetycznych z LHASA, SYNCHEM i SYNSUP	184

1. Założenia i cel pracy

1.1. Założenia

Problem planowania syntezy związków organicznych o nietrywialnej strukturze stanowi od ponad wieku przedmiot debaty chemików. Można ów problem podzielić na dwie części: 1) sztukę projektowania syntezy i praktycznego jej wykonywania; oraz 2) próbę ustalenia na czym polega istota projektowania syntezy.

W sztuce projektowania syntezy w XX w. i pierwszych dekadach XXI w. dokonał się bezprecedensowy i niepodważalny postęp. Świadczą o nim miliony wykonanych syntetycznie związków chemicznych, częstokroć przy użyciu stworzonych przez człowieka narzędzi i reagentów nieznanymi w naturze. Owocami tej sztuki są tworzywa sztuczne, leki i inne substancje biologicznie aktywne. Przydatności i ważności nauk chemicznych w adaptacji gatunku ludzkiego do rzeczywistości materialnej nikt poważnie nie kwestionuje.

Miliony związków chemicznych uzyskanych na pożytek ludzkości, dowodzi skuteczności stosowanych metod projektowania syntez chemicznych. Tymczasem wciąż jest żywa dyskusja nad tym, czym jest samo to projektowanie. Osią sporu jest opozycja stanowisk: *planowanie syntezy to sztuka* kontra *planowanie syntezy to nauka*. Oczywiście niezbędne jest zgromadzenie szeregu danych naukowych by podejmować się planowania syntezy, jednak sam proces projektowania, do dziś pozostaje spowity nimbem artyzmu i mistyki. Nadaje to najwybitniejszym projektantom cząsteczek unikatowy status – naukowców artystów. Użycie komputera do planowania syntez może się wydawać próbą podważenia tego statusu.

Korzeni tego podejścia upatruje się wśród pierwszych badaczy zasłużonych dla rozwoju syntezy totalnej z R.B. Woodwardem na czele. W latach 1935-1955 chemicy dysponowali nader skąpym zasobem reagentów i zbadanych reakcji chemicznych w porównaniu ze stanem współczesnym. Skutkowało to koncentrowaniem się na reaktywności poszczególnych cząsteczek i możliwych do wykonania na nich przemian. Studiowanie chemii organicznej polegało na studiowaniu precedensów – w takiej sytuacji każda cząsteczka wydawała się unikatowym wyzwaniem. Tymczasem wzrost ilości i dostępności uniwersalnych reagentów, grup ochronnych, technik analizy strukturalnej oraz samych pomyślnie wykonanych syntez totalnych sprawiła, że zaczęto spoglądać poza horyzont pojedynczego celu syntetycznego ku ogólniejszym prawom rządzącym planowaniem syntez. Następnym krokiem było zaproponowanie przez E.J. Corey'a podejścia retrosyntetycznego i wykazanie jego użyteczności na przykładzie dziesiątek zakończonych sukcesem syntez totalnych. Dla projektantów syntez oznaczało to ponadto przesunięcie ciężaru ich dziedziny ze sztuki ku technice. Skoro więc istnieją reguły o potwierdzonej skuteczności, to czy i jak wiele z kreatywnej aktywności chemika-projektanta, można zastąpić za pomocą reguł, a jaka część nie poddaje się uregulowaniu?

Hipoteza badawcza, stojąca u podstawy moich badań brzmi: „Zdolność projektowania wykonalnych planów syntetycznych – cząsteczek organicznych – nie jest umiejętnością wyłącznie ludzką, lecz możliwe jest stworzenie programu komputerowego zdolność tą posiadającego.” Moja praca opiera się na założeniu, że dotychczas kreatywne i w dużej mierze pozostawione ludzkiej intuicji składowe planowania syntezy chemicznej można uregulować.

1.2. Cel pracy

Pierwszym celem mojej pracy jest wykazanie, że po skodyfikowaniu reguł planowania syntezy, nie ma różnicy czy reguły te stosuje człowiek czy komputer.

Drugim celem mojej pracy jest wykazanie wykonalności syntetycznej planów wygenerowanych przez program komputerowy Chematica/SynthiaTM, oraz porównanie tych ścieżek syntezy z istniejącymi i stworzonymi przez człowieka dla tych samych związków. Pragnę w tym miejscu położyć nacisk na walidację eksperymentalną komputerowo projektowanych planów syntetycznych. Dotychczas żaden program komputerowy nie był w stanie stworzyć nowatorskich planów syntetycznych, które następnie wykonane w laboratorium okazałyby się porównywalne lub lepsze od tych stworzonych przez człowieka. W niniejszej pracy postaram się ustalić przyczyny takiego stanu rzeczy.

2. Wstęp literaturowy

2.1. Historia programów komputerowych wspomagających planowanie wieloetapowych syntez chemicznych

2.1.1. Wstęp

Proces projektowania syntezy chemicznej jest czynnością wyjątkowo złożoną. Wynika to z wielości parametrów koniecznych do uwzględnienia przy każdym etapie syntezy, oraz miejscem każdego etapu syntezy w całości planu syntetycznego. Drobny błąd na choćby jednym etapie skutkować może niewykonalnością całego planu. Ważnymi czynnikami wymagającymi uwzględnienia przy planowaniu syntezy są:

- 1) występowanie w celu syntetycznym podjednostek pochodzących z łatwo dostępnych bloków budulcowych,
- 2) identyfikacja wiązań chemicznych prowadzących do największego topologicznego uproszczenia struktury celu syntetycznego,
- 3) możliwość/konieczność wprowadzenia i usunięcia grup ochronnych na poszczególnych etapach,
- 4) występowanie konfliktów reaktywności pomiędzy różnymi grupami funkcyjnymi w cząsteczce,
- 5) występowanie konfliktów reaktywności pomiędzy stosowanymi reagentami, a grupami funkcyjnymi w cząsteczce,
- 6) identyfikacja występujących w celu syntetycznym elementów stereogenicznych, określenie sposobów i kolejności ich wprowadzenia do cząsteczki, oraz wzajemnego oddziaływania tych elementów stereogenicznych na siebie i wynikających z tego różnic reaktywności,
- 7) określenie kolejności wykonywanych reakcji chemicznych w planie syntetycznym,
- 8) określenie typu planu syntetycznego (liniowy lub zbieżny),
- 9) określenie łatwości znalezienia alternatywnych planów w razie niepowodzenia pojedynczego kroku syntetycznego,
- 10) identyfikacja występujących w planie syntetycznym związków nietrwałych lub o właściwościach utrudniających wykonanie i skalowanie syntezy.

Stworzenie chemicznie poprawnego planu syntetycznego wymaga starannego uwzględnienia każdego z powyższych czynników. Co więcej, dobry plan syntetyczny, stosując regułę brzytwy Ockhama, powinien zawierać tylko tyle etapów syntezy i oczyszczania ile istotnie jest koniecznych do uzyskania celu syntetycznego. Realia rynkowe dodają do powyższej listy stosowanie jak najtańszych dostępnych na rynku reagentów realizujących reakcje i stosowanie jak najtańszych bloków budulcowych. Wymogi zielonej chemii zwracają uwagę na kwestię jak najwyższej ekonomii atomowej planu syntetycznego, jak i możliwie najmniejszej ilości produkowanych odpadów. Chemik-projektant powinien mieć na uwadze wszystkie powyższe kryteria.

Czy komputer również jest w stanie uwzględniać wszystkie lub choćby większość z tych czynników i generować użyteczne plany syntez, stanowiące realną pomoc i usprawnienie pracy chemika? Pytanie to ma już ponad półwiekową tradycję. Odkąd komputery stały się powszechnym narzędziem pracy próbowano wykorzystać je również do zadań wykonywanych dotychczas przez chemików. Najwybitniejsi chemicy swoich czasów próbowali skutecznie „nauczyć” komputery

chemii organicznej. Cały szereg hucznie zapowiadanych programów był konsekwentnie negatywnie weryfikowany przez rzeczywistość. Skutecznie zmniejszało to motywację następnych śmiałków, utwierdzając opinię o wyjątkowości chemii organicznej i nieujmowalności jej złożoności dla niczego poza ludzkim umysłem.

Tworzone programy opierały się na różnych zasadach działania, tak więc w dziedzinę *komputerowo-wspomaganej syntezy* wpisuje się szereg programów o różnych funkcjach i zasadach działania. Funkcją wykonywaną przez taki program może być:

- 1) Analiza retrosyntetyczna,
- 2) Odkrywanie nowych reakcji chemicznych,
- 3) Przewidywanie produktu reakcji w oparciu o substraty i warunki reakcji,
- 4) Przewidywanie mechanizmu reakcji,
- 5) Proponowanie najlepszego związku wyjściowego dla syntezy,
- 6) Tworzenie planów syntetycznych.

Najczęściej programy do komputerowo-wspomaganej syntezy, łączyły w swym działaniu kilka z wyżej wymienionych funkcji. W niniejszej pracy interesują mnie w największym stopniu programy, które w założeniu, po wprowadzeniu przez użytkownika struktury celu syntetycznego, generowały gotowy plan syntetyczny lub jego elementy. Jeżeli program nie wykonywał tej funkcji, nie mieści się on wówczas w ramach tematycznych niniejszego omówienia.

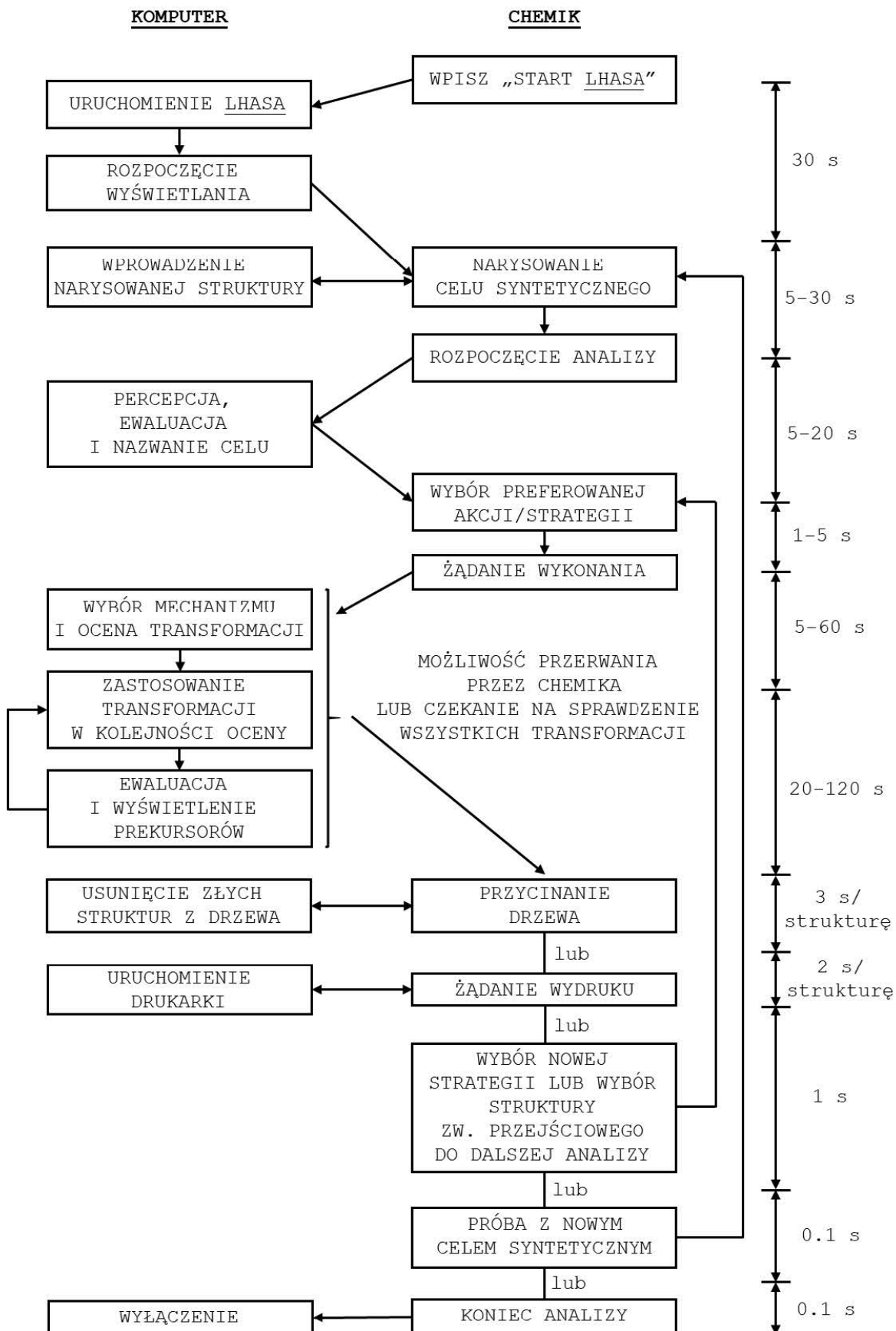
Poniżej przedstawię zarys historii programów komputerowych służących do wspomaganie syntezy. Szczególnym nacisk kłaść będę na rezultaty ich działania: walidację poprawności chemicznej i użyteczności prezentowanych programów oraz ocenę czy kiedykolwiek stanowiły one realną pomoc w pracy chemika-projektanta.

2.1.2. OCSS & LHASA

2.1.2.1. Założenia i ogólna zasada działania

Pierwszym programem służącym do planowania syntezy chemicznej był OCSS (Organic Chemical Simulation of Synthesis), tworzony od końca lat 60. XX w. na uniwersytecie Harvarda w zespole E.J. Corey'a. Projekt badawczy, który na początku lat 70. zmienił nazwę na LHASA (Logic and Heuristics Applied to Synthetic Analysis) był najdłużej rozwijanym programem wspomagającym syntezę organiczną. Przez niemal 30 lat (1969-1997) ukazało się blisko 30 publikacji opisujących postępy projektu i dodawane doń kolejne funkcjonalności. Tak bogaty materiał źródłowy pozwala precyzyjnie i chronologicznie śledzić rozwój projektu. Jednak ze względu na przejrzystość niniejszej rozprawy, w opisie skupię się na głównych założeniach programu LHASA i korzyściach uzyskanych z całości projektu.

Zaproponowane przez E.J. Corey'a w 1967 roku reguły retrosyntezy¹, stanowiące odpowiedź na pytanie: „Jak chemik wybiera ścieżkę syntezy spośród wielu możliwych?” stanowiły zarazem wymóg i warunek wstępny do stworzenia programu wspomagającego planowanie syntezy. Wymóg – ponieważ umiejętność planowania syntez w intuicyjnej i nieskodyfikowanej postaci, którą dysponowali ówczesni chemicy, nie nadawała się do wykorzystania w programie komputerowym. Warunek wstępny – gdyż chęć stworzenia programu wymusiła kodyfikację tych reguł. Oprócz reguł i wytypowania strategii syntezy, pomysł na program opierał się na wykorzystaniu eksperckiej bazy danych. Zaimplementowanie tych dwóch elementów, stanowiło rdzeń programu OCSS/LHASA. Pozostałe, rozwijane przez lata, funkcjonalności programu (tj. kolejne strategie dalekiego zasięgu, informacje o grupach ochronnych, percepcja pierścieni itd.), stanowiły wyłącznie dodatek do



Ryc.1. Schemat pracy z użyciem programu LHASA, z przypisaniem czynności/operacji wykonywanych przez program i przez użytkownika².

przyjętej jeszcze w latach 60. zasady działania.

W celu wprowadzenia do programu struktury celu syntetycznego posłużono się skonstruowanym ówczynie tabletem dotykowym, pozwalającym na narysowanie całej struktury cząsteczki i utworzenie przez program na jej podstawie tabeli wiązań chemicznych. Podkreślono³, że wystarczy 3 minutowe przeszkolenie by użytkownik, posługując się tylko jedną ręką, był w stanie wprowadzić za pomocą tabletu strukturę kortyzonu w czasie poniżej 30 sekund. W pierwszej wersji programu, nałożono jednak ograniczenie na wielkość struktury, która nie mogła zawierać więcej niż 36 atomów.

Po wprowadzeniu do programu struktury celu syntetycznego, należało wybrać jedną ze strategii retrosyntetycznych. Pozostawienie tego wyboru ludzkiej ocenie, stanowiło zakamuflowany sposób uniknięcia lub spowolnienia eksplozji kombinatorycznej przez ograniczenie już na wstępie przeszukiwanej przestrzeni transformacji. Następnie program, w oparciu o stworzoną na jego potrzeby ekspercką bazę danych o reakcjach chemicznych, w ciągu 20-120 sekund generował, a następnie przedstawiał pierwszą generację syntonów. Była ona automatycznie pomniejszana o struktury uznane przez program za naiwne, nieużyteczne i nierokujące. Użytkownik musiał ową pulę manualnie przeanalizować i wytypować warianty mogące prowadzić do wartościowych planów retrosyntetycznych. Ocenivszy pierwszą generację syntonów, użytkownik mógł albo wybrać jeden z nich jako kolejny punkt wyjściowy do dalszego rozwoju planu retrosyntetycznego lub wybrać inną strategię w nadziei wygenerowania przez nią atrakcyjniejszych syntonów.

To co zwraca uwagę w interakcji człowieka z programem LHASA, to pozostawienie człowiekowi dużej odpowiedzialności za rozwój ścieżki. Według autorów² (Ryc.1.) przeanalizowanie struktury każdego syntonu i ocena jego użyteczności zajmie użytkownikowi 3 sekundy na strukturę. Rola programu LHASA ogranicza się zatem wyłącznie do generowania chemicznie prawdopodobnych, kolejnych zestawów syntonów do oceny.

Zakładając, że umiarkowanej złożoności związek organiczny można uzyskać przeciętnie za pomocą 100 różnych reakcji chemicznych, to gdy do jego uzyskania potrzebujemy choćby pięciu etapów, ilość koniecznych do przeanalizowania planów retrosyntetycznych wynosi 10 miliardów. Manualne ich przeanalizowanie, wymagałoby od użytkownika ich wyświetlenia, następnie obejrzenia i zapamiętania, a dopiero wówczas porównania podług szeregu kryteriów wymienionych wcześniej. Zakładając, że ktokolwiek mógłby tego dokonać i idąc za Corey'em, poświęciłby 3 sekundy na strukturę, to samo ich obejrzenie zajęłoby ponad 950 lat. A mówimy tu o cząsteczce umiarkowanej złożoności.

Problem eksplozji kombinatorycznej planów retrosyntetycznych wykazuje również tak zwany *self-inflicted problem*. Mianowicie, im lepsze chemicznie plany syntetyczne chcemy generować, tym więcej reguł reakcji chemicznych musi posiadać nasza baza danych. Innymi słowy program musi posługiwać się bardziej wszechstronną chemicznie paletą transformacji. Niestety większa ilość transformacji, generować będzie więcej możliwych syntonów w każdym pokoleniu, co przyspiesza eksplozję kombinatoryczną. Chcąc ją opóźnić możemy albo posługiwać się uboższą bazą danych albo okroić zakres jej stosowalności, ze skutkiem w postaci mniejszej różnorodności chemicznych reakcji – tą drogą poszedł Corey^{4,5}.

2.1.2.2. Strategie, kombinacje taktyczne i ekspercka baza danych

Corey przyznał, że LHASA nie potrafiła *ad hoc*, w oparciu o strukturę celu syntetycznego, określić jaka jest optymalna dla niego strategia³. Jedynym kryterium wyboru transformacji do planu retrosyntetycznego jest wybór tych zapewniających jak najszybsze upraszczanie struktury w najbliższym kroku³.

Sercem programu LHASA była ekspercka baza danych – to dzięki zapisanym w niej transformacjom, możliwe było tworzenie jakichkolwiek ich sekwencji dla poszukiwanych celów syntetycznych. Nie wszystkie transformacje, traktowano jednak w ten sam sposób. Niektóre z nich zostały wyróżnione (na przykład reakcja Dielsa-Aldera czy anulacja Robinsona), a część pozostałych zestawiono w kombinacje taktyczne. Wszystko to miało na celu spowolnienie eksplozji kombinatorycznej. Skutkiem tego zróżnicowania było tworzenie tendencyjnych drzew retrosyntetycznych, których konstrukcja zależała od ocenianych przez chemika cech strukturalnych celu syntetycznego.

Tym sposobem wyróżniono 6 strategii analizy retrosyntetycznej⁶:

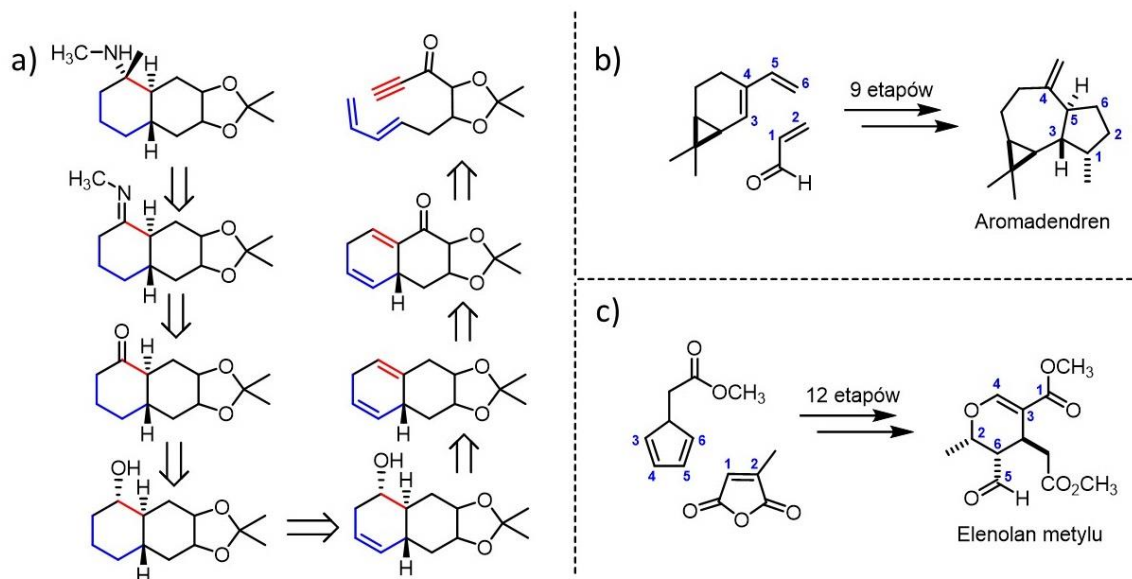
- 1) Strategie krótkiego zasięgu,
- 2) Strategie topologiczne,
- 3) Strategie dalekiego zasięgu,
- 4) Strategie stereochemiczne,
- 5) Strategie ukierunkowane na substrat,
- 6) Strategia sugerowana przez LHASA.

Nie wszystkie z tych strategii cechowała równa przydatność. Nie wszystkie też zostały w równym stopniu rozwinięte.

(1) *Strategie krótkiego zasięgu* dotyczyły głównie prostych struktur, gdzie i bez użycia LHASA, chemik z łatwością zaproponowałby krótką sekwencję transformacji przekształcających głównie grupy funkcyjne.

(2) *Strategie topologiczne* miały na celu zidentyfikowanie w strukturze docelowej wiązań, których utworzenie (rozłączenie w kierunku retrosyntetycznym) prowadziłyby uproszczenia struktury w większym stopniu niż innych wiązań. Prace nad tym modulem ze względu na jego uniwersalną aplikowalność w stosunku do różnych struktur podjęto⁴, jako jedno z pierwszych zadań projektu LHASA, już w latach 1970-1974. Wyróżniono cztery kryteria oceniające uproszczenie struktury w transformacji: 1) minimalizacja liczby podstawników i/lub grup bocznych; 2) minimalizacja liczby podstawników i/lub grup bocznych zawierających centra stereogeniczne; 3) minimalizacja liczby pierścieni; 4) minimalizacja liczby atomów mostkowych dla kilku pierścieni. Reguły te ograniczały się w zasadzie do związków karbocyklicznych, bowiem jako wiązanie strategiczne mogło być wytypowane wyłącznie wiązanie węgiel-węgiel, natomiast wiązania węgiel-heteroatom nie były brane pod uwagę. Analiza polegała na przeszukiwaniu drzewa retrosyntetycznego do głębokości 4 transformacji przy użyciu prostych przekształceń zmieniających i dodających grupy funkcyjne. Jeżeli ta stosunkowo płytka analiza nie znajdowała transformacji rozłączającej wytypowane wiązanie to program zarzucał tę analizę i przystępował do następnego wiązania z listy. Moduł ten został poddany walidacji za pomocą testu, w którym wytypowano 14 literaturowych syntez związków mostkowo-policyklicznych – w 10 przypadkach, znane z literatury syntezy, zawierały tworzenie wiązania wytypowanego przez LHASA jako strategiczne⁴.

(3) *Strategie dalekiego zasięgu* (określane też jako „powerful transformations”), stanowiły najbardziej zaawansowany moduł LHASA, generujący najdłuższe analizy (Ryc.2a.). Jego działanie polegało na autonomicznym generowaniu długich planów syntetycznych, bazujących na wytypowaniu określonej transformacji, a następnie głębokiej analizie retrosyntetycznej do momentu pojawienia się możliwości zastosowania tej transformacji. Eksponowana jako zaleta, długość tych planów budzi jednak wątpliwości. W abstrakcie pracy z 1974 r., o pierwszej strategii dalekiego zasięgu, napisanej dla reakcji Dielsa-Aldera, jest mowa o automatycznym przeszukiwaniu drzewa retrosyntetycznego do głębokości 15 etapów, w poszukiwaniu możliwości zastosowania tej reakcji¹⁰.



Ryc.2. Implementacja reakcji Dielsa-Aldera w programie LHASA. a) Przykład rezultatu wykorzystania strategii dalekiego zasięgu dla reakcji Dielsa-Aldera. Ograniczenie stosowania strategii dalekiego zasięgu dla reakcji Dielsa-Aldera, polegało na tym, że mogła być ona zastosowana wyłącznie wówczas, gdy LHASA rozpoznała w strukturze celu syntetycznego, karbocykliczny pierścień sześciocłonowy. Nie stosowała się do sytuacji, gdy w środku długiej syntezy wystąpiła reakcja Dielsa-Aldera, a następnie utworzony w niej pierścień sześciocłonowy ulegał dalszym reakcjom, tak, że nie był już obecny w finalnej strukturze. b) Ilustracja ograniczeń stosowalności reakcji Dielsa-Aldera w programie LHASA na przykładzie syntezy totalnej aromadendrenu^{7,8}; b) Ilustracja ograniczeń aplikacji reakcji Dielsa-Aldera w programie LHASA na przykładzie syntezy totalnej elenolanu metylu⁹. Dla przypadków b) i c) LHASA nie zastosowałaby reakcji Dielsa-Aldera.

Wartość ta została powtórzona w pracy przeglądowej Todd'a¹¹ z 2004. Co jednak faktycznie oznaczała wartość 15 etapów? Autorzy¹⁰ przyznają, że najbardziej czasochłonnym etapem pracy przy użyciu tej strategii, jest sprawdzenie przez chemika wszystkich zaproponowanych związków przejściowych i wybór następných do stworzenia kolejnej generacji. Automatycznym można tu nazwać wyłącznie generowanie każdej generacji z osobna, natomiast połączenie między nimi, ocena ekspercka i wybór kandydatów do dalszej analizy – pozostawały niezautomatyzowane.

Szczytem możliwości automatycznego generowania sekwencji był moduł z 1976 r., przeszukujący drzewo retrosyntetyczne do 4 etapów w głąb przy użyciu transformacji zmieniających grupy funkcyjne¹². Strategia ukierunkowana na znalezienie reakcji Dielsa-Aldera, cierpiała na poważne ograniczenie. By móc ją zastosować, w strukturze celu syntetycznego musiał się znajdować karbocykliczny pierścień sześciocłonowy – reakcja hetero-Dielsa-Aldera nie była rozpatrywana¹⁰. Zidentyfikowanie tego motywu w strukturze stanowiło podstawę do zaproponowania 6 teoretycznych par struktur dienu i dienofila, mogących stworzyć ów pierścień na 6 różnych sposobów. Zazwyczaj jednak możliwe było zidentyfikowanie od 0 do 3 takich par struktur i dla nich wykonywano analizę retrosyntetyczną. Konieczność zachowania pierścienia sześciocłonowego w produkcie końcowym, nakłada poważne ograniczenia na tę strategię. W czasie gdy implementowano reakcję Dielsa-Aldera do LHASA, publikowane były syntezy totalne (Ryc.2b-c), których LHASA z założenia nie byłaby w stanie odtworzyć.

W 1980 roku do strategii dalekiego zasięgu dodano anulację Robinsona¹³ i halolaktonizację¹⁴, jednak oprócz wygenerowania długich ścieżek, na podobnej zasadzie jak dla strategii dalekiego zasięgu Dielsa-Aldera, i tutaj nie wykonano ich walidacji. Ostatnią strategią dalekiego zasięgu, dla wariantu reakcji Dielsa-Aldera z chinonami¹⁵, dodano w 1997 roku, była to też ostatnia opublikowana praca o wspomaganiu retrosyntezy przy użyciu LHASA. Program

w założeniu miał posiadać szereg *strategii dalekiego zasięgu*, tak by choćby dla pierścienia sześciocłonowego móc samodzielnie dobrać najbardziej adekwatną¹⁰, faktycznie jednak LHASA nigdy nie osiągnęła tej zdolności.

(4) *Strategie stereochemiczne* miały być stosowane do celów syntetycznych posiadających centra stereogeniczne. W 1974 roku pojawiła się pierwsza praca¹⁰ zawierająca wzmiankę o rozpoznawaniu przez LHASA centrów stereogenicznych, konfiguracji E/Z wiązania podwójnego czy konfiguracji cis/trans. Jedyny przykład planu syntetycznego z LHASA wykorzystujący indukcję asymetryczną pojawił¹⁶ się w pracy z 1985 roku. Na dodanie do programu pełnych reguł Cahn-Ingolda-Preloga trzeba było czekać aż do 1994 roku, ale nawet wówczas nie podjęto prób walidacji modułu¹⁷.

(5) *Strategie ukierunkowane na substrat*^{6,16,18} rozwijany był w późnym okresie rozwoju LHASA. Oprócz prezentacji planu syntetycznego dla podstawionego pierścienia benzenu moduł ten nigdy nie był walidowany.

(6) *Strategie sugerowane przez LHASA* miały w założeniu dokonywać autonomicznej analizy struktury celu syntetycznego, a następnie w oparciu o nią wybrać z bazy danych skutecznych transformacji najskuteczniejszą dla danego przypadku. Ze względu na brak implementacji założonej listy skutecznych transformacji^{10,19}, moduł ten nie został zrealizowany w kształcie pierwotnie założonym. Działanie tej strategii ograniczało się więc do następującej procedury: w przypadku niewybrania przez użytkownika żadnej ze strategii 1-5, program próbował wszystkich posiadanych strategii jedna po drugiej do momentu gdy któraś z nich nie wygenerowała nowego związku przejściowego².

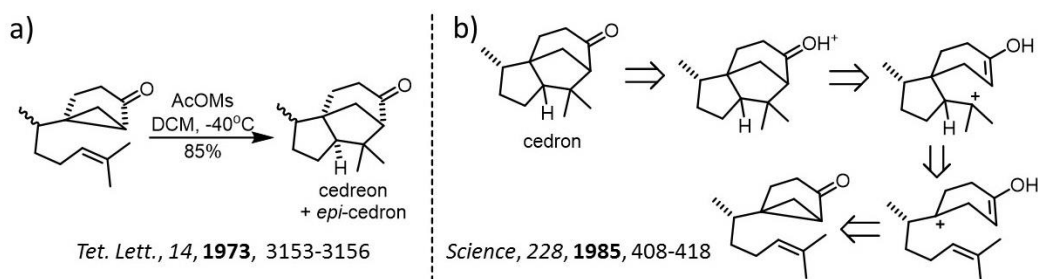
Oprócz 6 głównych strategii, istniał jeszcze szereg pomniejszych strategii, które tu omówię. Jednym ze sposobów na ograniczenie eksplozji kombinatorycznej⁵, było potraktowanie określonej sekwencji reakcji jako jednej, usztywnionej przemiany. W ten sposób, w ostatnich latach projektu, wprowadzono *kombinacje taktyczne*. Ich liczba w zależności od źródła oscyluje poniżej 500 (450⁵, 480²⁰, 495²¹, 500²²). Miały one na celu głównie tworzenie pierścieni, gdyż strategii Dielsa-Aldera dalekiego zasięgu nie dało się stosować do przypadków innych niż pierścien sześciocłonowy, a LHASA cierpiała na niedobór różnych metod adekwatnych do zamykania pierścieni. Zaproponowane⁵ przez tą strategię ścieżki, nie występowały wcześniej w literaturze, ani nie wykonano ich w laboratorium, dlatego nie można ich traktować jako przykładu walidacji.

Oprócz dotychczas opisanych strategii, w trakcie prac nad LHASA, pojawiła się propozycja strategii bazującej na odłączaniu łańcuchów bocznych¹⁹. Celem było uzyskanie pozbawionego podstawników pierścienia, zdefiniowanego jako rdzeń, do którego łatwiej wówczas można by zastosować którąś z transformacji dalekiego zasięgu. Oprócz opisu teoretycznego¹⁹, nie przedstawiono walidacji tej strategii i nie wspomina się już o niej po roku 1976.

Podobnie należy potraktować *strategie wyższego rzędu*²³ zaproponowane w 1976 roku. W razie napotkania przez program grupy funkcyjnej niemożliwej do ochrony, miały one dokonywać zmiany kolejności kroków syntetycznych w planie w celu uniknięcia konfliktów reaktywności. Ponowna wzmianka²⁴ o tych strategiach, pojawia się w 1985 roku, ponownie, jedynie jako propozycja dodania tego modułu w przyszłości. W późniejszych publikacjach brak jest informacji o jej dodaniu.

Na osobne potraktowanie zasługuje generowanie niektórych syntonów w postaci karbokationów, określone jako *transformacje mechanistyczne*¹⁶. Ta funkcjonalność pojawiła się już w pierwszej pracy³. Jej działanie polegało na wytworzeniu karbokationu (na przykład na skutek

rozcięcia wiązania węgiel-halogen), a następnie generowaniu szeregu produktów przegrupowania karbokationu, aż do osiągnięcia stabilnego związku przejściowego¹⁶. Opieranie analizy na przegrupowaniach karbokationów, wydaje się ryzykownym pomysłem. Ryzyko to wynika z małej możliwości kontroli tego procesu, jego dużej zależności od struktury i uzyskiwania mieszanin produktów. Jakie więc zalety przemawiały za pozostawieniem w programie tej drogi prowadzenia reakcji? Moim zdaniem, decyzja o umieszczeniu takiego modułu była obarczona dużym ryzykiem generowania kroków syntetycznych niemożliwych do realizacji. Z drugiej jednak strony był to jedyny prawdziwie kreatywny moduł, mogący prowadzić do przegrupowań nigdy wcześniej nie obserwowanych, chociaż teoretycznie możliwych. W 1973 roku Corey opublikował²⁵ syntezę totalną szeregu związków o rdzeniu cedrenowym, wykorzystujące atrakcyjne przegrupowanie karbokationu (Ryc.3a) budujące cały rdzeń tej klasy związków w jednej reakcji. Przykład zawierający budowanie tego szkieletu na drodze przegrupowania karbokationu (Ryc.3b) znalazł się¹⁶ też w pracy z 1985 roku przy opisie *transformacji mechanistycznych*. Zatem przyczyna zachowania tego obciążonego ryzykiem modułu mogła wynikać z założenia by program był co najmniej tak dobry jak jego twórcy²⁶ i by był w stanie odtworzyć *in silico* opracowane wcześniej przez nich spektakularne reakcje.



Ryc.3. Przykład tłumaczący wprowadzenie do LHASA modułu *transformacji mechanistycznych*, pozwalających na uzyskanie cedronu w wyniku przegrupowania karbokationu.

O ile LHASA została napisana w językach Fortran i C¹⁵, to ekspercką bazę reakcji stworzono w języku CHMTRN¹⁶. W tym sposobie zapisu, balans pomiędzy zrozumiałością dla chemika, a zrozumiałością dla procesora przesunięty był znacznie w pierwszą stronę. Dzięki temu, chemicy bez doświadczenia w programowaniu mogli w stosunkowo łatwo sposób przekształcać swoją wiedzę chemiczną na postać czytelną dla programu. Ponadto od 1971 roku używano programu TBLTRN¹⁴ tłumaczącego format CHMTRN na postać jeszcze bardziej skróconą i oszczędniejszą dla programu. Z drugiej jednak strony, stanowiło to ograniczenie dla możliwości inkorporacji bardziej złożonych transformacji chemicznych. Przykład jednej z transformacji z bazy danych znajduje się na Ryc.4.

W tak zapisanej regule reakcji, możemy dokładnie sprawdzić jakie grupy elektroakceptorowe autorzy wymienili jako umożliwiające reakcję. Brak jest na tej liście grupy sulfonowej, kwasu sulfonowego, grupy nitrowej, kwasu fosforanowego i jego estrów, tioestrów oraz amidów. Dla tych grup reakcja Michaela, choć chemicznie możliwa, nie zostałaby zaproponowana. Powód mógł być jednak inny. Moduł rozpoznający grupy funkcyjne w LHASA grupy funkcyjne, nie był w stanie rozpoznać grupy sulfonowej czy fosforanowej²⁸. Wynikało to z założenia, że grupa funkcyjna to atom węgla lub azotu połączony z heteroatomem. Jednak już inny heteroatom jak siarka czy fosfor, łączący się dalej z heteroatomem, nie był rozpoznawany. Stąd brak tych grup wśród słów znaczących w zapisie CHMTRN. Brak również w regule reakcji, grupy nitrowej, która w odróżnieniu od poprzednio wymienionych, była przez zapis CHMTRN rozpoznawana.

Ocena reakcji mogła się mieścić w przedziale 0-100, po zastosowaniu reakcji do konkretnego przypadku, ocena ulegała zmianie na skutek cech struktury warunkujących na jej aplikowalność.

```

TRANSFORM 117
NAME MICHAEL ADDITION OF HETERO NUCLEOPHILE
...HET-C-C-W => HET-H + C=C-W
...MARCH 585; HOUSE 596; B+P 468
...ORG. RXNS. VOL.5, 79-135 (1949)
...BULL. SOC. CHEM. FR. 254,325 (1962)
...PATH 2 BONDS
RATING 50          ...Old rating 40
GROUP*1 MUST BE KETONE OR CYANO OR ESTER OR ACID
OR LACTONE OR AMIDE*3 OR AMIDE*2 OR AMIDE*1
OR LACTAM OR VINYLW OR ALDEHYDE
GROUP*2 MUST BE ETHER OR AMINE*1 OR AMINE*2 OR AMINE*3
OR SULFIDE OR THIOL
STUDENT
REMOVES*STEREO CARBON2*1 ATOM*2
BROKEN*BONDS BOND2*1
...
KILL IF NO HYDROGEN ON ATOM*2
...REQUIRED FOR REACTION
KILL IF MULTIPLE BOND ON ATOM*2 OFFPATH OR: ON ATOM*3 OFFPATH
...WOULD PRODUCE ALLENIC PRECURSOR
IF BOND2*1 IS NOT IN A RING OF SIZE 5 THROUGH 7 &
THEN KILL IF BOND2*1 IS IN A RING
SUBTRACT 15 IF LEAVING GROUP ON ATOM*3
...POSSIBLE ELIMINATION
ADD 15 IF ANOTHER WITHDRAWING BOND ON ATOM*2
...EASIER ADDITION
SUBTRACT 15 FOR EACH WITHDRAWING BOND ON ATOM*3
...UNDESIRE MICHAEL POSSIBLE
SUBTRACT 10 IF ATOM*3 IS A TERTIARY*CENTER
IF NOT OLEFIN ON BOND*2 THEN KILL IF ATOM*2 IS NOT ENOLIZABLE
...STABLE ENOL PROVIDES DRIVING FORCE
IF SECOND GROUP IS ETHER THEN CONDITIONS NaOR
IF SECOND GROUP IS AMINE THEN CONDITIONS RNH2
IS SECOND GROUP IS SULFIDE OR: THIOL THEN &
CONDITIONS NaSR
...
BREAK BOND2*1
JOIN ATOM*2 AND ATOM*3
...

```

Ryc.4. Przykład reguły reakcji: addycja Michaela dla nukleofili tlenowych, azotowych i siarkowych, zapisanej w eksperckiej bazie LHASA.

Jeżeli zawierała się w przedziale od 0 do -50 wymagała ewaluacji chemika, reakcje z oceną poniżej progu odcięcia -50 nie były brane pod uwagę^{20,27}. Na podkreślenie zasługuje fakt, że oceny zostały przypisane reakcjom w sposób arbitralny, podobnie wartości kar i nagród dla reakcji za obecność elementów strukturalnych. Skutkiem tego mogło być niezbilansowanie programu i stronnicze faworyzowanie jednych reakcji kosztem innych w sposób zupełnie przypadkowy. Problem tendencyjności wyników, wynikającej z niezbilansowanych ocen, nigdy nie został poruszony w pracach o LHASA.

Wielokrotne wzmianki, pozwalają prześledzić również rozwój i przyrost bazy danych programu. W 1970 roku, LHASA zawierała²⁷ 250 reguł reakcji, 4 lata później ukazały się prace wskazujące na przyrost do co najmniej 270¹⁰ i następnie 280⁴. Po 16 latach trwania projektu, w 1985 roku, baza zawierała już 1100 reguł reakcji¹⁶. By w 1994 roku osiągnąć 2100²². Informacja na stronie internetowej Uniwersytetu w Nijmegen²¹ podaje w 2019 roku, że obecna wersja to LHASA 20.3, posiadająca 2271 reguł reakcji chemicznych w bazie.

2.1.2.3. Percepcja struktury i kontekstu strukturalnego

Program komputerowy mający operować cząsteczkami organicznymi musi mieć zdolność pozyskiwania informacji o właściwościach cząsteczki z jej struktury. Grupy funkcyjne stoją w centrum zainteresowania chemika, gdyż w przeciwieństwie do szkieletu węglowodorowego są dużo bardziej reaktywne. Dlatego niezdolność rozpoznawania różnorodnych grup funkcyjnych stanowiła poważne ograniczenie LHASA²⁸. Podobnie percepcja elementów stereogenicznych, która

bardzo długo nie została wprowadzona do programu w pełnej postaci^{10,16,17}. Wersja programu z 1972 roku²⁷ nie zawierała jeszcze chemii związków aromatycznych, dopiero w pracy z 1992 roku¹⁸ znajdujemy pierwszy i jedyny przykład z ich wykorzystaniem.

Percepcja zatłoczenia sterycznego i wykluczania z planu związków o zbyt dużym zatłoczeniu została poruszona²⁷ w pracy z 1972 roku. Oprócz możliwości obniżenia oceny reakcji (o ile każdorazowo taki kwalifikator został w reakcję wpisany), LHASA nie dysponowała żadną heurystyką pozwalającą na automatyczne odsiewanie zbyt zatłoczonych struktur. Jeszcze gorzej rzecz się miała ze związkami naprężonymi, gdyż z poziomu reguły reakcji nie można było przewidzieć i wykluczyć wygenerowania tego typu niemożliwych połączeń²⁸.

W pracach nad LHASA, dużo uwagi poświęcono pierścieniom karbocyklicznym^{3,29,30}, służyło to głównie usprawnieniu działania strategii dalekiego zasięgu Dielsa-Aldera. Pierścienie w cząsteczce podzielono też na *realne* i *pseudo-pierścienie*^{3,28}. Miało to ograniczyć ilość wyników, poprzez stosowanie reguł reakcji wyłącznie do pierścieni *realnych*. Jednak z powodu zbyt małej ilości metod tworzenia pierścieni pełen potencjał rozwiązań szacujących energię konformerów³⁰, czy korelujących konformację pierścienia z reaktywnością²⁹, nie został w pełni wykorzystany.

LHASA miała dostęp do dwóch baz danych o niekompatybilnościach warunków i grup²⁴. Pierwszą tabelę reaktywności 112 grup funkcyjnych wobec 60 warunków reakcji, opublikowano w 1976 roku²³, a 9 lat później ogłoszono jej wersję rozszerzoną do 138 warunków reakcji²⁴. Drugą bazą danych, obsługiwaną przez PROTECT, była tabela reaktywności 228 grup funkcyjnych, wobec 108 prototypowych warunków reakcji. Nietrudno zauważyć, że obie tabele powielają częściowo te same informacje, brak jednak informacji o próbie określenia wzajemnej relacji między nimi, czy o próbie ich unifikacji.

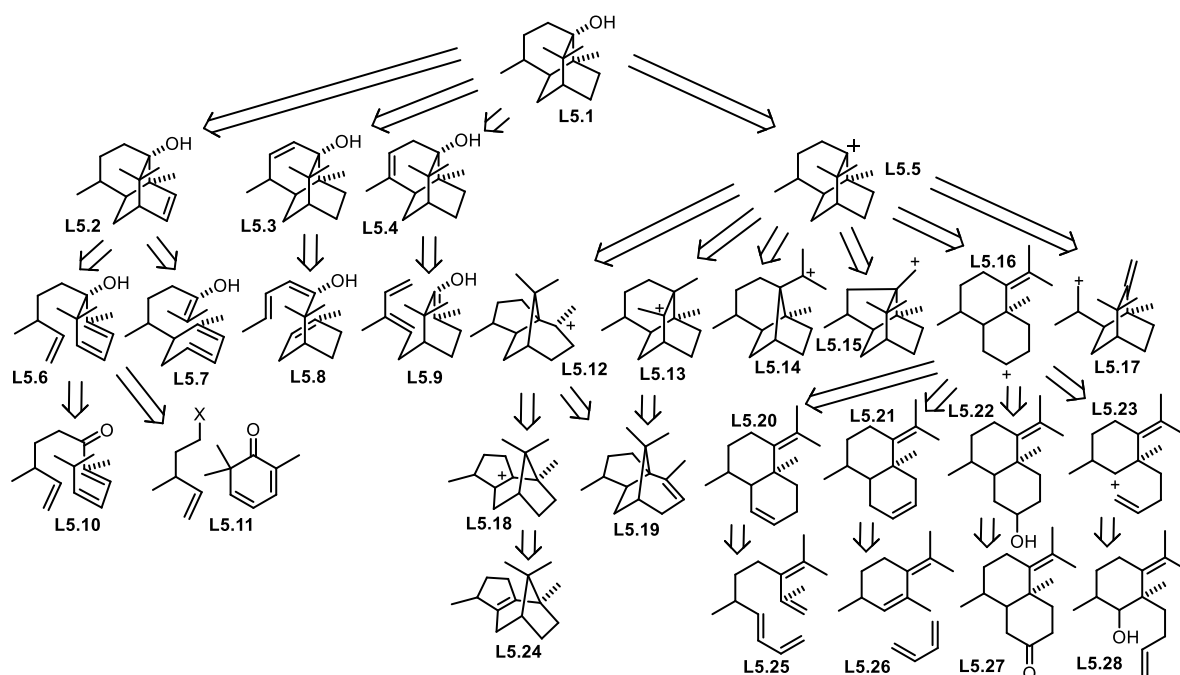
2.1.2.4. Rezultaty i ich ocena

Skala i ambicje projektu LHASA były imponujące, jednak nie powinno to przesłonić trzeźwej oceny, bezpośrednich rezultatów projektu. Spotkałem się z opinią, że skoro przez tyle lat chemik tak wybitny jak Corey pracował nad rozwojem programu komputerowego, to dla *jakiejś* cząsteczki musiał dokonać walidacji programu poprzez syntezę wcześniej wygenerowanego komputerowo planu. By obalić ten mit, poniżej przytaczam wszystkie związki, dla których w latach 1969-1997 opublikowano wygenerowaną przez LHASA analizy retrosyntetyczne. Najważniejsze z nich omawiam w poniższym rozdziale. Rozszerzone komentarze dla tu opisanych oraz omówienie mniej znaczących analiz retrosyntetycznych znajdują się w **Appendiks II**. Nie uwzględniam oczywiście trywialnych kilkukrokowych analiz, będących ilustracjami pracy któregoś z modułów LHASA – zresztą o ich syntetycznej weryfikacji Corey nie wspomina.

(I) Alkohol paczulowy (Ryc.5.) – W pierwszej pracy o LHASA³ z 1969 roku, na jego przykładzie pokazane zostało drzewo retrosyntetyczne o kilku generacjach. Oczywiście ze względu na skąpą wówczas bazę danych o regułach reakcji, przekształcenia sprowadzają się do: (I) różnych wariantów reakcji Dielsa-Aldera, aplikowanej dla każdego sześcioczłonowego pierścienia karbocyklicznego bez względu na jego właściwości elektronowe (reakcje: **L5.6**→**L5.2**; **L5.9**→**L5.4**; **L5.25**→**L5.20** etc.), oraz (II) do ryzykownych przegrupowań karbokationów (reakcje: **L5.12**→**L5.5**; **L5.15**→**L5.5** etc.).

Przedstawione drzewo retrosyntetyczne trudno traktować jako plan syntetyczny. Przedstawiono tu 14 różnych propozycji planów retrosyntetycznych, a oprócz związków nawiązujących do wcześniej już znanych syntez totalnych alkoholu paczulowego, pozostałe propozycje retrosyntetyczne, nie zostały wykorzystane w żadnej dotychczasowej syntezie. W pracy³ Corey zwraca uwagę, że drzewo retrosyntetyczne zawiera związek (**L5.24**) występujący wcześniej

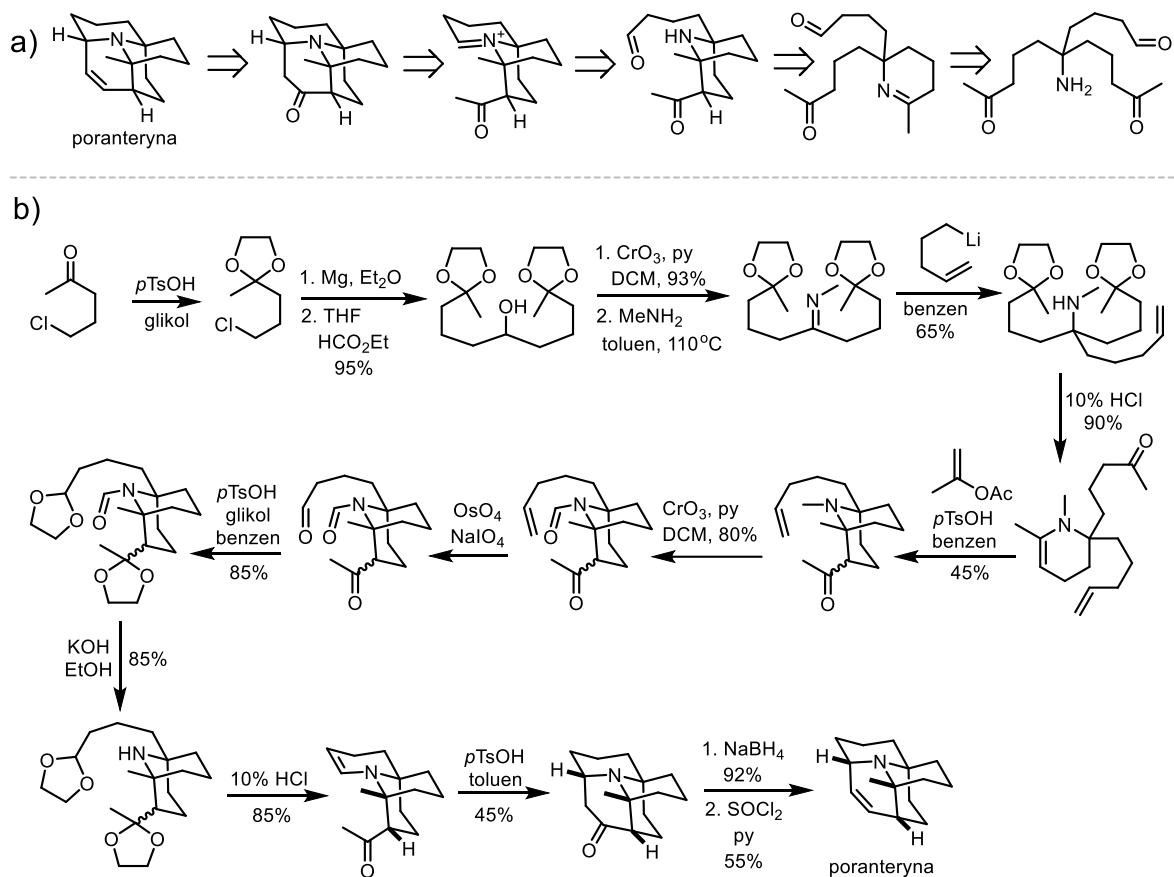
w syntezie Büchiego^{31,32} ukończonej w 1964 roku i dien **L5.11** z wcześniejszej syntezy Danishefsky'ego³³ z 1968 roku. Dien **L5.11** pojawia się też w późniejszych syntezach alkoholu paczulowego Näf'a³⁴ z 1974 roku, Bertranda³⁵ z 1980 roku i Magee³⁶ z 1995 roku. Co prawda w syntezach alkoholu paczulowego reakcją Dielsa-Aldera zastosowali Mirrington³⁷ w 1972 roku, Rao³⁸ w 1997 roku, Sun³⁹ w 2017 – wykorzystali jednak dieny nie występujące na przytoczonym drzewie retrosyntetycznym. Pozostałe podejścia: Yamady⁴⁰ z 1979 roku, Yu-Lin⁴¹ z 1991 roku i Srikrishny⁴² z 2005 roku, nie wykorzystują ani reakcji Dielsa-Aldera, ani proponowanych cięć retrosyntetycznych.



Ryc.5. Ilustracja wybranego drzewa retrosyntetycznego dla alkoholu paczulowego³.

(2) Poranteryna (Porantherine) – związek naturalny, którego syntezę totalną (Ryc.6b) Corey opublikował⁴³ w 1974 roku. Praca ta traktowała głównie o szczegółach syntetycznych, lecz w ostatnim jej akapicie odnotowano, że analiza retrosyntetyczna (Ryc.6a) została również uzyskana przy użyciu LHASA-10.

Fakt ten mógł sugerować^{22,43}, że to program stanowił podstawę planu syntetycznego, to znaczy, że najpierw przy użyciu LHASA stworzono plan syntetyczny, a następnie wcielono go w życie. Corey nigdy jednoznacznie tego nie twierdził, jednak już praca przeglądowa Wiese'go⁴⁴ z 1979 roku, właśnie w ten sposób zinterpretowała to zdanie. Synteza poranteryny opiera się na doborze grup ochronnych dla ketonów, aldehydu i aminy – tych informacji nie mogła dostarczyć LHASA w 1974 roku, gdyż wówczas nie posiadała jeszcze percepcji grup niekompatybilnych. Plan syntetyczny poranteryny wzbogacony o informację, które grupy funkcyjne należy ochronić został wymieniony dopiero w późniejszych pracach^{16,23} z 1976 i z 1985 roku – co ciekawe lista tych grup w obu publikacjach jest różna. Jako rozstrzygające warto tu przytoczyć stanowisko z jednej z ostatnich prac o LHASA²² z 1994 roku, gdzie autorzy jasno stwierdzają, że wyniki analiz programu były weryfikowane jedynie przez ich porównanie z precedensami literaturowymi.



Ryc.6. a) Ilustracja analizy retrosyntetycznej poranteryne, uzyskanej między innymi przy użyciu LHASA⁴³; b) Schemat faktycznie wykonanej przez Corey'ą syntezy totalnej poranteryne⁴³.

(3) W niniejszej części odnotowuję mniej znaczące analizy retrosyntetyczne, wraz z komentarzem o stopniu walidacji programu za ich pomocą. Ich pełniejsze omówienie wraz ze schematami znajduje się w **Appendiksie II**.

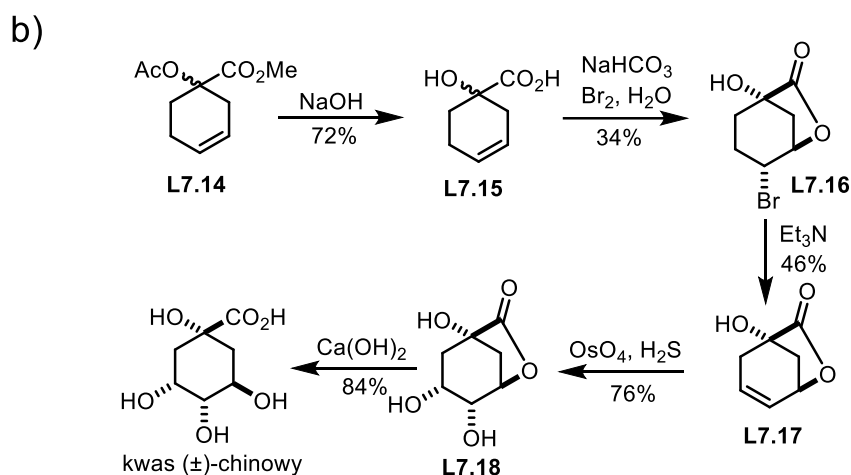
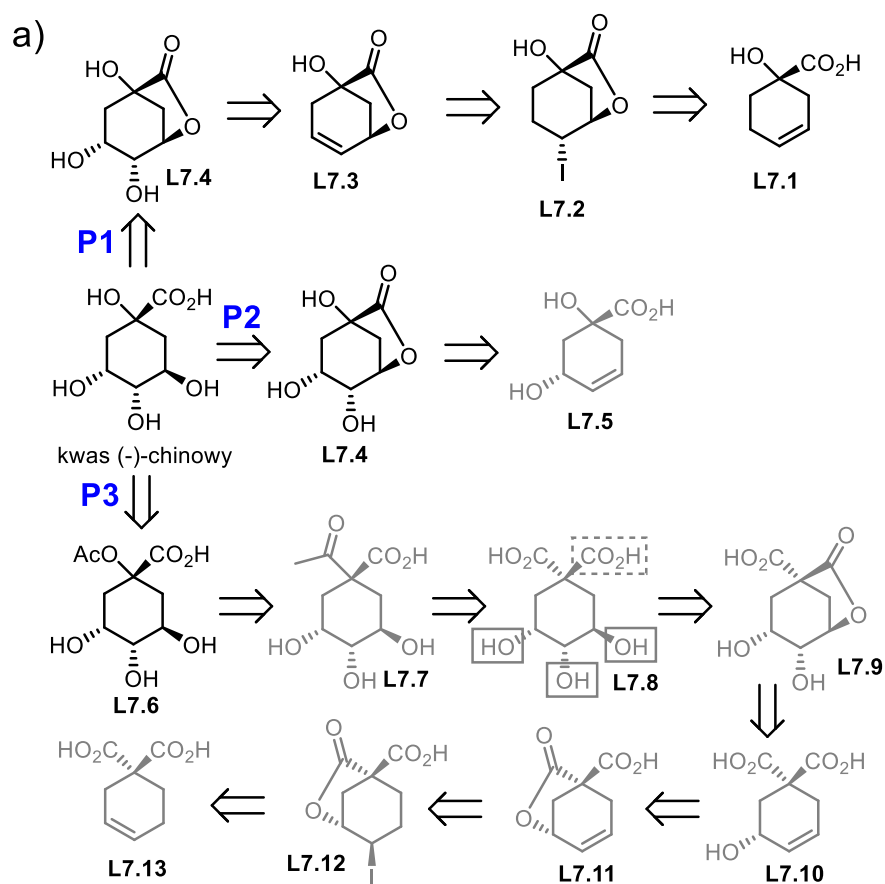
W pracy¹² z roku 1976 pojawia się analiza retrosyntetyczna thujopsenu, zawierająca trój etapową sekwencję odpowiadającą ostatnim etapom wcześniejszej syntezy totalnej Daubena⁴⁵. Jest to przykład walidacji przez porównanie części analizy z syntezą znaną z literatury.

Podobnie analiza retrosyntetyczna¹² ostatnich etapów syntezy prostaglandyny F_{2α} pokrywa się z końcówkami syntez totalnych, wykonanych wcześniej w grupie Corey'ą^{46,47} i Woodwarda⁴⁸. Cztery lata później, ukazała się praca dotycząca implementacji halolaktonizacji¹⁴, gdzie ponownie pojawia się plan syntetyczny prostaglandyny F_{2α}, tym razem uwzględniający ochronę grup funkcyjnych. Zawiera on elementy wcześniejszych syntez Corey'ą^{49,50}.

Następnym przykładem fragmentu analizy retrosyntetycznej¹² dla wcześniej wykonanej⁵¹ syntezy Corey'ą jest analiza caryophyllenu, wykorzystująca fragmentację Groba – reakcję niewystępującą w żadnej innej analizie z LHASA.

Analiza sativenu¹² zwraca uwagę odwrotną konfiguracją centrum stereogenicznego przy grupie izopropylowej w planie z LHASA i w związku uzyskanym wcześniej przez McMurry'ego⁵². Wygląda to na skutek pomyłki operatora wprowadzającego strukturę. Plan zaproponowany opiera się na przegrupowaniu karbokationu, nigdy potem nie wykonanym.

Trzy analizy dla waleranonu¹³ ilustrowały możliwości stosowania anulacji Robinsona. Powtarzają się one następnie pięć lat później, w pracy¹⁶ z 1985 roku. Jeden z planów przypomina



Ryc.7. a) Ilustracja 3 analiz retrosyntetycznych dla kwasu (-)-chinowego, uzyskanych przy użyciu LHASA¹⁴ w 1980 roku. Przerywaną ramką zaznaczono grupy niekompatybilne i wg. autorów niemożliwe do protekcji; ramką ciągłą zaznaczono grupy niekompatybilne i z możliwą protekcją.; b) Schemat syntezy totalnej kwasu (±)-chinowego, Wolinsky'ego⁹² z 1964 roku.

wcześniejszą syntezę totalną Marshall'a⁵³. Zastanawia fakt, że autorzy późniejszej pracy o LHASA, jako walidację programu, przedstawiają plan odrzucony w latach 60-tych przez Marshalla. Dwa pozostałe plany, nie zostały nigdy wykonane. Przykład prekursora wernolepiny¹⁴ pokazywał możliwość zastosowania przez LHASA halolaktonizacji. Autorzy nazwali cel syntetyczny *kluczowym prekursorem wernolepiny*. Tymczasem, żadna z 17 znanych⁵⁴⁻⁷⁰ syntez totalnych wernolepiny i jej fragmentów, nie wykorzystuje tego *kluczowego prekursora*. Wzmiankuje o nim

Grieco⁵⁴ rozważając jego użycie w pracy z 1977 roku. Już na wstępie zaznacza on jednak, że uzyskanie wynikającego z analizy retrosyntetycznej związku sześcioczłonowego o pięciu centrach stereogenicznych stanowi duże wyzwanie, które może być jednak zredukowane, gdy użyje się innego prekursora. Żadna z istniejących syntez totalnych wernolepiny nie przypomina tej zaproponowanej przez LHASA.

Apidikolina – licząca dwanaście etapów analiza retrosyntetyczna apidikoliny²⁰, zawiera niektóre z przejść znanych z dotychczasowego podejścia Trosta⁷¹. Nie została jednak opatrzona komentarzem dotyczącym ochrony grup funkcyjnych, czy stereoselektywności transformacji, kluczowych dla powodzenia tej syntezy. Pomimo wielu⁷¹⁻⁹¹ znanych syntez apidikoliny, plan LHASA proponuje przejścia nie występujące w żadnej z nich. Trudno plan ten uznać za przykład walidacji przez porównanie ze znanymi syntezami, gdyż autorzy oprócz dostrzeżenia analogii z początkiem syntezy Trosta⁷¹ nie dokonują porównania reszty planu.

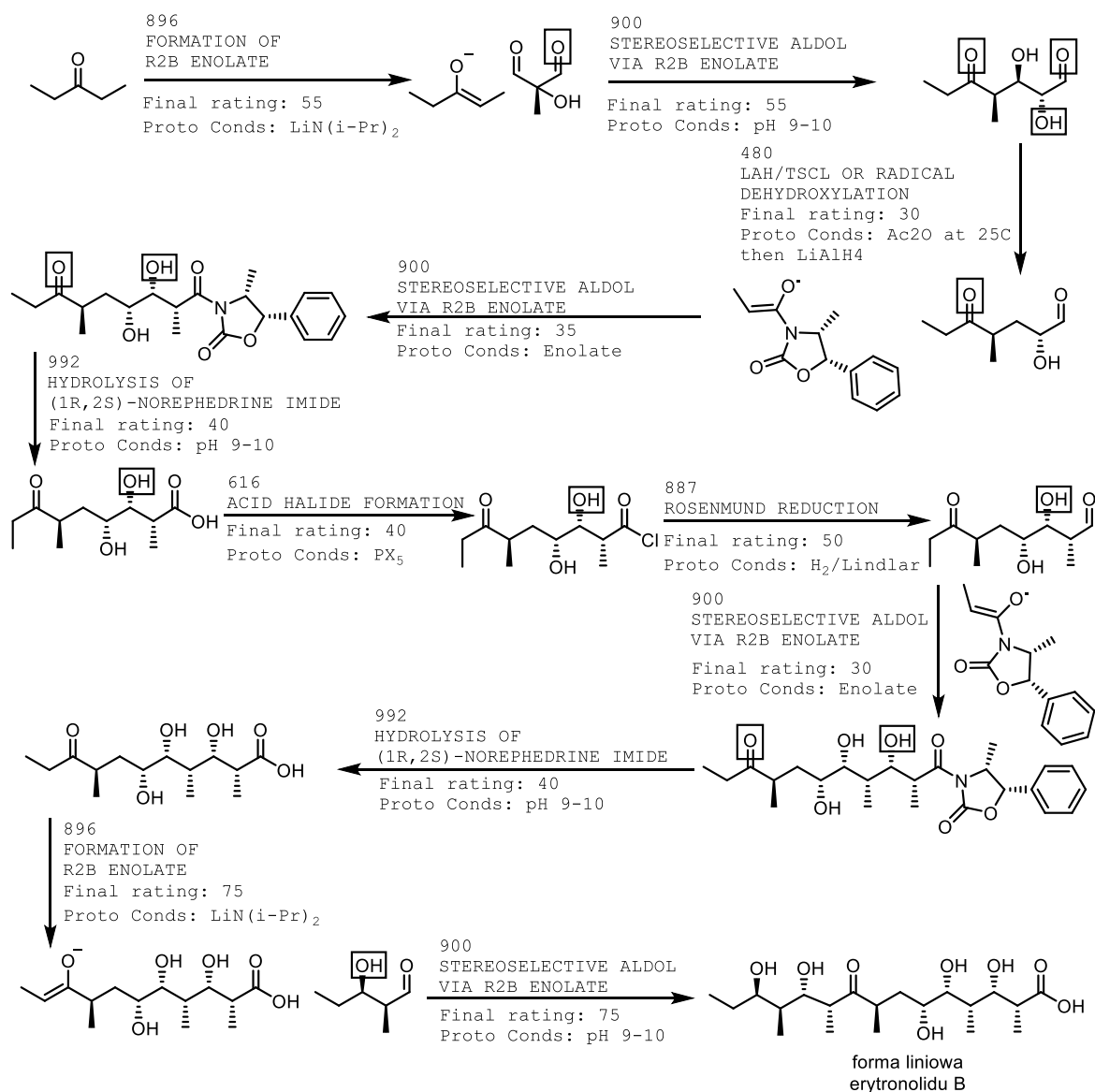
(4) Kwas (-)-chinowy – w pracy¹⁴ z roku 1980 zaproponowano dla niego trzy analizy retrosyntetyczne wygenerowane przy użyciu LHASA (Ryc.7a). Wszystkie z nich wykorzystują reakcję halolaktonizacji.

Plan P1 (górny panel Ryc.7a) zakłada wykorzystanie optycznie czynnego hydroksykwasu **L7.1**, autorzy nie podają jednak w jaki sposób uzyskać ten związek. Dopiero w 2000 roku w pracy Renauda⁹³ pojawia się jego 5-etapowa synteza adaptująca metodologię Seebacha⁹⁴ wykorzystującą dioksolanony, metatezę olefin oraz przejście analogiczne do **L7.1**→ **L7.2** z tą różnicą, że wykorzystano bromolaktonizację zamiast jodolaktonizacji. Również Pansere⁹⁵ w 2006 roku, przy użyciu pochodnych efedryny i następczej metatezy, uzyskał bromowy odpowiednik związku **L7.2**.

Jeżeli do przejścia wykonanego przez Renauda w 2000 roku, doda się etapy **L7.3**→**L7.4**→**kwasy chinowe** opublikowane przez Wolinsky'ego⁹² w 1964 roku dla racemicznego kwasu chinowego, wówczas plan P1 można uznać za zamknięty. Jednak największa trudność w syntezie kwasu chinowego, leży właśnie w syntezie prekursora zawierającego trzeciorzędowe centrum stereogeniczne. Ostatnie etapy planów P1 i P2 są identyczne, jednak przejścia **L7.5**→ **L7.4** nigdy nie wykonano, co dyskwalifikuje walidację planu P2.

Z najdłuższego planu P3 (dolny panel Ryc.7a) jedynie ostatnie przejście **L7.6**→**kwasy chinowe**, zostało wykonane przez Ogasawarę⁹⁶ w 1998 roku dla odpowiednika związku **L7.6** o zabezpieczonych grupach hydroksylowych, czego plan P3 nie sugerował. Etap **L7.13**→ **L7.12** generuje kluczowe centrum stereogeniczne, indukujące następne centra stereogeniczne w cząsteczce. Autorzy zauważają ten problem, jednak nie komentują w jaki sposób symetryczny kwas dikarboksylowy miałyby za pomocą tylko jednej wybranej grupy karboksylowej wejść w reakcję halolaktonizacji. Brak selektywności w sekwencji **L7.9**→ **L7.6**, autorzy proponują rozwiązać zamianą kolejności etapów, przeprowadzając reakcję Bayera-Villigera **L7.7**→ **L7.6** pozostawiając wciąż zamknięty lakton jak w związku **L7.9**. Miałyby być to forma ochrony i desymetryzacji grup karboksylowych. Podsumowując przedstawione analizy LHASA dla syntezy kwasu chinowego, nie stanowią atrakcyjnej alternatywy do znanego podejścia Wolinsky'ego z 1964 roku.

(5) Forma liniowa erytronolidu B – jej plan syntetyczny został zaprezentowany w pracy¹⁶ z 1985 roku. Jest to najpełniejszy plan syntetyczny kiedykolwiek zaprezentowany dla programu LHASA. Oprócz informacji o grupach koniecznych do ochrony, podane zostały również sugestie dotyczące warunków reakcji. Plan ten do wprowadzania kolejnych centrów asymetrii wykorzystuje powtórzoną czterokrotnie reakcją aldolową. Ścieżka ta, choć teoretycznie poprawna, wskazuje na bardzo lokalne traktowanie przez LHASA chiralności i percepcję kontekstu chemicznego ograniczającą się jedynie do kompatybilności/niekompatybilności grup funkcyjnych. Tutaj centra



Ryc.8. Ilustracja planu syntetycznego dla formy liniowej erytronolidu B, wygenerowanego¹⁶ przy użyciu LHASA w 1985 roku.

stereogeniczne stanowią dla programu inertyny fragment struktury – po ich wprowadzeniu i domyślnym zabezpieczeniu znajdujących się przy nich grup hydroksylowych przestają grać rolę. Występujący tu asymetryczny pomocnik Evansa, w rzeczywistości wykorzystano tylko na jednym⁹⁷ etapie syntezy totalnej erytronolidu B z 1989 roku. Grupa Corey'a ukończyła syntezę totalną⁹⁸ erytronolidu B w 1978 roku najdłuższy liniowy fragment syntezy liczył ponad 30 etapów. Zaprezentowany plan z LHASA zawiera tylko 11 etapów. Prowokuje to pytanie: skoro jest to krótszy plan (nawet uwzględniając dodatkowe etapy na wprowadzanie i usuwanie grup ochronnych), dlaczego do dziś nie został wykonany? Żadna dotychczasowa synteza erytronolidu B ani jego fragmentów⁹⁷⁻¹⁰⁶ nie przypomina tej wygenerowanej przy użyciu LHASA (Ryc.8.).

(6) Taksol – Na początku lat 90-tych, wśród zespołów zajmujących się syntezą totalną, trwał wyścig, kto pierwszy na drodze syntezy uzyska taksol¹⁰⁷⁻¹⁰⁹. W owym czasie, cząsteczka ta była uważana za jedną z najtrudniejszych do syntezy. Dlatego wybór takiego celu do analizy przez

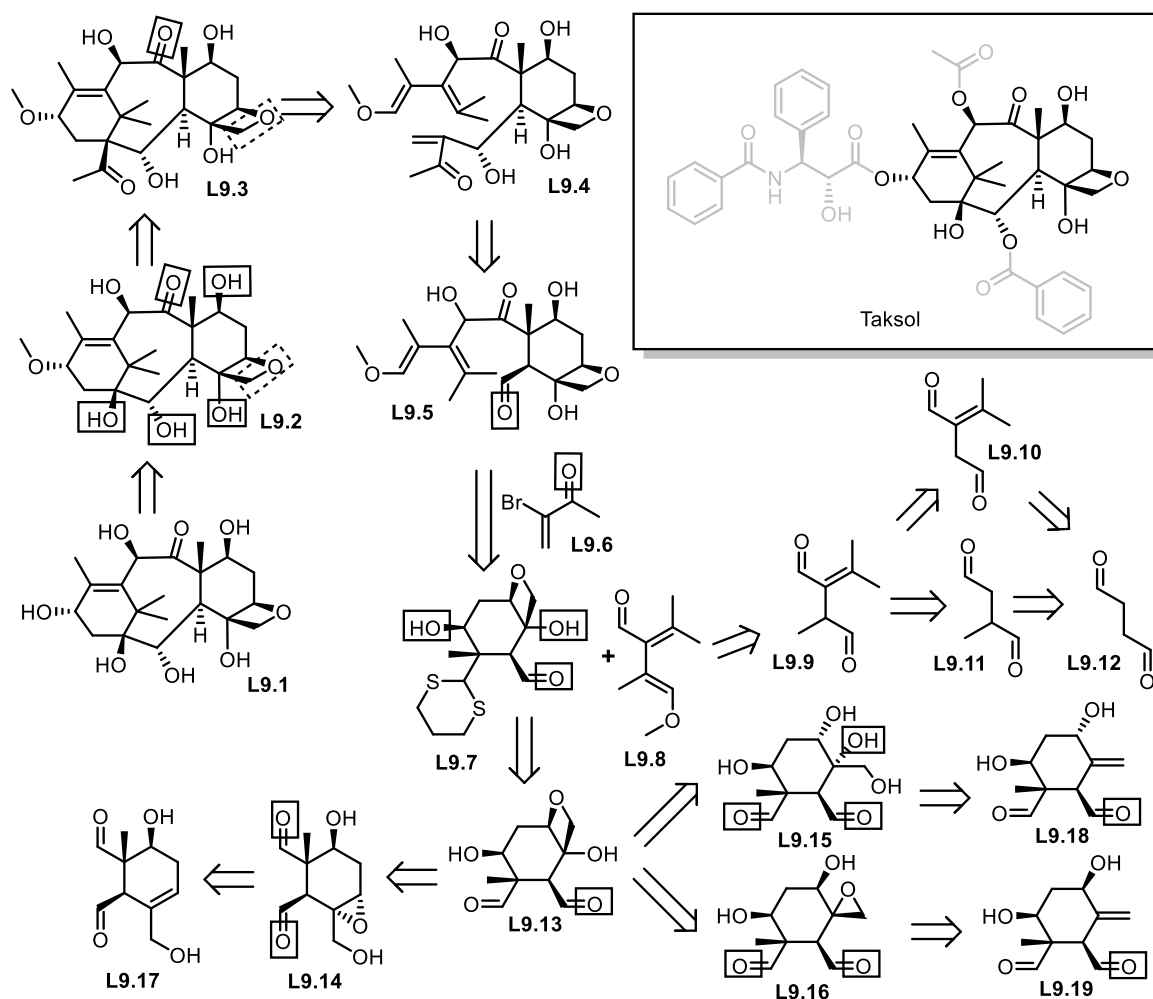
LHASA w pracy²² z 1994 roku był ambitnym celem. Była to też ostatnia praca o LHASA zawierająca plany syntetyczne.

Dla tworzonej od 25 lat LHASA, analiza retrosyntetyczna taksolu jest szczytowym osiągnięciem. Już na wstępie ułatwiono programowi zadanie, gdyż analizie retrosyntetycznej poddano nie cząsteczkę taksolu Ryc.9., a jej uproszczoną postać **L9.1**. Inspirując się tą analizą, autorzy zaproponowali plan syntetyczny dla związku, który jest jeszcze bardziej uproszczoną wersją związku **L9.1**, pozbawioną 3 centrów stereogenicznych i wrażliwego pierścienia oksetanowego.

Analiza retrosyntetyczna Ryc.9., opiera się na wykorzystaniu wewnątrzcząsteczkowej reakcji Dielsa-Aldera **L9.4**→**L9.3**, wybór tej transformacji nie dziwi, zważywszy na ilość pracy włożonej w opisane wcześniej zaimplementowanie strategii dla niej. Zaproponowany związek **L9.7** przypomina dialdehyd uzyskany przez Danishefsky'ego¹¹⁰ w 1992 roku. Ostatecznie nie posłużył on jako związek przejściowy w jego syntezach taksolu i bakkatyny z lat 1995¹¹¹-1996¹¹². Na uwagę zasługuje jednak fakt, że analiza LHASA dla dialdehydu **L9.13** kończy się na złożonych związkach **L9.17-19**, sposobu ich uzyskania nie zaproponowano. Dla porównania Danishefsky¹¹³⁻¹¹⁴, by uzyskać związek zbliżony do **L9.7** potrzebował 14 etapów.

W pracy²² z 1994 roku opisano 5 różnych dróg uzyskania **L9.5**. Stanowi to okazję zapoznania się z reakcjami tworzenia wiązania węgiel-węgiel, wytypowanymi przez LHASA. W planie 1) zaproponowana reakcja Dakina-Westa, wymagałaby użycia 6 ekwiwalentów¹¹⁵ złożonego strukturalnie bezwodnika. Plan 2) proponował wykorzystanie reakcji Claisena, niestety LHASA nie dostrzegła ryzyka racemizacji grupy hydroksylowej. Plan 3) zakłada użycie reakcji Michaela akroleiny z karboanionem w pozycji α do sulfotlenku i następnie zamianę sulfotlenku na keton w przegrupowaniu Pummerera. Propozycja ta budzi zastrzeżenia co do możliwości wygenerowania karboanionu na sulfotlenku w sposób selektywny po jednej stronie. Plan 4) proponuje reakcję związku Grignarda z aldehydem, jednak nie jest brana tu pod uwagę konkurencyjna reakcja z pierścieniem oksetanowym¹¹⁶. Do analizy retrosyntetycznej, wybrano reakcję umpolung ditianu **L9.7** z aldehydem **L9.8**. Jedno z końcowych przejść: **L9.5**→**L9.4** miało być realizowane za pomocą związku Grignarda uzyskanego z **L9.6** (Ryc.9.). Tymczasem, jak zauważają autorzy, o wiele dogodniejszym byłoby wykorzystanie reakcji Baylisa-Hillmana. Rzecz w tym, że LHASA nie mogła zaproponować jej użycia, gdyż nie zawierała w swej bazie tej reakcji.

Erik van Rozendaal w swojej pracy doktorskiej¹¹⁷ z 1994 roku, dotyczącej jedynej próby syntetycznej walidacji planów LHASA, opisuje syntezę odpowiedników związku **L9.8**. Związków **L9.10** i **L9.11** nigdy nie uzyskano – przyczyną tego może być niestabilność zaproponowanego przez LHASA jako związek wyjściowy aldehydu bursztynowego. Zamiast niego zastosowano diester etylowy kwasu bursztynowego. Enolan **L9.8** miał zostać przyłączony do **L9.7** za pomocą reakcji umpolung przy użyciu pochodnej ditianowej. Jednak autorzy²² wykazali, że nawet dla uproszczonego i pozbawionego pierścienia oksetanowego modelu związku **L9.7**, reakcja z aldehydami nie zachodzi. Zamiast tego zaproponowali inny sposób przeprowadzenia reakcji umpolung, wykorzystując pochodną cyjanohydryny. We wcześniejszej publikacji¹¹⁸ pokazali, że dla modelu związku **L9.7** pozbawionego grupy aldehydowej, reakcja umpolung zachodzi z szeregiem akrolein. Jednak Rozendaal¹¹⁷ wykazał, że dla przyjętego modelu i uzyskanych przez niego odpowiedników aldehydu **L9.8** reakcja umpolung z pochodną cyjanohydryny nie zachodzi. Oprócz jego pracy doktorskiej te wyniki nigdzie nie zostały opublikowane. Jest to o tyle ważne, że praca²² kończy się zapowiedzią syntezy modelu taksolu i ogłoszenia wyników w bliskiej przyszłości. Tymczasem już w tym samym roku 1994 autorzy wiedzieli, że nawet dla uproszczonego związku modelowego nie są w stanie zrealizować pierwszych etapów walidacji. Od ukazania się tej pracy minęło 25 lat, do dziś wyniki walidacji LHASA na modelu taksolu nie zostały opublikowane.



Ryc.9. Ilustracja analizy retrosyntetycznej dla makrocyklicznego rdzenia taksolu **L9.1**, wygenerowanej przy użyciu LHASA, pochodząca z pracy²² z 1994 roku.

Głównym celem pracy programu do wspomagania planowania syntezy (automatycznego lub semi-automatycznego), jest generowanie planów syntetycznych. Ich wartość można ocenić na kilka sposobów. Najpopularniejszym z nich jest odtwarzanie przez program istniejących ścieżek syntetycznych. Wcześniejsze ich wykonanie jest dowodem możliwości programu i poprawności planu. Jednak w przypadku programu semi-automatycznego, takiego jak LHASA, to człowiek decydował o wyborze kolejnych przejść, z tych, proponowanych przez program. Zatem znalezienie istniejących ścieżek nie udowodniało niczego poza tym, że prowadzące do nich reguły reakcji są zawarte w programie. Jeżeli chemik chcąc odtworzyć znaną ścieżkę nie znajdował wśród propozycji LHASA, związku przejściowego, należało wówczas dodać do eksperckiej bazy danych transformacji prowadzącą do niego regułę reakcji. Tylko jaki pożytek wynikał z takiej walidacji? Odwrómy tę kolejność: najpierw człowiek we współpracy z programem, znajduje plan syntetyczny, a dopiero następnie podejmuje się próbę jego wykonania w laboratorium. Jeżeli uda się w rzeczywistości wykonać taki komputerowy plan, wówczas możemy mówić o jednym przypadku pozytywnej walidacji. By mówić o pożytku synteza wykonana podług wskazań programu, musi być pod jakimś względem lepsza (nowatorska strategia, wyższa wydajność, mniejsza ilość etapów, łatwiejsze oczyszczanie, mniej produktów ubocznych etc.) od istniejących dotychczas, czyli opracowanych przez chemików, planów syntetycznych. Dla tak założonego kryterium, program LHASA można nazwać interesującym, ale nie pożytecznym.

2.1.2.5. Podsumowanie

It is important to test our understanding of synthetic analysis by writing and evaluating machine programs

E.J. Corey, 1969r.³

However, despite present limitations, including an incomplete knowledge base of about 2000 transforms, LHASA is capable of providing interesting suggestions of synthetic pathways for challenging targets. The present level of capability of LHASA can best be appreciated by its performance on specific problems.

E.J. Corey, 1990r.¹¹⁹

Projekt LHASA był najdłuższym w historii programem badawczym, mającym na celu „nauczenie” komputera chemii organicznej i planowania syntez¹²⁰. Jest on dobrym punktem odniesienia dla porównań z innymi programami do komputerowo wspomaganego planowania syntez. Przyjrząwszy się bezpośrednim rezultatom pracy programu, można wysnuć kilka wniosków i przestróg dla dalszych prób zmierzenia się z tym problemem.

Po pierwsze, LHASA opierała się na eksperckiej bazie danych o transformacjach chemicznych. Zalety i wady LHASA jak w soczewce skupiają się w definicji i ograniczeniach eksperckich baz danych. Zaletą programów opartych na eksperckich bazach danych²⁶ jest: 1) dobre uzasadnienie generowanych wyników, oraz łatwość prześledzenia drogi generowania wyniku; 2) generowanie wyników, których można byłoby się spodziewać po ocenie grupie ekspertów. Do wad natomiast należą: 1) kreatywność ograniczona do dokonywania predykcji dla nowych problemów; 2) systemy te nie są w stanie same odkrywać nowych reguł i wzorców w oparciu o surowe dane; 3) nie mogą się same uczyć. LHASA zatem mogła być co najwyżej tak dobra, jak baza danych, którą zawierała. Tymczasem, temu najważniejszemu dla działania programu modułowi, poświęcono przez 25 lat stosunkowo mało uwagi. Po tym czasie liczba transformacji w bazie wynosiła około 2000. Przytoczone rezultaty pracy programu wskazują, że było to zbyt mało dla generowania poprawnych planów syntetycznych. Zwraca uwagę fakt małej liczby reakcji tworzenia wiązania węgiel-węgiel. W żadnym z planów syntetycznych nie pojawia się chemia wykorzystująca sprzęgania katalizowane palladem, które w czasie tworzenia bazy danych LHASA stały się podstawowymi narzędziami syntetycznymi. Dodajmy do tego małą liczbę reakcji generujących centra stereogeniczne lub celowe stosowanie reakcji niestereoselektywnych w celu ich generowania. Wszystko to prowadzi do wniosku, że baza danych LHASA była zbyt mała w porównaniu do zadań, którym miała sprostać. Ostatecznie w 1994 roku¹²¹ pojawił się pomysł bardziej ogólnego używania reguł zawartych w bazie i w efekcie szybszej pracy programu. Niestety by móc to wcielić w życie bazę danych należałoby ujednotwić, przepisać od nowa i zaktualizować. Brak dalszych doniesień na ten temat każe przypuszczać, że nigdy tego nie zrealizowano.

Po drugie, percepcji struktury przez LHASA poświęcono najwięcej miejsca omawiając problemy: odróżniania pierścienia od przyłączonych doń łańcuchów bocznych, odróżniania pierścieni realnych od pseudo-pierścieni, analizie konformacyjnej cykloheksanu itp. W późniejszych publikacjach często brak dalszych wzmianek o nich i wpływie na poprawę uzyskiwanych planów syntetycznych. Tymczasem percepcja centrów stereogenicznych oraz konfiguracji wiązania podwójnego, kluczowych dla planowania syntezy, przez długie lata pozostawała nierozwiązanym lub rozwiązaniem połowicznie problemem. Oprócz jednego przypadku, dla liniowej formy erytronolidu B, w żadnej innej analizie retrosyntetycznej uzyskanej z LHASA nie pojawia się żadna inna reakcja stereoselektywna.

Po trzecie, scedowanie na użytkownika każdorazowej decyzji o wyborze transformacji, w praktyce ograniczało efektywną analizę retrosyntetyczną do ostatnich etapów syntezy. Rola programu sprowadzała się do dostarczania kolejnych generacji, prawdopodobnych chemicznie prekursorów analizowanego związku. Manualne poruszanie się po drzewie retrosyntetycznym, przesunęło ograniczenia programu na użytkownika: jego zdolność analizy chemicznej, oceny transformacji i pamięć. Ograniczenia ludzkiej percepcji były również sposobem uniknięcia i zamaskowania problemu eksplozji kombinatorycznej.

Po czwarte, ograniczenia sprzętowe i algorytmiczne. W latach 60-90-tych, kiedy rozwijana była LHASA, moc dostępnych procesorów stanowiła realne ograniczenie zadań im stawianych. Program mógł skończyć się błędem, gdy zachowano zbyt dużą liczbę gałęzi drzewa retrosyntetycznego, które należało na bieżąco usuwać w celu zwolnienia pamięci². Ponadto zespół pracujący nad programem składał się głównie z chemików próbujących „uczyć” komputer chemii organicznej tak, jak uczy się człowieka. Tymczasem specyfika i możliwości tego nowego narzędzia, wymagały zastosowania i tworzenia algorytmów operujących informacją chemiczną w sposób niedostępny dla człowieka i wcześniej przezeń niestosowany. Takie możliwości (sprzętowe i algorytmiczne) upowszechniły się dopiero w XXI wieku.

Trudno oprzeć się wrażeniu, że prace nad LHASA polegały na latach dodawania nowych funkcjonalności i wstrzymywaniu się z ostateczną walidacją zanim się ich nie doda, a kiedy części te stworzono i połączono w jednym programie okazało się, że suma modułów składowych nie sprawia, że program stał się chemicznie użyteczny. Zacytowane powyżej słowa ojca LHASA z jego mowy noblowskiej z 1990 roku¹¹⁹, ich ostrożny sceptycyzm i powściągliwe aspiracje są najlepszą, moim zdaniem, bezpośrednią oceną rezultatów tego projektu.

Oprócz bezpośrednich rezultatów, projekt LHASA, zaowocował również kilkoma osiągnięciami pośrednimi. Przede wszystkim był ważną pobudką do sformułowania retrosyntetycznych reguł analizy chemicznej, co zostało wyrażone w zacytowanych powyżej słowach Corey'a³ z 1969 roku, z pierwszej pracy o LHASA. Ponadto na potrzeby programu konieczne było uporządkowanie i skodyfikowanie wiedzy o chemii grup ochronnych i niekompatybilnościach reaktywności grup funkcyjnych i warunków reakcji. Zaowocowało to opublikowaniem w 1980 roku, klasycznego już i wznawianego do dziś podręcznika¹²² Theodory Greene *Protective groups in organic chemistry*. W przedmowie do jego pierwszego wydania czytamy: „*This work has been carried out in association with professor Elias J. Corey, who suggested the study of protective groups for use in computer-assisted synthetic analysis*”. Kolejną korzyścią dla współczesnych chemików była stymulacja prac nad reprezentacją struktur chemicznych w programach komputerowych. Pracujący nad LHASA w latach 80-tych Stewart Rubenstein¹⁶ wszedł następnie we współpracę z innym harwardzkim profesorem – Davidem Evansem, co z kolei zaowocowało powstaniem powszechnie dziś wykorzystywanego przez chemików programu ChemDraw¹²³. Te pośrednie, rezultaty projektu LHASA zasługują na uznanie, gdyż, o ile sam program nie stał się powszechnie używanym narzędziem, to prace nad przydaniem komputerowi umiejętności planowania syntezy stały się przyczyną stworzenia innych, powszechnie dziś wykorzystywanych, narzędzi chemicznych.

Projekt LHASA nigdy nie został oficjalnie zamknięty. Jego macierzysta strona internetowa: <http://lhasa.harvard.edu/> - jest już zamknięta, natomiast strona uniwersytetu w Nijmegen: cheminf.cmbi.ru.nl/cheminf/lhasa/ podaje informację, że ostatnia wersja programu to LHASA 20.3, zawierająca 2271 transformacji i 495 kombinacji taktycznych. Na stronie dostępna jest instrukcja obsługi programu, datowana na listopad 2002 roku. Można tam, również przeczytać, że program jest ciągle rozwijany przez grupę na Harvardzie. Jednak fakt, że harwardzka strona projektu została już zamknięta pozwala przypuszczać, że dalsze prace nad LHASA nie są już prowadzone.

2.1.3. SYNCHEM i SYNCHEM2

2.1.3.1. Wprowadzenie

SYNCHEM był programem rozwijanym od roku 1968, na Uniwersytecie Stony Brook, w grupie H. L. Gelerntera. Leżące u podstaw programu założenia są podobne jak w przypadku LHASA, jednak istnieje o wiele mniej publikacji opisujących jego działanie. Pierwsze doniesienie pochodzi z 1973 roku¹²⁴ – znajdują się tam dwie wygenerowane przez program analizy retrosyntetyczne (dla witaminy A i twistanonu). Najważniejszą pracą opisującą SYNCHEM i SYNCHEM2, jest publikacja w *Science* z 1977 roku¹²⁵. Omawia ona najbardziej spektakularne, wygenerowane przez SYNCHEM analizy retrosyntetyczne. Ostatnie prace pochodzą z lat 90-tych, jednak nie zawierają już one planów syntetycznych, a jedynie referują próby wykorzystania uczenia maszynowego do stworzenia bazy danych o reakcjach. Po roku 1998 nie pojawiają się już nowe doniesienia o tym programie.

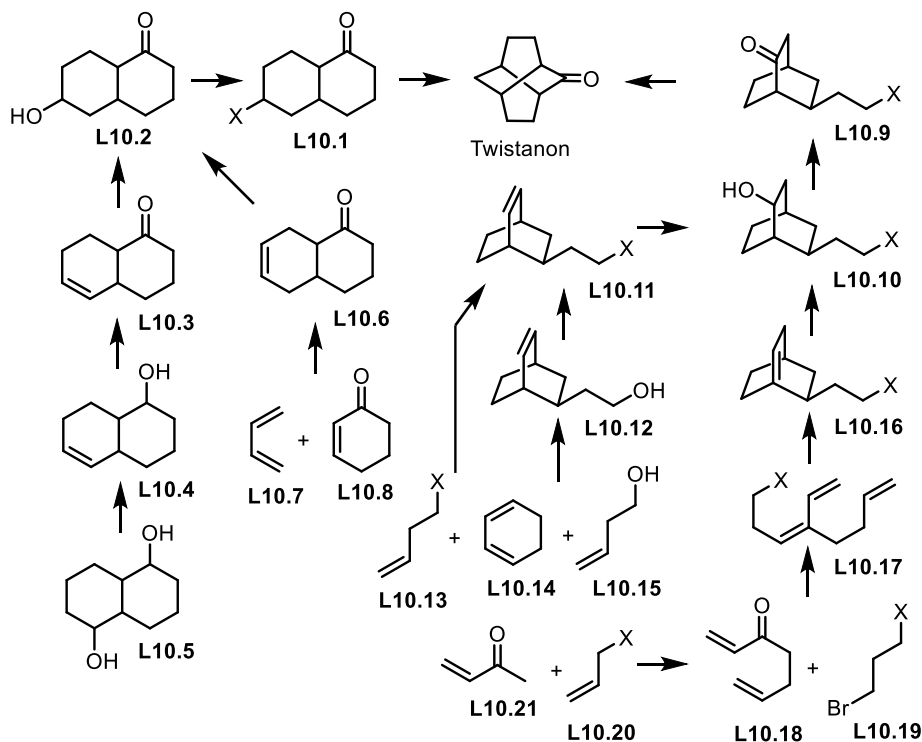
SYNCHEM w założeniu miał działać bez ludzkiego nadzoru, projektując plany syntetyczne nie gorsze niż te, których można by się spodziewać od doktoranta pierwszego roku¹²⁵. Dlatego też w pierwszej wersji nie zaimplementowano percepcji stereochemii, gdyż autorzy nie spodziewali się, że program kiedykolwiek będzie się mierzył z tak trudnymi problemami. Również dobór złożoności celów syntetycznych był ograniczony do tych znajdujących się poniżej ryzyka eksplozji kombinatorycznej. Wielokrotne zmiany, patche, nadpisy i poprawki w programie doprowadziły do jego dysfunkcyjności. Wszelkie próby naprawy kończyły się fiaskiem, między innymi z powodu zaniedbań w prowadzeniu dokumentacji tworzonego programu¹²⁵. Dlatego, korzystając z wcześniejszych doświadczeń, w 1974 roku autorzy zakończyli prace nad SYNCHEM, a rozpoczęli nad SYNCHEM2, który został już, między innymi, przystosowany do percepcji stereochemii. Omawiając wyniki autorzy nie zawsze precyzują czy uzyskano je przy pomocy SYNCHEM czy SYNCHEM2, dlatego dla uproszczenia będę odtąd posługiwał się nazwą SYNCHEM na obie wersje tego programu.

Informację strukturalną o celu syntetycznym wprowadzano za pomocą liniowej notacji Wiswessera (WLN), lecz również reprezentacja macierzowa lub TSD (topological structural description) były akceptowane. Cel syntetyczny był sprawdzany na obecność użytecznych syntetycznie grup funkcyjnych oraz motywów strukturalnych – nazwanych *syntemami*. Reakcje w bazie danych nazywano *reaction schemata*, które mogły być faworyzowane lub nie, w zależności od użytej heurystyki rozpoznającej obecność grup aktywujących dla danej „reaction schemata”, lub rozpoznając zatłoczenie steryczne. Odbywało się to podobnie jak zmiana oceny w LHASA. SYNCHEM mógł również zasugerować potrzebę użycia grupy ochronnej. Po wprowadzeniu struktury celu syntetycznego, SYNCHEM generował zestaw podcelów (*subgoals*) dla tej struktury, a poszukiwania ścieżki kończyły się na jednym z 3000 związków (zapisanych w formacie WLN) występujących w zaimplementowanym katalogu firmy Aldrich.

Problemem była ograniczona swoboda w poruszaniu się użytkownika pomiędzy graficzną reprezentacją cząsteczek a formatem WLN. Skomplikowane reguły WLN okazały się trudne do odczytania nie tylko dla chemików, lecz również ich konstrukcja – ubolewali twórcy SYNCHEM – udaremnia każdą próbę zaprojektowania rozsądnego i wydajnego algorytmu do ich przetwarzania. Te problemy z operowaniem formatem WLN, spowodowały jego zmianę na inny, stworzony specjalnie na potrzeby tego projektu system SLING (SYNCHEM Linear Input Graph). Ze względu na prostotę zapisu był on dla komputera dogodniejszy od WLN, który zachowano tylko do wyświetlania finalnego związku. Ponadto zapis WLN nie niesie informacji o elementach stereogenicznych w strukturze, a nowo stworzony SLING składał się z 2 zapisów dla każdej cząsteczki. Jeden opisywał konstytucję wiązań, drugi zawierał informację o stereochemii.

2.1.3.2. Rezultaty

Pomimo mniejszej niż w przypadku LHASA liczby doniesień literaturowych, rezultaty pracy programu zostały zwalidowane na przykładzie kilku tuzinów związków¹²⁴. W 1973 roku twórcy SYNCHEM oceniali, że poprawność chemiczna generowanych planów wahała się od dobrych i wykonalnych po naiwne. Dlatego nie zostały one nigdy opublikowane, a pracę SYNCHEM możemy dziś oceniać na podstawie planów syntetycznych dla pięciu związków, ogłoszonych w latach 1973 i 1977.



Ryc.10. Ilustracja planów syntetycznych dla twistanonu wygenerowanych przez program SYNCHEM^{124,125}.

Twistanon (tricyklo[4.4.0.0]dekan-2-on) jest bezpośrednim prekursorem cyklicznego węglowodoru - twistanu, uzyskanego pierwszy raz w 1962 roku przez Whitlocka¹²⁶. W pracy z 1973 roku¹²⁴ znajduje się jego plan syntetyczny (Ryc.10.), powtórzony następnie w pracy z 1977 roku¹²⁵. Składa się on z dwóch głównych wariantów retrosyntezy, bazujących na wewnątrzcząsteczkowym α -alkilowaniu ketonu **L10.1** lub **L10.9**.

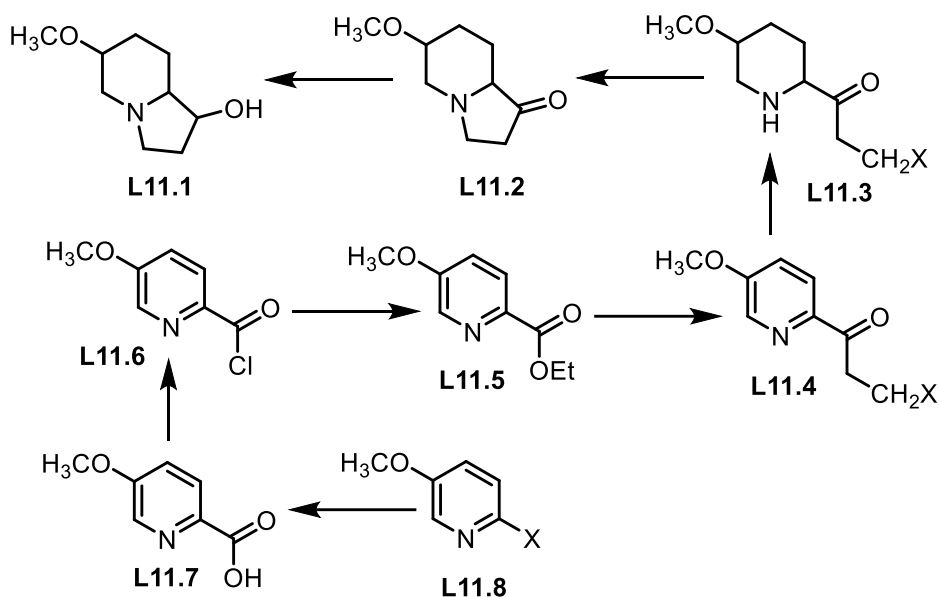
Propozycja alkilowania **L10.9** powtarza pierwszą syntezę twistanu Whitlocka¹²⁶ z 1962 roku, gdzie grupa przedstawiona przez SYNCHEM jako X była estrem mesylowym. Dalsza analiza tego planu prowadzi nas do osobliwego związku **L10.16**. Zaproponowanie takiej cząsteczki ujawnia brak percepcji przez SYNCHEM naprężonych struktur łamiących regułę Bredta¹²⁷. Tego typu naprężone bicykliczne szkielety można uzyskać jedynie w specjalnych warunkach¹²⁸, jednak do dziś brak przykładów ich zastosowań syntetycznych. Należy również zauważyć, że przejście **L10.11**→**L10.10** nie jest regioselektywnym sposobem wprowadzenia grupy hydroksylowej. To do którego z dwóch atomów węgla przyłączy się tlen jest decydujące dla uzyskania planowanego produktu. Brak kontroli powstającego centrum stereogenicznego jest tutaj mniejszym problemem wobec następczego utlenienia alkoholu. Warto tu przytoczyć dla porównania metodę zastosowaną przez Whitlocka, który wykorzystując reakcję jodolaktonizacji, w regio- i stereokontrolowany sposób zaktywował to wiązanie podwójne.

Druga propozycja, kończąca się alkiłowaniem **L10.1** stanowi wariację podejścia Deslongshamps z 1966 roku¹²⁹, który również oparł syntezę na szkielecie dekaliny. Tutaj dekalinę uzyskuje się w reakcji Dielsa-Aldera **L10.7** + **L10.9** → **L10.6**, lub kończy analizę na dihydroksydekalinie **L10.5**. Deslongchamps uzyskał podobny diol w redukcji odpowiedniego naftalenu. Na uwagę zasługuje tu selektywne hydroksylowanie podwójnego wiązania w **L10.6** Brak tu informacji o regioselektywności wprowadzenia grupy hydroksylowej na wybrany jeden z dwóch możliwych atomów węgla. Na uwagę zasługuje sekwencja **L10.5**→**L10.4**→**L10.3**→**L10.2**. Wobec braku stosowania grup ochronnych, program dyssymetryzuje diol **L10.5** by móc przeprowadzić selektywne utlenienie jednej już grupy hydroksylowej w przejściu **L10.4** → **L10.3**. Autorzy nie komentują czy ta sekwencja to tylko zbieg okoliczności czy celowy sposób obchodzenia przez SYNCHEM konfliktów reaktywności. Należy podkreślić, że zaproponowana dyssymetryzacja diolu **L10.5** nigdy nie została wykonana. Produktem, jakiego można się by spodziewać po eliminacji zaproponowanego diolu nie jest **L10.4**, lecz produkt o bardziej podstawionym wiązaniu podwójnym. SYNCHEM zdaje się nie dostrzegać ryzyka zajścia eliminacji w kierunkach innych niż założone.

Podsumowując, przykład planu syntetycznego dla twistanonu pokazuje możliwości generowania kilkietapowych sekwencji przy użyciu SYNCHEM. Program posługuje się w tym celu wyłącznie prostymi reakcjami zamieniającymi grupy funkcyjne. Nie bierze przy tym pod uwagę, że nawet proste reakcje takie jak eliminacja grupy hydroksylowej, czy hydroksylowanie wiązania podwójnego, mogą biec na różne sposoby w zależności od cząsteczki, w której się znajdują, a brak kontroli tych procesów skutkować będzie niewykonalnością planu syntetycznego. Przedstawione plany zawierają elementy przypominające istniejące już syntezы twistanonu i tylko w tym zakresie, przez powtórzenie, jest to walidacja pozytywna. Pozostałe zaproponowane przez SYNCHEM etapy budzą wątpliwości co do możliwości ich przeprowadzenia i tu walidacja wypada negatywnie.

W pracy z 1973 roku¹²⁴ znajdujemy plan syntetyczny dla witaminy A, o którym twórcy w kolejnych pracach już nie wspominają. Tworzenie wiązań węgiel-węgiel jest tu realizowane wyłącznie przy użyciu reakcji Wittiga, co pozwala podejrzewać niedobór w bazie danych reakcji tworzenia wiązań węgiel-węgiel.

Eter slafradiolu – w pracy¹²⁵ z 1977 roku przedstawiono plan syntetyczny dla metylowego eteru slafradiolu **L11.1** (Ryc.11.). W odróżnieniu od planów dla witaminy A i twistanonu, ten zawiera dużo bardziej przekonujące przejścia retrosyntetyczne. Uwagę zwraca jednak transformacja **L11.5**→**L11.4**. O ile każdy z etapów jest opatrzony dokładną nazwą transformacji jakiej należałoby dokonać, to ów przejście zostało nazwane: *naive reaction [...] that could be accomplished by a variety of multistep sequences*. Jest to o tyle zastanawiające, że w całym tym planie syntetycznym, to właśnie to przejście jest najmniej „nawne” i próby przekształcenia estru w β-haloketon wymagałyby wydłużenia całej ścieżki o kilka etapów. W literaturze nie ma dogodnych metod pozwalających zrealizować ów przejście za pomocą sekwencji krótszej niż 3-4 etapy. Praca przeglądowa Weisego⁴⁴ z 1979 roku błędnie podaje przykład tego związku jako walidację SYNCHEM przez stworzenie planu przez program i następczą syntezę według tego planu. Syntetyczna realizacja tego planu nigdy nie została opublikowana, a sami autorzy¹²⁵ podkreślają, że co najwyżej przypomina on ścieżkę syntezy **L11.1** zrealizowaną kilka miesięcy wcześniej na Uniwersytecie w Illinois. Można więc zaklasyfikować ten przykład jako częściową walidację przez powtórzenie istniejącej syntezy.

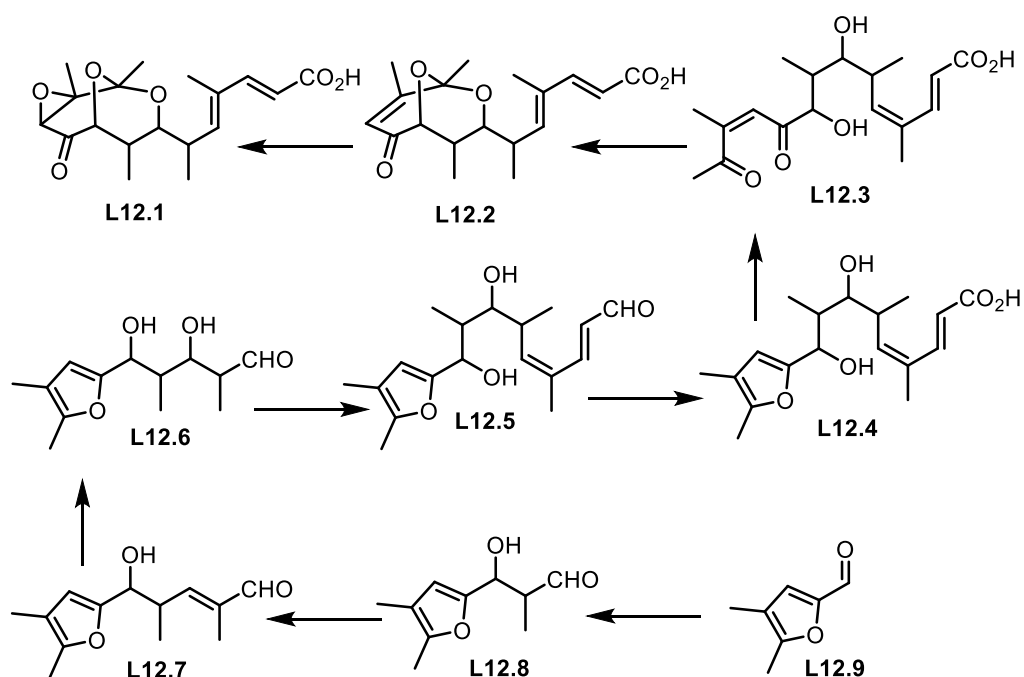


Ryc.11. Ilustracja planu syntetycznego dla eteru metylowego slafradiolu **L11.1**, wygenerowanego przez program SYNCHEM¹²⁵.

Pochodna chinuklidyny – praca¹²⁵ z 1977 roku podaje 13-etapową ścieżkę prowadzącą do podjednostki strukturalnej występującej w alkaloidach kory chinowca. Należy tu jednak zaznaczyć, że plan automatycznie wygenerowany przez SYNCHEM kończy się na związku czwartym od końca. Ze względu na zapis reakcji Wittiga wskazujący na niekompatybilność grupy karboksylowej w tej reakcji, ostatnie etapy nie mogły zostać wygenerowane – dopisano je później. Sekwencja wydaje się planem poprawnym, choć nieco zbyt długim w porównaniu ze złożonością celu syntetycznego. W przypadku szkieletu alkaloidów chinowych, wykorzystywanych przede wszystkim jako ligandy zapewniające stereokontrolę reakcji, brak uwzględnienia właściwości stereochemicznych cząsteczki w planie syntetycznym jest szczególnie dojmujący.

Tirandamycic acid jest najbardziej złożoną cząsteczką analizowaną przy pomocy SYNCHEM, rezultaty opublikowano¹²⁵ w 1977 roku. Kwas ten jest produktem degradacji naturalnie występującego antybiotyku tirandamycine¹³⁰ (brak nazwy polskiej). Za analizą przedstawioną na Ryc.12. kryje się pouczająca historia mutualnego projektowania syntezy. Problem syntezy wspomnianego antybiotyku i jego pochodnych stanowił w owym czasie przedmiot zainteresowania¹³¹ grupy Rineharta z Uniwersytetu w Illinois. Stamtąd też pochodził impuls wykorzystania programu do zaprojektowania syntezy **L12.1** lub choćby pomocy w jego analizie retrosyntezy. Wprawdzie dostrzeżenie przez SYNCHEM wewnątrzcząsteczkowej ketalizacji **L12.3**→**L12.2** jest kluczowe w analizie **L12.1**, jednak na tym pierwsze analizy programu się zatrzymywały. Dopiero obserwacja poczyniona przez F.W. Fowlera ze Stony Brook, który dostrzegł w motywie 1,4-diketonu pochodną utlenienia pierścienia furanu, pozwoliła, po wprowadzeniu tej reakcji do bazy danych, dokończyć analizę.

Tutaj, podobnie jak w przypadku planu syntetycznego witaminy A, dominującą reakcją wydłużania łańcucha węglowego jest reakcja Wittiga (**L12.8**→**L12.7**, **L12.6**→**L12.5**). Spotykamy



Ryc.12. Ilustracja planu syntetycznego dla **L12.1** (tirandamic acid) wygenerowanego przez program SYNCHEM¹²⁵.

również reakcję aldolową **L12.9**→**L12.8** – w tym przypadku na docenienie zasługuje dobór substratów ułatwiający jej selektywne przeprowadzenie. Nietrywialnym zadaniem byłoby natomiast przeprowadzenie selektywnego epoksydowania na ostatnim etapie **L12.2**→**L12.1**. Spośród 3 obecnych w cząsteczce wiązań podwójnych dwa są sprzężone z grupami elektronoakceptorowymi, co utrudnia chemoselektywne przeprowadzenie reakcji. Ponadto zaproponowana reakcja, oprócz reaktywności wobec wybranego wiązania podwójnego, musiałaby dokonać ataku w sposób stereoselektywny, od strony mostka ketalowego. Można by posłużyć się wówczas koordynacją za pomocą mostkowego atomu tlenu, jednak uzyskanie podwójnej selektywności w tej reakcji w dalszym ciągu można określić mianem nietrywialnego zadania. Oprócz zaproponowania transformacji generujących żądaną konstytucję atomów, uzupełnienie o komentarz, jak należałoby wykonać reakcję w sposób stereoselektywny – okazuje się niezbędne dla połowy etapów. Takiego komentarza ani autorzy ani SYNCHEM nie udzielają.

Ostatniej walidacji SYNCHEM dokonano¹²⁵ przy pomocy grupy ekspertów. Test przeprowadzono przy współpracy grupy pracowników odpowiedzialnych za syntezę chemiczną w dobrze znanej firmie farmaceutycznej (jej nazwy nie ujawniono). Do badania wytypowano związek chemiczny o prostej strukturze (której nie ujawniono), ale ze względów komercyjnych istotny dla wspomnianej firmy. Testowany SYNCHEM w ciągu 30 minut wygenerował 8 planów syntetycznych dla analizowanego związku, podczas gdy grupa ekspertów w ciągu kilku godzin zaproponowała około 12 planów. Spośród tych stworzonych przez człowieka planów, wszystkie, do których program posiadał odpowiednie reguły reakcji w bibliotece, również zostały znalezione. Ponadto SYNCHEM zaproponował jedną ścieżkę, która nie znajdowała się na eksperckiej liście propozycji. Została ona określona jako potencjalnie wartościowa. Była to unikatowa w historii programów do wspomagania retrosyntezy metoda walidacji. Niestety zreferowano ją w sposób ogólnikowy, nie dając możliwości oceny choćby tej jednej, nieoczywistej dla ekspertów, ścieżki syntetycznej.

2.1.3.3. Podsumowanie

Rezultaty pracy programu SYNCHEM nie odbiegały swym zaawansowaniem od tych osiągniętych przez LHASA. Dużą różnicę stanowiła natomiast deklarowana samodzielność programu i zdolność generowania planów syntetycznych bez nadzoru człowieka. Autorzy jednak nie wyjaśnili jak dokładnie odbywała się ta praca pozbawiona ludzkiego nadzoru.

Potrzeba było około dekady rozwoju, od końca lat sześćdziesiątych do roku 1977, by SYNCHEM osiągnął swój szczyt możliwości. W następnych latach nie ukazała się już żadna analiza retrosyntetyczna wygenerowana przez SYNCHEM. Od końca lat 70-tych do końca 90-tych, ukazało się jeszcze kilka prac¹³²⁻¹³⁴ związanych z projektem, jednak świadczą one tylko o kolejnych problemach napotykanym przez twórców. Na podkreślenie zasługują dwa z nich. Po pierwsze zbyt mała baza danych (biblioteka) transformacji stanowiła oczywiste ograniczenie programu. Po ponad 20 latach rozwoju, w 1990 roku zawierała¹³³ ponad 1000 „reaction schemata”. Twórcy bez rezultatu próbowali wykorzystać uczenie maszynowe¹³³ do stworzenia większej, eksperckiej bazy danych w sposób automatyczny. Po drugie percepcja stereochemii, którą miał zapewniać SYNCHEM2, nigdy nie została owocnie zaimplementowana. Pomimo podjęcia tego tematu¹²⁵, żaden z przedstawionych rezultatów pracy programu, nie zawiera choćby jednej transformacji stereoselektywnej.

SYNCHEM był rozwijany w tym samym czasie co LHASA, w tym samym roku (1998) również ukazał się ostatni komunikat o postępach w projekcie. Zaprezentowane w *Science* z 1977 roku rezultaty pracy budziły nadzieję – w szczególności analiza skomplikowanego tirandamycic acid. Dalsze losy SYNCHEM dowiodły jednak jak wiele czynników jest niezbędnych do stworzenia planu syntetycznego i jakim wyzwaniem może być implementacja percepcji stereochemicznej. Dziś program SYNCHEM ma już znaczenie tylko historyczne.

2.1.4. SYNSUP

2.1.4.1. Wprowadzenie

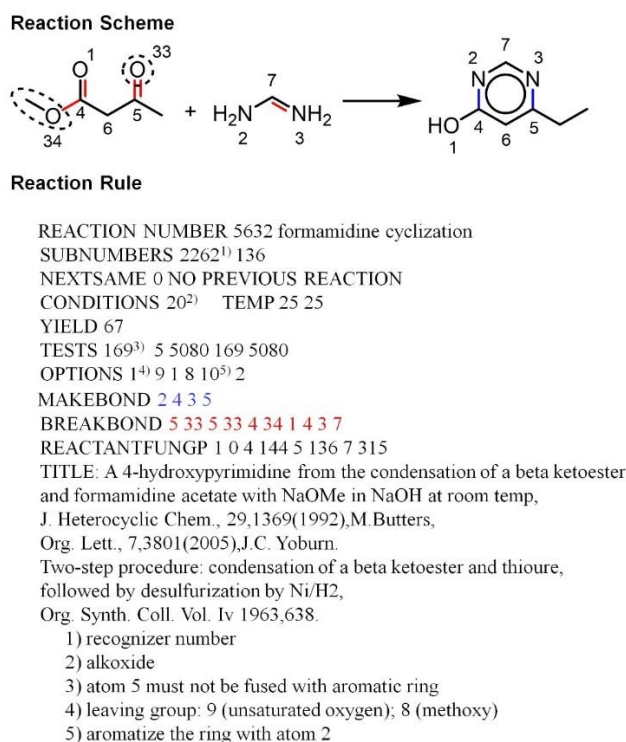
Już pierwsze doniesienie³ o LHASA z 1969 roku mogło budzić wątpliwości dotyczące zasady działania programu, a zwłaszcza manualnego poruszania się po drzewie retrosyntetycznym. O wiele bardziej zaawansowanym podejściem było generowanie przez program całych ścieżek – taki cel podjął¹³⁵ w 1971 roku Malcolm Behrson z uniwersytetu w Toronto. Już rok później ogłosił pierwsze doniesienie¹³⁶ proponujące zasadę działania programu wykluczającego potrzebę jakiegokolwiek ludzkiej interwencji w trakcie pracy. Jednak nie przedstawił wówczas ani jednej analizy retrosyntetycznej ilustrującej jego działanie, a program nie miał nawet swojej nazwy.

Po 12 latach od ogłoszenia pracy Behrsona do pracy nad projektem włączyła się japońska firma chemiczna Sumitomo Chemical Company¹³⁷. Nadano wówczas programowi nazwę SYNSUP-MB¹³⁷. Nazwa jest skrótem od SYNthesis Sumitomo Program¹³⁸, inicjały MB oznaczają przypuszczalnie udział Malcolma Behrsona. W późniejszych pracach używano już tylko nazwy SYNSUP. Pierwotnie program był napisany w języku LISP¹³⁶, ale już wówczas Behrson zasygnalizował potrzebę przepisania go na inny język, ostatecznie więc SYNSUP napisano w języku C++¹³⁹.

Pobudką do rozwoju tego programu do analizy retrosyntetycznej była potrzeba tworzenia „map reakcji”¹³⁸, to znaczy możliwie wielu wariantów syntez dla wybranych związków. W przemysłowym reżimie pracy nad procesem wytwarzania nowych cząsteczek, Sumitomo chciało

mieć pewność¹⁴⁰, że podczas opracowywania planów syntezy ich chemicy projektanci nie przeoczą jakiegoś ważnego wariantu oraz by do testów syntetycznych typować tylko najlepsze z nich.

Podczas pracy z programem, jako zapytanie wprowadzano strukturę analizowanego związku oraz specyfikowano warunki wyszukiwania za pomocą prostych wytycznych¹³⁷. Na strukturę nałożone było ograniczenie w postaci rozmiaru cząsteczki (nie więcej niż 72 atomy z wyłączeniem atomu wodoru)¹⁴¹ i nie więcej niż 5 centrów stereogenicznych¹⁴². Ograniczenia te zostały niezmienione, aż do końca projektu SYNSUP. Wytyczne pozwalały na ograniczenie obszaru poszukiwań ścieżki syntetycznej do najlepszych, najbardziej upraszczających strukturę¹⁴⁰ transformacji, występujących w każdej generacji. Można tu mówić o przykładzie podejścia depth-first¹⁴⁰ (w przeciwieństwie od breadth-first) jako sposobie uniknięcia eksplozji kombinatorycznej¹⁴². Podejście to było okupione kosztem w postaci ryzyka niezalezienia najlepszych planów



Recognizer

```

case 65: //phenolic OH, oxygen
...
case 282: //N-C=N, central carbon
makesb(mol,2262, // 4-hydroxypyrimidine
1,i,1, //sb[L][1] = i
1,j,7, //sb[L][7] = j
33,1,7,4, //sb[L][4] = NEAREST NBR OF ATOM 1 TO ATOM 7
113,4,6,1,0,0,2,2, //ATOM 4:ARYL C, 2 HETERONBR
3,4,7,2, //sb[L][2] = COMMON NBR OF 4 AND 7
113,2,7,1,0,0,0,0, //ATOM 2:UNSAT'D N, NO HETERONBR
57,2,2582, //RINGINDEX MUST BE 2582
13,4,1,2,6, //sb[L][6] = A THIRD NBR OF 4 NOT 1 OR 2
113,6,6,1,0,1,0,1, //ATOM 5:ARYL C,0-1 H, 0-1 HETERONBR
7,6,4,5, //sb[L][5] = OTHER RING NBR OF 6 BESIDES 4
113,5,6,1,0,1,1,1, //ATOM 5: ARYL C,0-1 H, 1 HETERONBR
3,5,7,3, //sb[L][3] = COMMON NBR OF 5 AND 7
113,3,7,1,0,0,0,0, //ATOM 3:UNSAT'D N, NO HETEROATOMS
0); //SUBSTRUCTURE END

```

Ryc.13. Przykład zapisu reguły reakcji chemicznej w programie SYNSUP¹³⁸.

syntetycznych ukrytych za maksimami lokalnymi w postaci reakcji naiwnych i nietypowanych do rozwinięcia. Narzuca się tu analogia z programami szachowymi posiadającymi sztywno zapisany warunek, by nie poświęcać hetmana, podczas gdy zniesienie tego warunku mogłoby pozwolić na spektakularne ruchy¹⁴². Najważniejszą wytyczną mającą wpływ na pracę programu było narzucenie ilości etapów w ścieżce. Za pomocą tego parametru sterowano wielkością odpowiedzi programu¹⁴², unikając tym samym generowania trudnego w analizie nadmiaru planów syntetycznych. W późniejszych wersjach SYNSUP w razie ustawienia zbyt małej wartości warunku STEPLIMIT i nieznaledzenia żadnej ścieżki automatycznie zwiększał jego wartość i kontynuował pracę¹³⁸. Lista 11 warunków pojawia się w pracy¹³⁷ z roku 1990 – oprócz narzucenia długości ścieżki pozwalają one również na narzucenie liczby reakcji nietrywialnych (tworzących wiązania C-C) lub wytypowania w celu syntetycznym wiązań, które nie mają być cięte w kierunku retrosyntetycznym. W podobnej postaci tabela z warunkami pojawia¹³⁸ się też w pracy z 2009 roku.

Reguły reakcji chemicznych stanowią podstawowy składnik klasycznych programów do planowania syntezy chemicznej. Nie inaczej było w przypadku SYNSUP. Przykład reguły reakcji stosowanej w tym programie zaprezentowano na Ryc.13. Reguła ta zawierała szereg informacji: od podstawowych o nazwie reakcji i jej ilustracji, przez odnośniki literaturowe i typowe warunki, aż po dane o jej wymaganiach sterycznych i elektronowych¹⁴². Twórcy SYNSUP, świadomi fundamentalnej różnicy między „reaction rule database” i „reaction example database”, jako krytyczny dla rozwoju programu określili właśnie proces powstawania eksperckiej bazy danych¹⁴². Wczesna wersja bazy danych z 1990 roku zawierała 2000 reakcji¹³⁷. Zakładając, że systematyczne jej tworzenie rozpoczęto od rozpoczęcia współpracy z Sumitomo, to w porównaniu z LHASA stworzenie 2000 reguł reakcji w 6 lat robi wrażenie. Już 2 lata później donoszono o około 2500 regułach reakcji¹⁴², w tym reakcjach enancjoselektywnych i diastereoselektywnych. Po tych pierwszych doniesieniach o działaniu SYNSUP następuje 14-letnia przerwa. W 2006 roku wspomniano o ponad 5500 regułach¹⁴⁰, w 2007 roku powtórzono wartość 5500¹³⁹, w 2009 roku mówi się o 5800 regułach¹⁴³ i w kolejnej pracy z tego roku już o ponad 5900¹³⁸. Co ciekawe w ostatniej pracy o SYNSUP z 2010 roku znów jest mowa o 5800 regułach reakcji¹⁴¹. Liczby te dwukrotnie przekraczają liczbę reguł posiadanych przez LHASA – dodając do tego automatyczne generowanie całych planów syntetycznych i percepcję stereochemii, można było mieć duże oczekiwania wobec tego programu.

Na docenienie zasługuje fakt ostrożnego podejścia do tworzenia reguł reakcji. Problem niepowtarzalności procedur syntetycznych jest zmorą każdego praktykującego chemika syntetyka. Dlatego wybór rzetelnego źródła informacji dla tworzonych reguł reakcji jest sprawą najwyższej wagi. W tym celu twórcy SYNSUP skorzystali z najbardziej wiarygodnego źródła, jakim jest *Organic Synthesis*. Posługiwali się też podręcznikami dla reakcji powszechnie uznanych. W ostatniej dekadzie rozwoju programu skupiono się w szczególności na dodawaniu reguł reakcji tworzenia pierścieni heterocyklicznych.

Oprócz prostego rozpoznania możliwości zastosowania reguły reakcji w konkretnym przypadku, SYNSUP posiadał narzędzia potrafiące w sposób bardziej zniuansowany ocenić reaktywność. Dlatego np. podczas epoksydowania wiązania podwójnego wykorzystywał informację, że najbardziej podatne na tą reakcję są niepodstawione wiązania podwójne, a podczas reakcji z użyciem nukleofila, że reakcji ulegnie szybciej chlorek kwasowy niż pierwszorzędowy chlorek alkilowy. Ten poziom wiedzy o reaktywnościach określono jako poziom I, w roku 1990 SYNSUP posiadał już jego elementy¹³⁷. Poziomu II, dostrzegającego na przykład różnice w reaktywności wynikające z konformacji, SYNSUP nigdy nie osiągnął. Następne narzędzie służyło zapobieganiu generowania nonsensownych związków przejściowych. Dzięki temu program nie proponował związków acylolitolowych lub generowania aromatycznych związków Grignarda posiadających

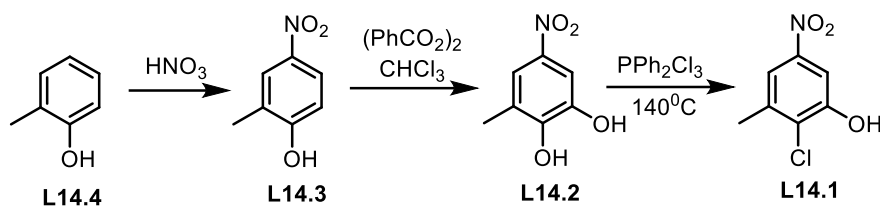
w pozycji *orto* atom halogenu¹³⁷. Nie były też proponowane związki nietrwałe, takie jak aldehyd propynowy, niepodstawiony cyklopentadienon lub cyklobutadieny. Za każdym razem, gdy program wirtualnie generował związek przejściowy, sprawdzał czy nie znajduje się on na liście cząsteczek niemożliwych. Naruszeniu reguł Bredta zwykle zapobiegało się na etapie sprawdzania wykonalności reakcji, zanim jeszcze generowany był związek przejściowy¹³⁷ – szczegółów tego procesu nie podano.

Do rozpoznawania kontekstu chemicznego i niekompatybilności pomiędzy grupami funkcyjnymi w cząsteczce oraz między grupami a warunkami reakcji, służyły tabele reaktywności. Pojawiają się one już w wersji z 1990 roku¹³⁷, w pracy z 2009 roku¹³⁸ mowa natomiast o tabeli opisującej wzajemne oddziaływanie 150 cech strukturalnych wobec 180 warunków reakcji. Jednak, według autorów, była to wciąż zbyt mało subtelna percepcja reaktywności. Jako przykład podają tu różne rodzaje eterów: etery *tert*-butylowe czy benzytowe są stosowane jako grupy ochronne i ulegają reakcji w warunkach odpowiednio kwasowych lub H₂/Pd-C, które trudno uznać za szkodliwe dla wszystkich eterów. Dlatego autorzy uzupełnili SYNSUP informacjami pozyskanymi z ekstrakcji baz danych w sposób uwzględniający powyżej zasygnalizowaną konieczność wieloznacznej percepcji w obrębie jednej grupy funkcyjnej¹⁴¹. SYNSUP zawierał również zainkorporowane tablice pK_a – służyły one do określania ilości ekwiwalentów zasady dla każdego konkretnego przypadku¹³⁷.

Rezultat pracy programu zawierał kompletną sekwencję syntetyczną, ilustracje poszczególnych reakcji, ich nazwy, najpopularniejsze warunki i odnośniki literaturowe do precedensów ich stosowania¹³⁵. Ścieżki kończyły się na związkach znajdujących się na liście umieszczonej w programie. Zawierała ona około 900 000 związków chemicznych, z czego 30% posiadało przypisaną cenę i tylko 12% numer CAS. Po wygenerowaniu ścieżki było niemożliwe sprawdzenie, czy jest ona znana w literaturze. W tym celu SYNSUP musiałby mieć znacznie większą bazę danych. Jest to o tyle ważne, że jak w 1990 roku przyznali twórcy: program często odtwarzał warianty znanych w literaturze ścieżek¹³⁷. W 1992 roku¹⁴², pomimo 8 lat rozwoju, SYNSUP w dalszym ciągu nie proponował ścieżek znacząco lepszych od tych już znanych. Odwrotny test wypadał jeszcze gorzej – analizując ścieżki literaturowe okazywało się, że w każdej ścieżce występowały zazwyczaj 1-2 typy reakcji, których nie zawierała jeszcze tworzona baza danych. Jest to przykład problemu czarnego łabędzia (*black swan*), gdzie rzadki przypadek – tu rzadko stosowanej reakcji – okazuje się niezbędny do przeprowadzenia niektórych syntez. Innymi słowy, baza danych o transformacjach, oprócz najbardziej popularnych przekształceń, musi zawierać również mniej typowe reakcje, gdyż to one warunkują kompletność ścieżek. Twórcy SYNSUP, po stworzeniu około 2500 reguł reakcji, byli świadomi tego problemu¹⁴².

2.1.4.2. Rezultaty

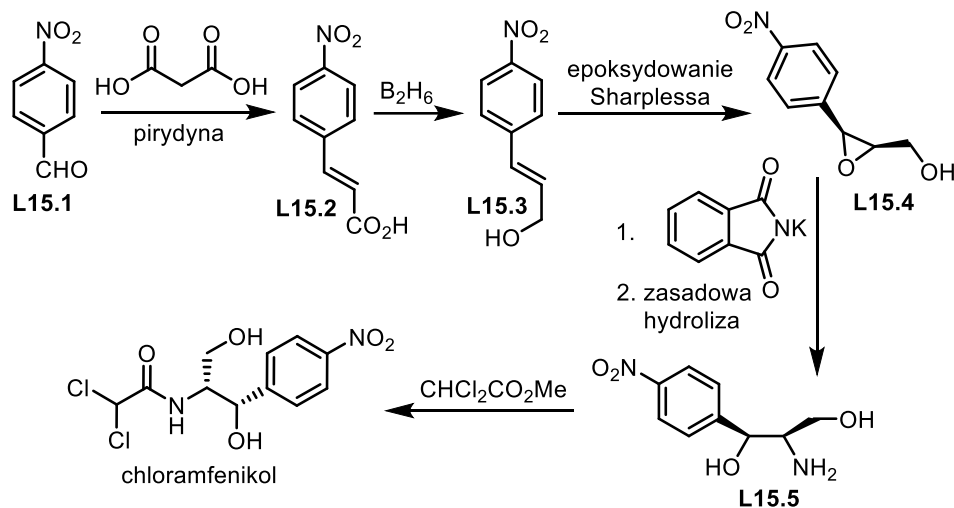
Poniżej omówię wybrane rezultaty pracy programu SYNSUP ze szczególnym naciskiem omawiając rodzaje transformacji chemicznych, do których był zdolny. W odróżnieniu bowiem od LHASA, dla którego nigdy nie zaprezentowano sekwencyjnych funkcjonalizacji układów aromatycznych, dla SYNSUP już w 1990 roku¹³⁷ przedstawiono rezultaty dla tej klasy przekształceń. Podobnie ze stereokontrolowanymi reakcjami, których dla SYNCHM (pomimo deklaracji) nigdy nie zaprezentowano, a dla LHASA wykazano jedynie w ograniczonym stopniu i dla jednego związku.



Ryc.14. Ilustracja planu syntetycznego dla czteropodstawionej pochodnej benzenu **L14.1** wygenerowana za pomocą SYNSUP i ogłoszona w 1990 roku¹³⁷.

Na Ryc.14. zaprezentowano plan syntetyczny będący dobrą ilustracją pracy z SYNSUP. Do wygenerowania go, konieczne było dodanie reguły reakcji selektywnego chlorowania przy użyciu Ph_2PCl_3 – umożliwiło to uzyskanie planu krótszego, jednak jak ostrożnie zauważają sami autorzy, selektywność przejścia **L14.2**→**L14.1** budzi ich wątpliwości.

Praca o SYNSUP¹⁴² z tematycznego numeru *Recueil des Travaux Chimiques des Pays-Bas* z 1992 roku, poświęconego właśnie programom wspomagającym analizę retrosyntetyczną, plasowała ów program na wyróżniającej się pozycji. Przedstawiono tam między innymi plan syntetyczny antybiotyku – chloramfenikolu (Ryc.15.) – wykorzystujący nowatorskie w tamtym czasie, enancjoselektywne epoksydowanie Sharplessa¹⁴⁴ **L15.3**→**L15.4**. W porównaniu z istniejącym podejściem¹⁴⁵ z lat 40-tych, przedstawiony plan wydaje się atrakcyjną alternatywą. Autorzy twierdzą¹⁴², że wprawdzie nie jest to wykonany eksperymentalnie plan, ale jest on krótszy od dotychczasowych. Nie jest to do końca prawda, gdyż podejście z 1975 roku¹⁴⁶ ma tyle samo etapów co przedstawione. W porównaniu do poprzedniej pracy również nastąpił postęp, gdyż w planie (Ryc.15.) nie ma już przejść budzących duże zastrzeżenia. Co więcej, operowanie



Ryc.15. Ilustracja planu syntetycznego dla chloramfenikolu, wygenerowanego przez program SYNSUP i ogłoszonego w 1992 roku¹⁴².

elementami stereogenicznymi w strukturze w przejściach **L15.3**→**L15.4** i **L15.4**→**L15.5** odbywa się swobodnie i w uzasadniony sposób.

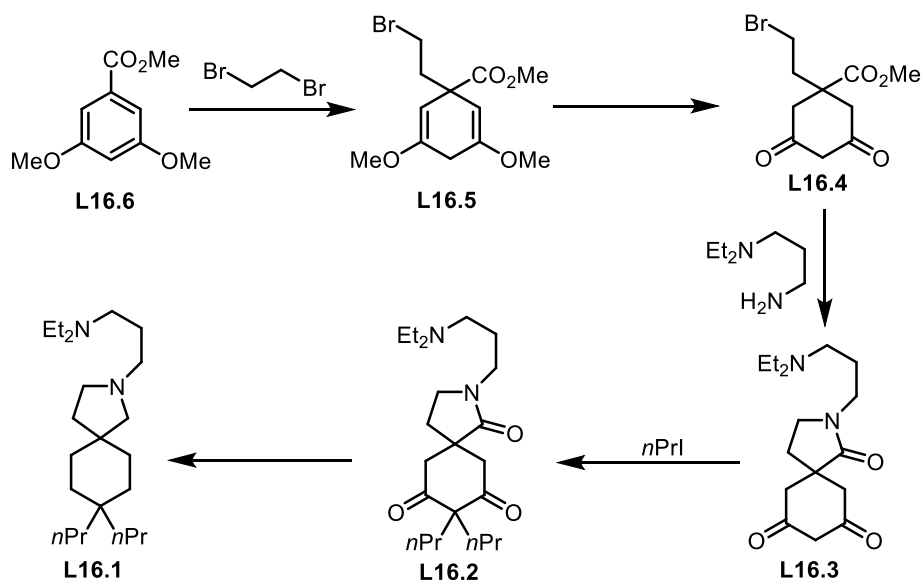
Oprócz przedstawionego przypadku, praca z 1992 roku podaje jeszcze inne analizy ilustrujące działanie SYNSUP, ale do prowadzonej dyskusji nie wnoszą one nic więcej.

Po ponad dekadzie dalszych prac, w 2006 roku ogłoszono kolejną pracę walidującą działanie SYNSUP¹⁴⁰. Przez ten czas ekspercka baza danych reakcji powiększyła się do 5500. Cząsteczka,

poddawana analizie retrosyntetycznej, azaspiran **L16.1** stała się obiektem zainteresowania Sumitomo ze względu na znaną aktywność biologiczną¹⁴⁷. SYNSUP wygenerował dla jej struktury ponad setkę planów syntetycznych. Pośród znalezionych przez program znalazły się syntezы znane już w literaturze – potraktować to należy jako znaną metodę walidacji, deklarowaną już w pracach z lat 90-tych.

Autorzy, wbrew istniejącym podejściom, wybrali nową strategię syntezy (Ryc.16.), bazującą na dimetoksyarylu **L16.6**. Posiada on grupę estrową stanowiącą podstawę do budowy pierścienia piroliny. Co ciekawe, w związku **L16.5** występują (w postaci zabezpieczonej) 2 grupy ketonowe, zlokalizowane w sposób umożliwiający (po ich demetylowaniu) selektywne wprowadzenie dwóch grup *n*-propylowych. Wykonywany eksperymentalnie plan syntetyczny (Ryc.17.) różni się od wygenerowanego przez SYNSUP w dwóch punktach. Po pierwsze, zaproponowanego dialkylowania przy użyciu jodku *n*-propylowego nie realizowano na aktywnej grupie metylenowej układu 1,3-dikarbonylowego, lecz przeprowadzono w obecności enoli utrwalonych jako etery metylowe (**L17.7**→**L17.6**). Po drugie, w planie komputerowym hydrolizy metylowych eterów enoli dokonywano przed redukcją grupy karbonylowej laktamu, tymczasem w wykonanej syntezie redukcja laktamu **L17.6**→**L17.5** następuje w połowie ścieżki, separując redukcję ketonów jako osobny problem. Obejście planowanego przejścia pozwoliło na uniknięcie powstawania ketiminy, będącej produktem reakcji konkurencyjnej dla zamierzonego alkiłowania pierwszorzędowego bromku.

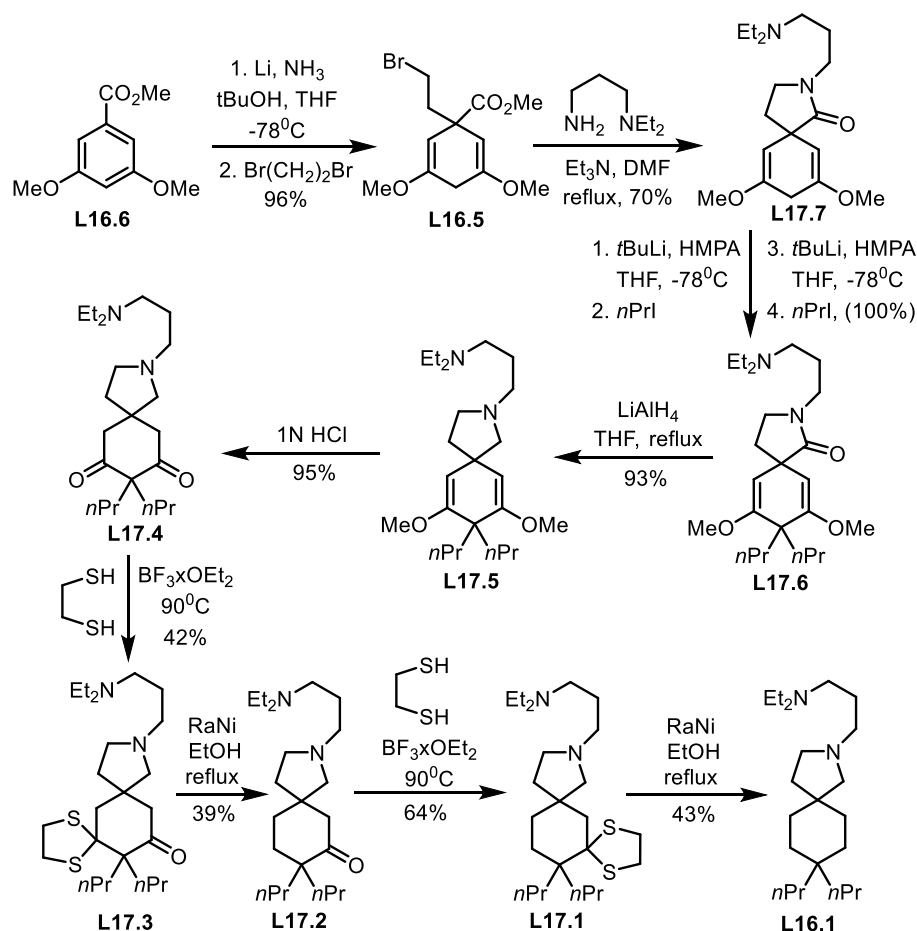
Może zastanawiać brak propozycji przez SYNSUP warunków redukcji grupy karbonylowej w laktamie i dwóch grup ketonowych (**L16.2**→**L16.1**). To właśnie realizacja syntetyczna tego przejścia sprawiła twórcom największe problemy. Próba wykonania redukcji Mozingo w dwóch etapach okazała się niewykonalna z powodu, niemożności jednoczesnego wprowadzenia dwóch grup ditioketalowych do **L17.4**. Najoczywistszym wytłumaczeniem tego faktu, wydaje się zbyt duże zatłoczenie steryczne w otoczeniu czwartorzędowego atomu węgla. Możliwe było jedynie wprowadzenie jednej grupy ditioketalowej na raz, co po uwzględnieniu umiarkowanej wydajności



Ryc.16. Ilustracja wybranego planu syntetycznego dla: azaspiranu **L16.1**; wygenerowanego przez program SYNSUP i ogłoszonego w pracy z 2006 roku¹⁴⁰.

procesu, znacznie obniżyło atrakcyjność tego podejścia. Inne sposoby redukcji grup karbonylowych, jak redukcja Wolffa-Kiznera, redukcja tosylohydrzonów, czy redukcja ketonów do alkoholi i próba redukcji odpowiedniego diolu po przeprowadzeniu w ditosylan nie przyniosły żadnego rezultatu.

Syntezę zaproponowaną przez SYNSUP zrealizowano, co stanowi rzadki przypadek uporu w syntetycznej walidacji komputerowego planu eksperymentalnego. Jednak całkowita wydajność tak przeprowadzonej syntezy, wynosząca jedynie 3%, w porównaniu ze znaną w literaturze o wydajności 25%, dyskwalifikuje ten plan jako konkurencyjną alternatywę¹⁴⁰. Jak zaznaczyłem wcześniej, jeżeli synteza wykonana w oparciu o program komputerowy jest gorsza od istniejących, program można nazwać interesującym i działającym poprawnie, lecz jeszcze nie pożytecznym. Tak też było w tym przypadku. Przedstawiona synteza jest jedyną ogłoszoną tego typu walidacją dla SYNSUP.



Ryc.17. Ilustracja walidacji syntetycznej zmodyfikowanego planu retrosyntetycznego dla: azaspiranu **L16.1**; wygenerowanego przez program SYNSUP i ogłoszonego w pracy z 2006 roku¹⁴⁰.

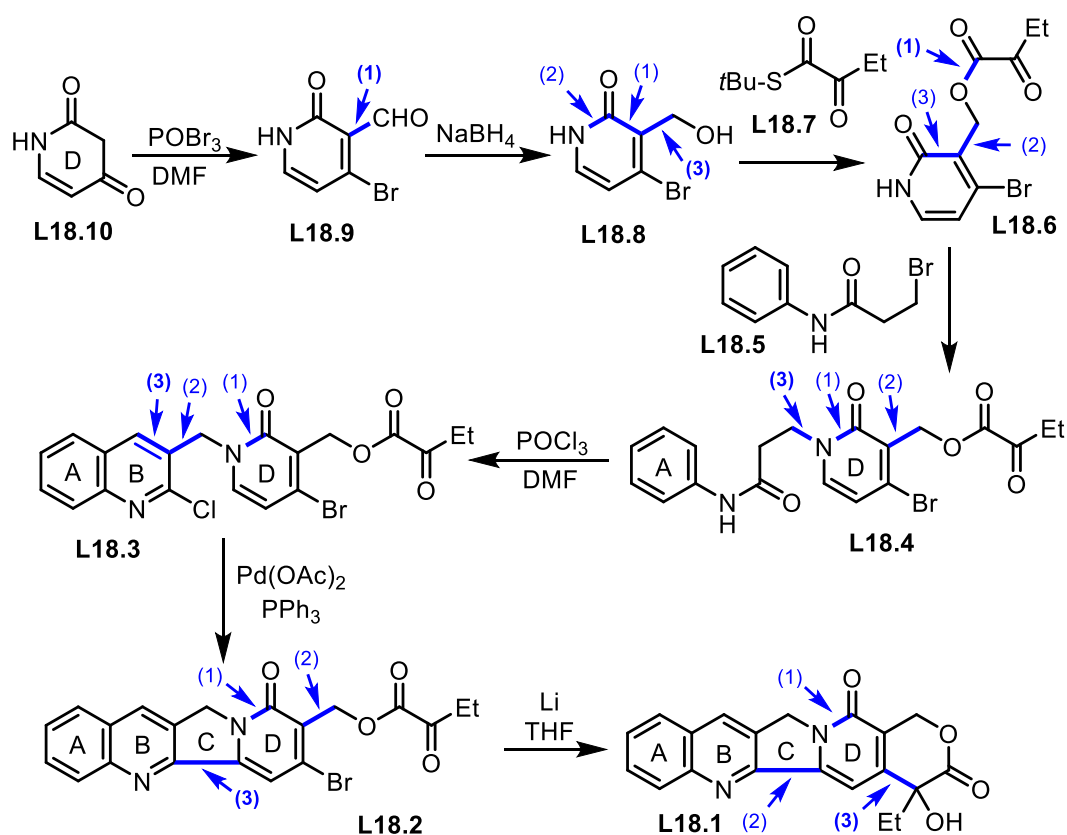
Ostatnie prace o SYNSUP z lat 2007-2010, donoszą o udoskonalaniu istniejącej zasady działania programu. Najbardziej znacząca spośród nich jest praca z 2009 roku¹⁴³. Opisuje ona nowy sposób na ograniczenie eksplozji kombinatorycznej. Kierowanie się największym uproszczeniem w każdej generacji, przy bazie danych zbliżającej się do 6000 reakcji, przestało być skutecznym sposobem na ograniczenie ilości generowanych planów. Kolejne zawężenie opierało się więc na typowaniu w każdej generacji wiązań najbardziej wartościowych retrosyntetycznie i tylko reakcje dokonujące ich tworzenia były brane pod uwagę. Twórcy wskazują tu na dwa podejścia: *selective search* w opozycji do *exhaustive search*. *Exhaustive search* jest dla opisywanego problemu

niewykonalny z powodów sprzętowych i algorytmicznych. Stąd pobudka do poszukiwania narzędzi wspierających *selective search*, dotychczasowe podejście typu *depth-first* już nie wystarczało, stąd zwrot ku jeszcze bardziej wyszukany metodom selekcji. Poprawność podejścia typującego wiązania sprawdzono na dwa sposoby.

Pierwszy sposób polegał na wytypowaniu grupy celów syntetycznych należących do różnych klas związków, których sposoby tworzenia potwierdziłyby poprawność typowania wiązań strategicznych. W tym celu posłużono się 25 produktami 27 syntez totalnych opisanych w pierwszym tomie klasycznej pracy Nicolaou¹⁴⁸. To porównanie typowania wiązań wypadło poprawnie dla SYNSUP.

Wykazawszy chemiczną poprawność typowania wiązań strategicznych, podjęto próbę wygenerowania planów syntetycznych dla 3 opisanych w książce związków. Jest wielce prawdopodobne, że SYNSUP w 2009 roku nie zawierał wszystkich transformacji wykorzystywanych w referowanych przez Nicolaou syntezach totalnych. Jednak dla 3 związków udało się przy użyciu komputera odtworzyć plany syntez bardzo zbliżone do znanych w literaturze¹⁴³. Są to zarazem ostatnie przykłady planów syntetycznym z SYNSUP, toteż analiza rodzajów transformacji pozwala na ocenę do jakiego stopnia rozwinięto program i jaka jakość analiz kryła się za niespełna 6000 reguł reakcji.

Kamptotecyna **L18.1** – posiada wielopierścieniowy rdzeń wymagający w analizie (Ryc.18.) uwzględnienia reakcji pierścieni heterocyklicznych. Na uwagę zasługuje przejście **L18.4**→**L18.3** budujące pierścień chlorochinoliny, ponieważ dla LHASA czy SYNCHEM nie obserwowaliśmy ani jednego przykładu tworzenia aromatycznego pierścienia heterocyklicznego – ta klasa związków



Ryc.18. Ilustracja planu syntetycznego dla kamptotecyny; wygenerowana przez program SYNSUP i ogłoszona w pracy z 2009 roku¹⁴³.

wymaga osobno zapisywanych reguł reakcji. Pewne wątpliwości może budzić propozycja przeprowadzenia **L18.3**→**L18.2** w obecności bromku winylowego, chętniej wchodzącego w reakcję z palladem niż chlorek arylowy. Zaproponowana tu metoda wyklucza stereokontrolowany przebieg reakcji **L18.2**→**L18.1**. Oprócz tych zastrzeżeń plan wygląda przekonująco. Co więcej, jego wygenerowanie bez zastosowania typowania wiązań zajmowało programowi niemal 10 godzin. W tym czasie znajdowano 42 ścieżki o maksymalnej długości 7 etapów. Tymczasem po wprowadzeniu typowania wiązań skrócono ten czas do 32 sekund (sic!), znajdując 4 ścieżki o maksymalnej długości 8 kroków. Ta znaczna poprawa czasu pracy, okupiona jednak jest ograniczeniem różnorodności rozwiązań. Podobna analiza dla innych związków wykazała, że nie jest to rezultat przypadkowy.

Na powyższym przykładzie widać znaczący postęp, jaki dokonał się w ciągu dekad rozwoju programu SYNSUP, zarówno jeżeli chodzi o czas potrzebny na wygenerowanie odpowiedzi, jak i jakość analiz retrosyntetycznych.

2.1.4.3. Podsumowanie

Construction of such programs is a mammoth enterprise.

Twórcy SYNSUP, 1992r.¹⁴²

Przedstawione powyżej przykłady dają podstawę do wyciągnięcia wniosku, że połączenie dwóch czynników: 1) liczącej kilka tysięcy reguł reakcji bazy danych i 2) zastosowanie narzędzi ograniczających eksplozję kombinatoryczną i, co za tym idzie, skracających czas pracy programu dało w efekcie najskuteczniejszy program do analizy retrosyntetycznej w swoich czasach. Warto odnotować, że Sumitomo Chemical Company było zainteresowane również innymi programami wspierającymi analizę chemiczną. W 1990 roku dla kilku związków testowych porównano¹³⁸ SYNSUP i LHASA, a w 1998 roku SYNSUP i SYNCHEM. W ocenie Sumitomo oba testy wykazały, że SYNSUP jest nie gorszy niż konkurenci. Niestety nie opublikowano wyników tego porównania ani zaproponowanych przez programy analiz retrosyntetycznych. Wyniki porównania można tłumaczyć tym, że w latach 90-tych SYNSUP dopiero początkował, natomiast LHASA i SYNCHEM były rozwijane około 15 lat dłużej od niego. Wyników takich jakie dawał SYNSUP w latach 2006-2009, LHASA ani SYNCHEM nigdy nie osiągnęły.

W związku z rozpoczęciem projektu CCP1 (Computer Chemistry Program 1) Sumitomo od 1990 roku brało też udział w rozwoju programu AIPHOS¹³⁵. Ponadto firma do różnych celów wykorzystywała inne programy do analizy chemicznej, takie jak EROS i jego następca WODCA, oceniono jednak, że nie nadają się dobrze do analizy retrosyntetycznej¹³⁸. CAMEO wykorzystywano do przewidywania produktów reakcji i pKa, testowano również CHIRON, bazy REACCS i ISIS¹³⁸. Tak różnorodne rozeznanie w dziedzinie, pozwoliło na wyciąganie wniosków z błędów popełnionych przy tworzeniu innych programów.

SYNSUP był nie tylko projektem czysto naukowym – rozwój we współpracy z przemysłem zmuszał twórców do wdrożenia programu do użycia. W firmie Sumitomo do SYNSUP miało dostęp 2000 pracowników, każdy mógł wysłać zapytanie mailem i mailem otrzymywał też wyniki analizy. Rocznie takich zapytań było około 700. Ile z tych pomysłów podsuniętych przez SYNSUP zostało wykorzystanych, dziś się już nie dowiemy.

Pod koniec pierwszej dekady XXI w. SYNSUP wykorzystano do analizy popularności wykorzystania reagentów i związków budulcowych w syntezach różnych leków na przykładzie grupy antydepresantów¹³⁹. Z punktu widzenia chemii procesowej, obniżenie kosztu używanych odczynników i związków budulcowych jest wartym zainteresowania problemem. Rezultaty okazały się dosyć trywialne, bowiem najpopularniejsze okazało się wykorzystanie tlenku węgla, metanolu,

bromoetanu i formaldehydu¹³⁸. Niemniej jest to przykład wykorzystania programu do analizy chemicznej w innym niż pierwotnie założony – celu.

Drugim ważnym uzupełnieniem programu było stworzenie modułu rozpoznającego konflikty reaktywności między grupami funkcyjnymi a grupami i warunkami oraz zaimplementowanie go do ekstrakcji tych informacji z baz danych. W oparciu o to stworzono tabelę włączoną do SYNSUP¹⁴¹. Autorzy podkreślają wyższość i wszechstronność swojego podejścia nad ręcznie tworzoną tabelą z LHASA, która nie była aktualizowana i jak wiemy zawierała arbitralne rozstrzygnięcia o braku konfliktu reaktywności, pomimo braku podstawy eksperymentalnej do takich twierdzeń. Podejście twórców SYNSUP pozwalało automatycznie aktualizować i poszerzać tabelę w oparciu o ukazujące się w literaturze dane eksperymentalne. Było to uzasadnione potrzebą i wartościowe uzupełnienie programu. Jednak nie zdołano już wykorzystać go do poprawy generowania analiz retrosyntetycznych. Po 2010 roku brak już doniesień o SYNSUP.

Można odnieść wrażenie, że SYNSUP zszedł ze sceny w momencie swojego szczytu możliwości. Przypuszczam, że złożyło się na to kilka przyczyn. Malcolm Behrson – *spiritus movens* projektu, pracujący nad nim od samego początku w 1971 roku, w 2007 roku zakończył współpracę z Sumitomo¹³⁸ – według oświadczenia firmy od tej pory Sumitomo i Behrson dalej osobno rozwijają swoje wersje programu. Behrson w tym momencie miał 82 lata (zmarł w 2014 roku w wieku 89 lat). Przypuszczalnie więc, był osobą najlepiej wiedzącą jak SYNSUP działa i jak można wprowadzać doń zmiany. Bez jego udziału projekt trudno było kontynuować. Drugą przyczyną były zastrzeżenia użytkowników z Sumitomo¹³⁸ – zgłaszali oni, że SYNSUP proponuje nieodpowiednie reakcje na ścieżkach, generuje dużo podobnych ścieżek (konsekwencja podejścia *depth-first*) i nadal brakuje mu ważnych reakcji. Również autorzy jednej z ostatnich prac z 2007 roku¹³⁹, sygnalizowali swoje wątpliwości co do poprawności chemicznej generowanych planów. Wobec niedostatecznie dużej bazy danych o reakcjach, autorzy rozważali użycie¹³⁸ narzędzi do automatycznego generowania baz reguł reakcji z programów ARChem Route Designer, SYNCHEM czy AIPHOS, jednak obecność wielu powtarzających się reakcji w bazach i błędów w oparciu o które będą tworzone reguły dla nieistniejących reakcji, powstrzymywała twórców SYNSUP przed tym krokiem. Ta przeczność i dbałość o jakość reguł reakcji w bazie danych, zamiast pogoni za ich ilością, zasługuje na najwyższe docenienie. Niemniej mówiąc o regułach reakcji, należy zaznaczyć, że wybrany język zapisu bynajmniej nie ułatwiał ich tworzenia, a uzupełnienie programu o bardziej wyszukane reakcje, wymagałoby najpewniej zmiany języka zapisu, a co za tym idzie przepisania całej istniejącej bazy danych, co po ponad dwóch dekadach rozwoju oznaczałoby zaczęcie niemal od początku. W podobnej ślepej uliczce znalazła się LHASA w latach 90-tych.

Jako ostatnią przyczynę można wskazać rozwój SYNSUP pod auspicjami przemysłu. Jeżeli weźmie się pod uwagę zastój w gospodarce japońskiej¹⁴⁹⁻¹⁵⁰ trwający od lat 90-tych, to decyzja zarządu Sumitomo o zamknięciu projektu rozwijanego od 26 lat i w dalszym ciągu nie generującego lepszych planów syntetycznych niż te już znane¹³⁸ – staje się zrozumiała.

Reasumując, prześledzenie historii SYNSUP dawało z jednej strony nadzieję, że możliwe jest poradzenie sobie z eksplozją kombinatoryczną jeżeli tylko dostatecznie umiejętnie będzie się poruszać po drzewie retrosyntetycznym. Z drugiej strony stanowiło ostrzeżenie, że ponad dwie dekady ręcznego tworzenia reguł reakcji – to za mało. Takie ostrzeżenie mogło działać deprymująco na następców próbujących zmierzyć się z tym wyzwaniem.

2.1.5. AIPHOS/KOSP

2.1.5.1. Wprowadzenie

AIPHOS/KOSP – były to dwa japońskie programy, ściśle ze sobą związane. AIPHOS (Artificial Intelligence for Planning and Handling Organic Synthesis) był projektem szerszym, zawierającym inne programy (np. KOSP - Knowledge base-Oriented system for Synthesis Planning¹⁵¹) programy należące do tzw. rodziny programów AIPHOS. Rozwijany¹⁵²⁻¹⁵³ był pierwotnie przez prof. Kimito Funatsu na Uniwersytecie Technicznym w Toyohashi w późnych latach 80-tych, następnie został skomercjalizowany i jeszcze w listopadzie 2010 roku był dostępny na stronie Fujitsu Limited¹⁵⁴. Program miał spełniać dwie funkcje: 1) generowania wszystkich możliwych prekursorów dla zadanego związku docelowego, 2) przewidywania produktów i ścieżek reakcji¹⁵². Podobnie jak LHASA, AIPHOS działał interaktywnie^{151,155}, generując zestaw prekursorów dla zadanego celu syntetycznego i pozostawiając użytkownikowi wybór następnego związku startowego. Oprócz struktury prekursora AIPHOS proponował dla każdego zaproponowanego przejścia warunki reakcji¹⁵².

W przeciwieństwie do LHASA, AIPHOS dysponował stworzoną automatycznie bazą danych o transformacjach¹⁵⁶. Uzyskano ją przez identyfikację atomów, którymi różnią się substrat i produkt, definiując tym samym rdzeń reakcji^{11,151}. Następnie podobne rdzenie reakcyjne, połączono w klastry¹⁵⁶ będące podstawą transformacji w bazie. Jako różnicę między AIPHOS i KOSP twórcy wskazują¹⁵¹ jakość analiz retrosyntetycznych – o ile te z AIPHOS miały być bardziej abstrakcyjne, to KOSP miał oferować bardziej empiryczne i użyteczne propozycje. W praktyce jednak, z powodu małej ilości zaprezentowanych przykładów działania obu modułów, trudno dostrzec między nimi jakościową różnicę.

AIPHOS zawierał dwie bazy danych RKB (Reaction Knowledge Base) i LGKB (Leaving Group Knowledge Base)¹⁵¹. RKB była tak naprawdę bazą RDB pozyskaną z programu SYNLIB i uzupełnioną o mapowanie atomów¹⁵¹. W RKB warunki reakcji zostały opisane dwustopniowo, na pierwszym stopniu znajdowało się 16 (A-Q) warunków głównych, na drugim stopniu, pod każdą literą, kryła się lista warunków szczegółowych. I tak przykładowo warunki B – były to warunki zasadowe, na drugim stopniu bardziej szczegółowo podzielone na 9 warunków konkretnych jak LDA czy *t*BuOK. Warunki z drugiego stopnia pochodziły¹⁵¹ z tabel podręcznika Greene¹²², które przypomnijmy, stworzono na potrzeby LHASA. Jednak informacja o warunkach nie pociągała za sobą propozycji zastosowania grup zabezpieczających¹⁵⁷ i aż do końca rozwoju programu nie dodano tej funkcjonalności. Wprawdzie twórcy już w pierwszych pracach zaznaczają¹⁵², że dla programu istotny jest dobór warunków tak, by po rozpoznaniu konfliktu reaktywności zaproponować warunki ten konflikt omijające, jednak poza deklaracją brak jest konkretnych przykładów działania tego modułu.

Do zapisu struktury i przetwarzania informacji chemicznych, AIPHOS wykorzystywał tabelę połączeń¹⁵³. Po wprowadzeniu przez użytkownika struktury celu syntetycznego w postaci graficznej, program generował kanoniczną tabelę wiązań między atomami i dalej w takiej postaci przetwarzał „cząsteczkę”. Jest to o tyle ważne, że tabele połączeń i zapisy macierzowe są bardziej czasochłonne w przetwarzaniu od zapisów liniowych, co stanowi immanentne ograniczenie programu.

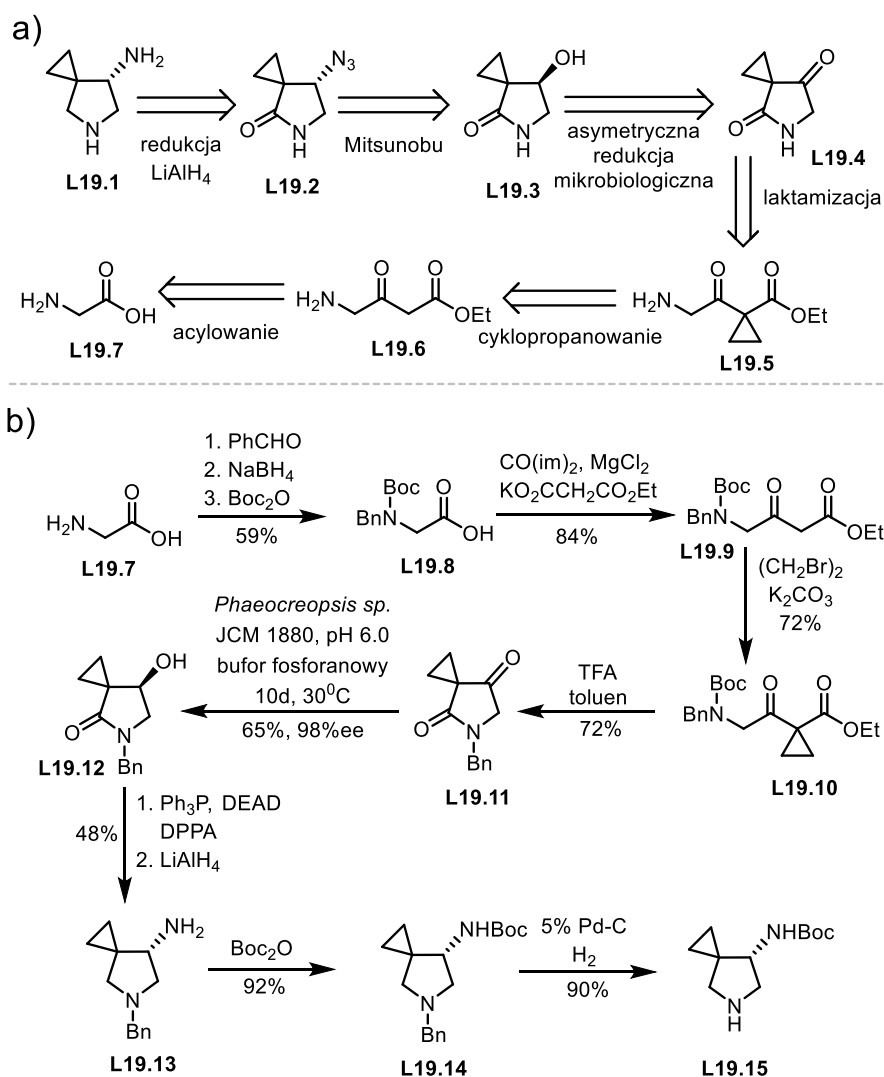
AIPHOS posiadał elementy percepcji struktury, wydzielono 111 fragmentów strukturalnych¹⁵⁵, takich jak karbocykliczne pierścienie 3-6 członowe, atomy węgla o różnej rzędowości itp. Ponadto rozpoznawał 40 grup funkcyjnych, podzielonych na 9 kategorii¹⁵⁵.

Program miał proponować istotne cięcia retrosyntetycznie istotne. Twórcy powołują się przy tym na inspirację, którą była zdefiniowana przez Corey'a strategia topologiczna¹⁵⁵. Po wytypowaniu

miejsca strategicznego program dokonywał jego wirtualnego cięcia. Użytkownik również mógł wytypować miejsce cięcia, o ile AIPHOS nie rozpoznał go jako potencjalnie interesujące¹⁵⁵. Na skutek wirtualnego cięcia powstawały „luźne końce”, którym, by odpowiadały realnym cząsteczkom, program dopisywał grupy opuszczające z bazy LGKB¹⁵⁵. Tak uzyskana cząsteczka była następnie sprawdzana na obecność fragmentów strukturalnie niestabilnych, takich jak nadtlutki¹⁵², zapisanych na załączonej do programu liście. W przypadku ich znalezienia, taki prekursor nie był proponowany. Analizy kończyły się na materiałach startowych, podzielonych na 3 klasy: aromatyczna, niearomatyczne i chiralne. Pochodziły one z bazy ACD (Available Chemicals Directory), zawierającej około 130 000 struktur, przeważnie o szkieletach 3-10 węglowych.

2.1.5.2. Rezultaty

Istnieje zaledwie kilka opublikowanych przykładów działania AIPHOS/KOSP. Już w pierwszej pracy z 1988 roku¹⁵² znajduje się analiza retrosyntetyczna chloropromazyny. Pokazuje ona niedostatki jakie u swych początków cechowały AIPHOS. O ile identyfikacja miejsc cięcia



Ryc.19. a) Ilustracja analizy retrosyntetycznej dla fragmentu DU-6859a **L19.1** zaproponowanego przez program KOSP i ogłoszonego w 1998 roku¹⁵⁷, b) walidacja syntetyczna planu wygenerowanego przez KOSP dla **L19.1**.

została zaproponowana w sposób chemicznie zasadny, to jednak zaproponowane prekursory budzą wątpliwości co do wykonalności syntetycznej.

Najbardziej zaawansowanym przykładem walidacji działania AIPHOS, jest synteza fragmentu związku DU-6859a¹⁵⁷. W przypadku omówionych wcześniej programów, rzadko spotykaliśmy walidację poprzez próbę syntezy według komputerowo-generowanego planu, a co dopiero dla związku wymagającego syntezy asymetrycznej. Dlatego tym bardziej przykład ten zasługuje na uwagę. Analiza retrosyntetyczna (Ryc.19a) zakłada uzyskanie pierwszorzędowej grupy aminowej w **L19.1**, z alkoholu **L19.3** przez azydek **L19.2**. Przejście to wiąże się z inwersją konfiguracji, co program uwzględnia. Alkohol **L19.3** proponowano uzyskać na drodze mikrobiologicznej redukcji ketonu **L19.4** za pomocą drożdży piekarskich. Faktycznie redukcji drożdżami nie udało się zrealizować, a ostateczny dobór warunków pozostawiono chemikom.

Nietrudno dostrzec, że ilość faktycznie wykonanych etapów syntetycznych (Ryc.19b) jest większa od tych zawartych w planie. Wynika to z braku uwzględniania grup ochronnych przez AIPHOS, toteż dobór grup pozostawiono adaptatorom. W całym planie kluczowy jest etap wprowadzenia centrum asymetrii. W celu wykonania redukcji mikrobiologicznej przetestowano około 400 mikroorganizmów (bakterii, drożdży, pleśni), redukcja ketonu nastąpiła tylko w przypadku 4 szczepów. Sprawdzenie powstającego centrum stereogenicznego wykluczyło kolejne 2 szczepy, generujące centrum odwrotne od żądanego. Po optymalizacji z pierwotnych 8% wydajności, uzyskano 65% (98%*ee*). Jednak konieczność prowadzenia reakcji przez 10 dni może budzić zastrzeżenia, co do aplikowalności tego podejścia w większej skali. Jako walidacja przykład ten jest wyjątkowy na tle ówczesnych, ze względu na zastosowanie reakcji asymetrycznych. Z drugiej jednak strony, w porównaniu z wcześniejszym patentem¹⁵⁸ z 1996 roku, nie przedstawia atrakcyjniejszej alternatywy.

Pozostałe przykłady pracy programu, nie osiągają już poziomu poprzedniego przykładu. W pracy z 1999 roku¹⁵¹ przedstawiony jest plan syntezy taksotominy C, powtarzający syntezę znaną z literatury i jako taki może być potraktowany jako walidacja poprawności. Jednak znów, plany te różnią się brakiem grup ochronnych.

Ostatnie prace o AIPHOS/KOSP z 2007 roku¹⁵⁹ i 2010 roku¹⁵⁷ wykorzystują przewidywanie prekursorów do jednoetapowych syntez α,β -nienasyconych estrów i chromanonów. Następnie, zaproponowane reakcje były przedmiotem analiz za pomocą DFT, mających na celu wytypować te o największej szansie powodzenia. Testowanie *in silico* miało ograniczyć niepotrzebne syntezy w nierokujących warunkach. Dla jednoetapowych syntez podejście to może być zasadne, jednak próba zastosowania go do wieloetapowych syntez i oceny rozbudowanych drzew retrosyntetycznych obecnie nie jest jeszcze wykonalna, chcąc uzyskiwać rezultaty w czasach rzędu minut.

2.1.5.3. Podsumowanie

Ostatnie prace o AIPHOS/KOSP, pokazują, że twórcy szukali dla programu zastosowań innych niż analiza retrosyntetyczna, w której pomimo ponad 20 lat rozwoju, nie osiągnął on zadowalających rezultatów. Jednym z planów ulepszenia KOSP¹⁵¹ było wprowadzenie nieinteraktywnego generowania sekwencji retrosyntetycznych, ale to wymagałoby narzędzi selekcyjnych zmniejszających rozmiary drzew analiz¹⁶⁰. Celu tego nigdy nie osiągnięto.

Przyczyn niepowodzenia projektu AIPHOS można upatrywać przede wszystkim w zbyt ograniczonej percepcji struktury. W latach 2007-2010 AIPHOS/KOSP był komercyjnie dostępny w Fujitsu Limited. Obecnie ten znak handlowy ma status: porzucony¹⁶¹.

2.1.6. Inne programy

Omówione dotychczas programy LHASA, SYNCHEM, SYNSUP i AIPHOS/KOSP – stanowiły najlepsze, w mojej ocenie, osiągnięcia pierwszego okresu prac nad programami do komputerowo wspomaganą analizy retrosyntetycznej. Rezultaty ich pracy nadają się do dyskusji poprawności chemicznej. Nie należy jednak zaniedbywać proporcji i zapominać o głównym celu, jakim miało być generowanie planów syntetycznych lepszych od istniejących i wykonywalnych w laboratorium. Do etapu walidacji laboratoryjnej dochodziło jedynie incydentalnie, co pasuje te programy w kategorii ciekawych prób, lecz jeszcze nie pozwala ich nazwać użytecznymi dla chemika narzędziami.

Równolegle do nich, rozwijanych było szereg innych programów. Nie odniosły one jednak takiego sukcesu i stopnia zaawansowania, toteż ze względu na ograniczone ramy niniejszej rozprawy, odnotuję jedynie istnienie pozostałych programów, takich jak: CHIRON¹⁶²⁻¹⁶⁴, COMPASS¹⁶⁵⁻¹⁶⁷, FLAMINGOES¹⁶⁸⁻¹⁷⁰, LILITH¹⁷¹⁻¹⁷³, RETROSYN¹⁷⁴, SECS¹⁷⁵⁻¹⁷⁷, SESAM¹⁷⁸, SOPHIA¹⁷⁹⁻¹⁸⁰, SOS¹⁸¹⁻¹⁸², SST¹⁸³, SYNGEN¹⁸⁴⁻¹⁸⁸, SYNLIB¹⁸⁹, SYNLMA¹⁹⁰, TOSCA¹⁹¹, TRESOR¹⁹², WODCA¹⁹³⁻¹⁹⁴, WTDWTR¹⁹⁵. Jako, że nie zawierały one interesujących chemicznie przykładów walidacji, nie będą omawiane.

Słowo o funkcji programów. Obiektem mojego zainteresowania i tematem tej pracy, jest komputerowe wspomaganie syntezy chemicznej poprzez automatyczne generowanie planów syntetycznych lub choćby analiz retrosyntetycznych. Tak przyjęty temat, nie mieści w sobie programów spełniających inne funkcje, takie jak:

- 1) przewidywanie produktu reakcji chemicznej przy zadanym substracie, lub substracie i warunkach reakcji – taką funkcję pełniły programy: CAMEO¹⁹⁶, EROS¹⁹⁷⁻²⁰⁰, ROBIA²⁰¹⁻²⁰²,
- 2) przewidywanie nowych reakcji chemicznych: ARGENT²⁰³, IGOR²⁰⁴, IGOR2²⁰⁵⁻²⁰⁶, SYMBEQ²⁰⁷,
- 3) analiza mechanistyczna reakcji chemicznych: ICAR²⁰⁸⁻²⁰⁹, RAIN²¹⁰.

Powyżej wymienione programy, tradycyjnie omawiane są^{11,44,156,194,205,211-220} wspólnie z tymi przeznaczonymi do planowania syntezy, tu jednak z przyczyn podanych powyżej, nie zostaną potraktowane wspólnie.

Na granicy wspomaganą analizy retrosyntetycznej leży proponowanie substratu dla zadanego związku, jednak gdy substrat ten generowany jest nie w oparciu o regułę reakcji lecz o (często pozbawiony grup funkcyjnych) układ wiązań szkieletowych. Rezultaty takich analiz mają ograniczoną użyteczność syntetyczną, gdyż poza zaproponowaniem rdzenia cząsteczki, pozostała praca analityczna w dalszym ciągu pozostaje po stronie chemika. Takie cechy miały analizy generowane przez programy SESAM czy CHIRON.

Pomimo dekad prób, stworzenie w XX wieku programu komputerowego dokonującego analizy retrosyntetycznej i tworzącego wartościowe plany syntetyczne zakończyło się klęską. Problem okazał się trudniejszy niż pierwotnie zakładano, zyskując sobie miano jednego ze Świętych Graali chemii organicznej²²¹. Komercyjne wersje programów z tego pierwszego okresu¹³⁵ (WODCA, ChemSpire, Arthur, AIPHOS/KOSP) są już niedostępne, a strony internetowe projektów zostały zamknięte.

2.2. Współczesne programy komputerowe wspomagające planowanie wieloetapowych syntez chemicznych

2.2.1. Wstęp

Wprowadzony podział na programy współczesne i historyczne, jest czysto arbitralny, trudno bowiem wskazać w dziedzinie programów komputerowych do analizy retrosyntetycznej jasną cezurę. Uważam, że daje się ją jednak pośrednio wyczytać z rozwoju dziedziny. Do końca pierwszej dekady XXI wieku, wieloletnie i zakrojone na szeroką skalę projekty stworzenia takiego programu, zostały po cichu zakończone. To znamienne, że prace nad LHASA spektakularnie ogłaszane w latach 60-tych, do końca lat 90-tych wytraciły entuzjazm, a prac – choć oficjalnie nigdy nie zarzucono – to zaprzestano ich publikacji, podobny los spotkał SYNCHEM. SYNSUP, który swymi początkami sięga lat 70-tych. Pomimo dotrwania do 2010 roku, ze względu na jego genealogię, SYNSUP należy nazwać programem najbardziej zaawansowanym, lecz należącym do poprzedniej epoki. Kolejnym kryterium ułatwiającym uporządkowanie dorobku w omawianej dziedzinie, jest stan prac nad poszczególnymi programami. Do historycznej kategorii zaliczyłem programy nierozwijane już od lat i o małych szansach na wznowienie, mających dziś już wyłącznie status wzmianki w artykułach przeglądowych. Tymczasem w niniejszym rozdziale o programach współczesnych, pomieściłem przykłady projektów z ostatniej dekady, po których można się spodziewać dalszych doniesień literaturowych.

Po klęskach wielkich projektów na początku XXI wieku, w dziedzinie zapanował marazm^{26,218}. Problem komputerowo wspomaganą analizy retrosyntetycznej wydawał się niemożliwy do rozwiązania za pomocą dotychczas stosowanych podejść. Ręczne tworzenie bazy danych o regułach reakcji było zbyt mozolnym, wieloletnim przedsięwzięciem, w którym, jak pokazały przypadki LHASA i SYNSUP, nawet wielotysięczne bazy danych nie były wystarczające, by proponować plany syntetyczne konkurencyjne ze zdolnościami chemików. Zatem jak wiele reguł reakcji potrzeba było wprowadzić do takiej bazy, by zacząć generować atrakcyjne chemiczne plany? Na to pytanie nie posiadano ówczesnie odpowiedzi. I w tym zawierało się ryzyko, którego nie chciano już podejmować, tłumacząc: „Skoro nawet noblista, E. J. Corey przez 30 lat, nie rozwiązał tego problemu, to jest on albo zbyt trudny albo niemożliwy do rozwiązania”.

Skoro ręczne tworzenie bazy danych okazało się podejściem zbyt żmudnym i nie przynoszącym zamierzonych efektów, próbowano skonstruować takie reguły w sposób automatyczny²²²⁻²²³. Pokusa wykorzystania już istniejących baz danych i wyekstrahowania z nich wiedzy chemicznej sięga swymi początkami programów SYNCHEM i AIPHOS. Jednak okazuje się, że i to podejście z kilku powodów nie spełnia pokładanych w nim oczekiwań. Po pierwsze, tak wygenerowane reguły dobrze radzą sobie z reakcjami typu S_N2 , estryfikacji lub syntezy specyficznych pierścieni heterocyklicznych, lecz nie sprawdzają się przy bardziej skomplikowanych reakcjach o mniejszej liczbie precedensów, gdzie algorytm nie ma kompetencji i dostatecznej liczby przykładów by ocenić które fragmenty rdzenia są istotne dla zajścia reakcji, a które nie mają na reakcję decydującego wpływu. Stanowi to swoisty szklany sufit dla programów o automatycznie generowanych bazach reguł. Z jednej strony program przecież działa, generuje plany syntetyczne, więc – wyciąga się wniosek – przyjęta zasada działania musi być poprawna. Tylko, że obecnie, żaden z tak skonstruowanych programów nie przebił wspomnianego szklanego sufitu i nie był w stanie w sposób automatyczny stworzyć bazy reguł skomplikowanych reakcji²¹⁸.

Rozdział niniejszy poświęcam ostatnim osiągnięciom w dziedzinie komputerowego wspomaganie syntezy chemicznej.

2.2.2. Reaxys Synthesis planner & SciFinder SciPlanner

Dwa najbardziej popularne programy wśród praktykujących syntezę chemików, ze względu na daleko idące podobieństwa, potraktuję łącznie. Zarówno Reaxys jak i SciFinder, dysponują programami wspomagającymi planowanie syntezy²²⁴⁻²²⁵. Ich działanie polega na znalezieniu, wśród posiadanych precedensów literaturowych, sekwencji znanych reakcji układających się w ścieżkę syntezy dla zadanego produktu. Posługiwanie się taką bazą danych posiada zarówno zalety jak i wady.

Do niewątpliwych zalet tego podejścia należy ugruntowana pewność o poprawności treści precedensów, odpowiedzialność za nią sędowana została przez twórców bazy na publikacje źródłowe. Jednak pomiędzy treścią publikacji czy patentu a reakcją z bazy danych pojawia się margines mogący być źródłem błędu. Oczywiście ilość ogłaszanych codziennie doniesień literaturowych z samej tylko chemii organicznej wyklucza ręczne przepisywanie treści publikacji na format bazy danych. Ze względu na ilość pracy służą do tego algorytmy (*data mining*) – i tu tkwi jedno ze źródeł błędów w bazie danych. Innym mogą być błędy w samych publikacjach, które nie zostały znalezione w procedurze recenzji.

Do wad podejścia bazującego na literaturze zaliczyć należy wyłączenie opierania się na literaturze. Stanowi to oczywiste ograniczenie – na przykład jeżeli poszukujemy planu dla pochodnej związku naturalnego, który występuje tylko jako ester metylowy – to taki plan zgodnie z zasadą znajdziemy (o ile związek ten został wcześniej co najmniej raz syntetycznie uzyskany). Natomiast drobna zmiana estru z metylowego na etylowy, całe przedsięwzięcie uniemożliwi, ponieważ pochodnej etylowej w bazie nie będzie. Z chemicznego punktu widzenia jest to drobna różnica. Jednak to co dla chemika jest drobną zmianą dla Reaxys Synthesis Planner i SciPlanner jest barierą nie do pokonania. Wynika to z nieoperowania przez te programy wyabstrahowanym pojęciem reguły reakcji/transformacji, a jedynie precedensami. Z tego samego źródła poprawność chemiczną czerpała pierwsza wersja Chematici²²⁶⁻²²⁹, będąca modulem sieciowym łączącym precedensy literaturowe w sieć i na niej operująca.

Jak Reaxys Synthesis Planner i SciPlanner radzą sobie z błędami w ich bazach danych? Oczyszczanie bazy danych z błędów chemicznych (bez względu na źródło ich pochodzenia), poprawiające jej jakość, winno stanowić jeden z priorytetów dla zarządzających bazą. Tak się jednak nie dzieje. W przypadku Reaxys²³⁰, jedną pobudką do usuwania błędów są zgłoszenia użytkowników. Sam Elsevier w systemowy sposób nie poprawia jakości bazy Reaxys. Prawdopodobną przyczynę takiego podejścia (oprócz braku presji rynku) omawiam w punkcie 2.2.6..

Podsumowując, programy planujące syntezę w Reaxys i SciFinder ściśle zależą od precedensów literaturowych. W celu powtórzenia istniejących syntez i uzyskiwania związków znanych (czasem na alternatywnej drodze), dobrze się sprawdzają. Jednak z założenia nie mogą służyć syntezie nowych prototypowych związków ani planować syntezy związków znanych, których synteza miałaby przebiegać z wykorzystaniem związków nieznanych w literaturze.

2.2.3. ARChem

ARChem²³¹ jest przykładem współczesnego, komercyjnie dostępnego programu służącego do generowania planów retrosyntetycznych. Posługuje się on wyabstrahowanym pojęciem reguły reakcji. W przeciwieństwie do poprzednich, eksperckich podejść jego baza reguł reakcji była tworzona na drodze ekstrakcji z istniejących baz precedensów literaturowych. Przykłady takich ekstrakcji przeprowadzono na bazach Beilstein, Accelrys, Reaxys, CIRX²³².

Tworzenie reguły reakcji odbywa się w tym programie na 5 etapach: 1) identyfikacja rdzenia reakcji, czyli grupy atomów które zmieniają wiązania między substratem, a produktem; 2) rozszerzenie rdzenia na znaczące sąsiednie atomy; 3) łączenie rozszerzonych rdzeni w klastry; 4) generowanie uogólnionej reguły reakcji przez każdy klaster; 5) przekształcenie uogólnionej reguły reakcji w kompletną regułę reakcji. Ważnym kryterium jest liczność rozszerzonych rdzeni w klastrze, bowiem im ich więcej, tym jakość reguły będzie większa. W testach ekstrakcji próg liczności rdzeni w klastrze wynosił od 2 do 5. W oparciu o wyekstrahowane reguły tworzone są plany syntetyczne – w ramach ich walidacji, stworzono za pomocą ARChem plan dla Zatoosteronu, był on podobny do planu znanego w literaturze. Plany kończyły się na związkach komercyjnie dostępnych w Aldrich, Lancaster i Acros, a baza związków końcowych liczyła około 120 000 molekuł.

Niestety program ten nie posiadał percepcji stereochemii, regiochemii, ani konfliktów reaktywności między grupami funkcyjnymi, więc jego użyteczność w planowaniu syntezy jest ograniczona. W 2014 roku został zakupiony przez Wiley²³³⁻²³⁴, od tego czasu brak dalszych doniesień o tym programie.

2.2.4. ICSynth

Programem bardzo podobnym do ARChem jest ICSynth²³⁵. Rozwijany od 2005 roku, również wykorzystuje wyabstrahowane reguły reakcji. W jego przypadku pochodzą one z baz danych SPRESI i SOS²³². Program nie rozpoznawał konfliktów reaktywności, a percepcję elementów stereogenicznych sami autorzy określali jako wymagającą usprawnienia. Walidację programu przeprowadzono we współpracy z chemikami z firmy AstraZeneca. Program miał spełniać rolę generatora pomysłów retrosyntetycznych. Jeden z tych pomysłów został wykonany syntetycznie w postaci czteroetapowej syntezy oksaspiroketonu – podjednostki budulcowej w serii prototypowych leków na Alzheimera. Po 2015 roku brak doniesień o tym programie.

2.2.5. Sieć neuronowa Seglera-Wallera

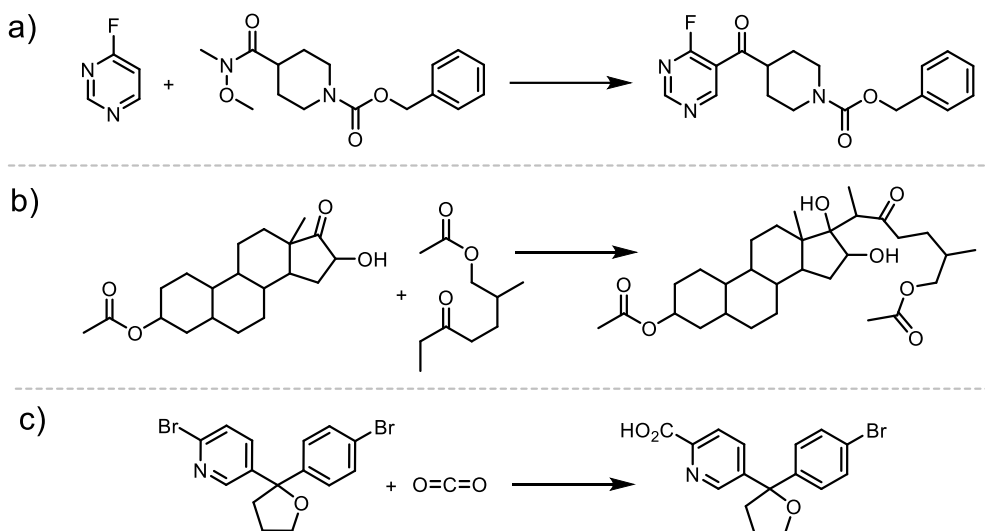
W 2018 roku, w *Nature*²³⁶, ukazała się praca opisująca wykorzystanie sieci neuronowych do analizy retrosyntetycznej. Reguły reakcji zostały automatycznie wyekstrahowane z bazy Reaxys. Wytypowano wszystkie reakcje jednoetapowe opublikowane do 2015 roku, tak utworzono zbiór 12,4 miliona precedensów reakcji. Następnie zidentyfikowano w nich rdzenie reakcyjne poprzez rozpoznanie atomów zmieniających wiązania między substratem a produktem. Rdzenie reprezentujące jednakowe reakcje zebrano w zbiory, uzyskując dwa zbiory reguł reakcji. *Rollout Set* zawierał reguły reakcji stworzone w oparciu o zbiory precedensów liczące ponad 50 przykładów reakcji, taki wymóg spełniło 17 134 reguł. Drugi, ogólniejszy, *Expansion Set* stworzono w oparciu o zbiory liczące trzy i więcej precedensów, uzyskując zbiór 301 671 reguł.

Posługując się tak stworzonymi regułami reakcji, zestawiono trzy różne sieci neuronowe (N) z metodą Monte Carlo zastosowaną do drzew retrosyntetycznych (MCTS), uzyskując 3N-MCTS. Narzędzie to pozwalało przy podaniu struktury celu syntetycznego, wygenerować dla niej analizę retrosyntetyczną. Poprawność działania sprawdzono na zbiorze 301 671 reakcji, sieci przewidywały dla nich produkt reakcji dla zadanych substratów – poprawną predykcję osiągnięto w 31% przypadków. Gdy wzięto pod uwagę pierwsze 10 najlepszych wyników, szansa, że pośród nich znajdzie się poprawny produkt reakcji wzrosła do 63%. Dalsze zwiększenie zbioru wyników do 50 najlepszych, podniosło tę szansę do 72%.

Kolejną metodą sprawdzenia sieci Seglera-Wallera, było wygenerowanie przez nią analiz retrosyntetycznych dla 9 związków chemicznych, których syntezy nie były znane przed 2015 rokiem (podejście to miało wykluczyć możliwość, że znajdowały się one w zbiorze uczącym), a których

syntezy opublikowano później. Zestawiono obok siebie te dwie ścieżki syntetyczne (równej długości), a następnie posługując się podwójnie ślepą próbą przedstawiono je grupie 45 doktorantów z Chin i Niemiec, by je ocenili i wybrali lepszą. Test ten wykazał, że grupa oceniająca nie dostrzegła dużej różnicy między syntezami istniejącymi, a stworzonymi automatycznie ścieżkami. Jak dotąd jest to najlepszy rezultat uzyskany przez sieć neuronową Seglera-Wallera²³⁷.

Omawiane podejście jest kolejnym bazującym na pojęciu wyabstrahowanej reguły reakcji chemicznej. Warto zaznaczyć, że abstrahowanie reguły opiera się wyłącznie na kryteriach strukturalnych (zmiana w rdzeniu reakcyjnym), natomiast informacje o cechach stereochemicznych nie są zapisywane i nie wchodzi w skład reguły. Podobnie informacja o warunkach reakcji jest tracona w tym procesie. Skutkuje to tym, że na przykład reguła reakcji estryfikacji jest tworzona w oparciu o precedensy estryfikacji Fischera, Steglich, Mitsunobu i inne – wszak zmianie ulegają zawsze te same atomy rdzenia. Dlatego deklarowana przez twórców możliwość rozpoznawania kontekstu chemicznego przez identyfikację w obecności jakich grup funkcyjnych reguła reakcji jest stosowana, a jakie są z nią chemicznie niekompatybilne – nie może działać poprawnie. Twórcy sami zwracają uwagę, że zastosowane podejście byłoby trudne do zastosowania przy analizie retrosyntetycznej związków naturalnych. W ich przypadku ignorowanie centrów stereogenicznych jak i brak percepcji efektów sterycznych nie jest już możliwy.



Ryc.20. Przykłady reakcji zaproponowanych przez sieć neuronową Seglera-Wallera²³⁶.

Na Ryc.20. przedstawiono trzy przykłady reakcji chemicznych proponowanych przez sieć Seglera-Wallera i wybrane ze ścieżek poddawanych walidacji. Omówienie tych przypadków pozwoli ocenić na ile poprawnie chemicznie dokonywane jest ekstrakowanie reguł, ich uogólnianie i łączenie w klastry reguł.

Reakcja przedstawiona na Ryc.20a, wygląda na przykład wykorzystanie amidu Weinreba do syntezy ketonu. Jednak wątpliwości budzić może zastosowany nukleofil. MCTS nie przedstawia w swoich ścieżkach żadnych informacji o warunkach reakcji, dlatego należy się domyślić jaka reakcja i w jakich warunkach mogłaby tu zajść. O ile partner elektrofilowy nie budzi tu wątpliwości o tyle nukleofil już tak. Czy miałby być to przykład orto-litowania? W literaturze brak jest jakichkolwiek przypadków dla podobnych pierścieni. Rodzi to pytanie, jak bardzo ogólne było w tym przypadku abstrahowanie reguły reakcji?

Przykład z Ryc.20b, ilustruje przypuszczalnie reakcję aldolową między dwoma ketonami. Pomijając brak uwzględnienia centrów stereogenicznych (co w przypadku szkieletu stereoidu trudno

zaniedbać), groźba braku selektywności wydaje się tu poważniejszym problemem. W przypadku reakcji aldolowej między dwoma ketonami, możliwe jest wcześniejsze wytypowanie, który z nich będzie elektrofilem, a który nukleofilem. Jest to możliwe dzięki wcześniejszemu utrwaleniu partnera nukleofilowego w postaci eteru silylowego enolanu. Chcąc przeprowadzić tę reakcję, należy użyć silnej zasady w celu wytworzenia anionu enolanowego, co w przypadku takiego substratu grozi wewnątrzcząsteczkową reakcją z grupą acylową. Problem ten świadczy o nie braniu przez program pod uwagę kontekstu strukturalnego. Co więcej, obie strony ketonu są sterycznie niemal nierozróżnialne, dlatego selektywna aktywacja tylko jednej ze stron może się okazać niewykonalna.

Reakcja Ryc.20c. zaproponowana przez sieć Seglera-Wallera jest przypuszczalnie reakcją metalowania bromku arylowego za pomocą magnezu i następczego tworzenia kwasu karboksylowego. Przeprowadzenie tej reakcji selektywnie na jednym bromku arylowym, w obecności drugiego może okazać się niewykonalne. Jest to kolejny przykład świadczący o ignorowaniu kontekstu strukturalnego i przypuszczalnych konfliktów reaktywności w proponowanych transformacjach.

Sieć Seglera-Wallera jest zdolna generować kilkietapowe analizy retrosyntetyczne. Niestety brak uwzględniania przez nią aspektu stereochemicznego, regioselektywności reakcji, konfliktów reaktywności, konieczności stosowania grup zabezpieczających oraz opierania się o reguły reakcji generujące transformacje budzące wątpliwości – każde uznać, że podejście to nie jest jeszcze użytecznym narzędziem wspierającym pracę koncepcyjną.

2.2.6. Podsumowanie

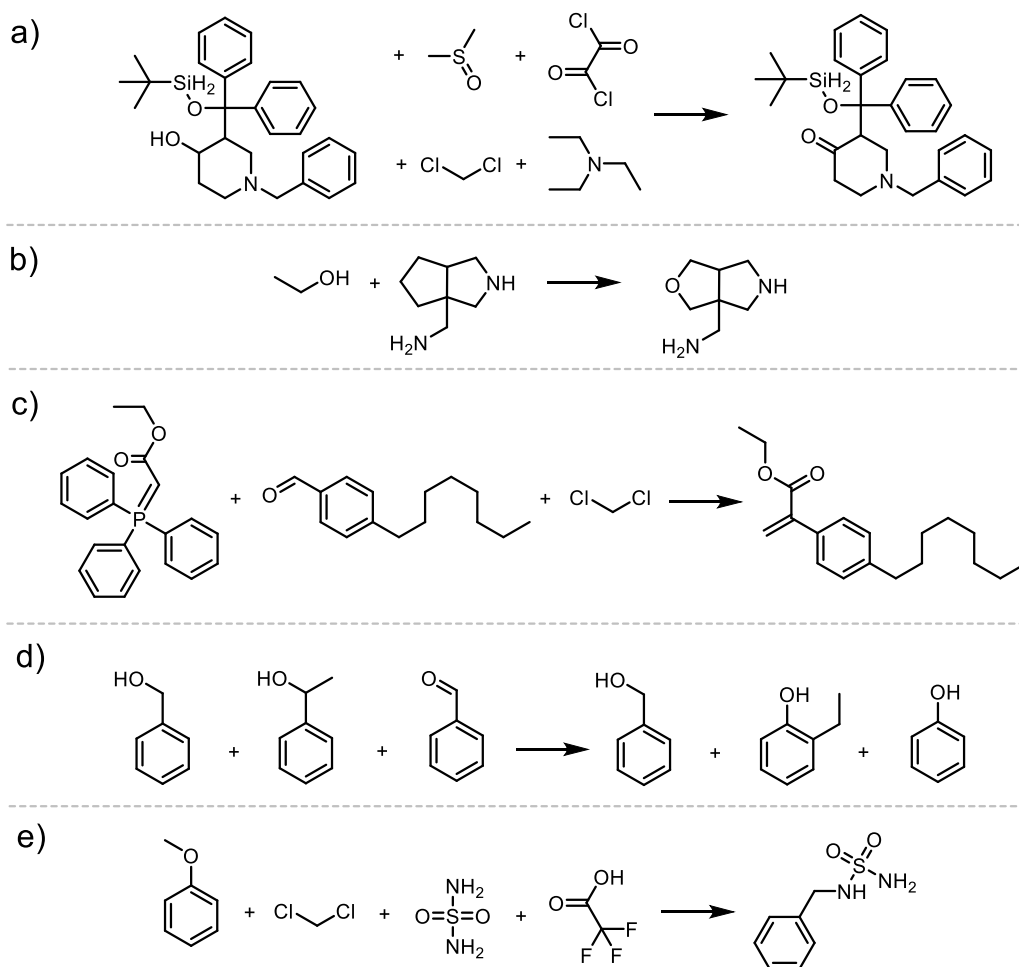
Przytoczone przykłady współczesnych prób stworzenia programu do planowania syntezy chemicznej, świadczą dobitnie o trudności problemu. Wykorzystanie współczesnych komputerów, metod sztucznej inteligencji i uczenia maszynowego w dalszym stopniu nie przyniosło przełomu w omawianej dziedzinie. Wspomaganie syntezy organicznej za ich pomocą rozwijane jest na kilku polach. Po pierwsze, podejmowane są próby przewidywania reakcji chemicznych²³⁸⁻²⁴⁶ – podejście to w dalszej perspektywie pozwoliłoby przewidywać sekwencje reakcji organicznych, a w rezultacie tworzyć całe plany syntetyczne. Zbliżonym zagadnieniem jest określanie reaktywności grup i wiązań w celu syntetycznym, dzięki czemu wskazywane jest wiązanie proponowane do retrosyntetycznego rozcięcia²⁴⁷⁻²⁵⁰. Połączenie tego typowania z regułami reakcji zdolnymi proponować transformacje chemiczne tworzące ów wiązania jest kolejną formą analizy retrosyntetycznej *in silico*. Wreszcie podjęto różne bezpośrednie próby analizy retrosyntetycznej *in silico*²⁵¹⁻²⁶², często bazujące na automatycznie generowanych regułach reakcji pozyskanych za pomocą uczenia maszynowego, bez korekty dokonywanej przez ludzkiego eksperta. Ostatnim typem wspomaganiania planowania syntezy jest ocena związków przejściowych lub reakcji stosowanych w planie syntetycznym²⁶³⁻²⁷⁰. Wygenerowane komputerowo lub znane z literatury plany syntetyczne mogą być przedmiotem oceny i hierarchizacji według różnych kryteriów, jak choćby zachowanie ekonomii redoks²⁶⁸ lub najszybszy przyrost złożoności w kolejnych etapach^{267,269}.

Jedno z wyzwań stojących przed chcącymi automatycznie uczyć maszyny chemii organicznej wymaga dokładniejszego omówienia. Chcąc automatycznie abstrahować z baz danych reguły transformacji chemicznych, napotykamy na problem błędów w tychże bazach. Dlaczego zatem i tu nie posłużyć się algorytmem, odsiewającym reakcje (zastosowania reguł) poprawne od niepoprawnych? Albo innymi słowy, jak obserwując nieznaną grę (chemię) rozpoznać gdy ktoś błędnie stosuje reguły tej gry? Sytuację taką opisuje Wittgenstein:

„w jakich przypadkach powiadamy, iż w jakąś grę gra się według pewnej określonej reguły [?] Reguła może być pomocą w nauce gry. Podaje się ją uczniowi, ćwicząc z nim

jej zastosowania. – Bądź też jest ona narzędziem w samej grze. – Albo: reguła nie ma zastosowania ani w nauce gry, ani w samej grze; nie występuje też w spisie reguł. Uczymy się grać patrząc, jak grają inni. Mówimy jednak, iż gra się tu według tych a tych reguł, ponieważ postronny obserwator mógłby je odczytać z praktyki gry – jak prawo przyrody, któremu podlegają posunięcia w tej grze. – Jak jednak w tym wypadku obserwator odróżni błąd graczy od posunięcia prawidłowego?²⁷¹ ”

W przypadku chemii organicznej, reguły reakcji chemicznych są właśnie regułami w samej grze. Czy współczesne algorytmy mając do dyspozycji zbiór precedensów reakcji chemicznych, potrafiłyby rozpoznać w nim przykłady błędne? Reakcje, które nie zaszły lub w których zapisie brakuje atomów? Omówię to zagadnienie na przykładzie bazy reakcji pochodzącej z USPTO (United States Patent and Trademark Office).



Ryc.21. Przykłady błędów chemicznych w reakcjach pochodzących z bazy USPTO.

Kluczową kwestią w automatycznym tworzeniu reguł reakcji jest poprawność danych zawartych w bazach danych, z wykorzystaniem których tworzone są automatycznie reguły reakcji²⁶¹. Najpopularniejsze wśród chemików syntetyków bazy z danymi o reakcjach chemicznych: baza amerykańskich patentów USPTO oraz bazy Reaxys i SciFinder zawierają błędy powstałe najczęściej podczas transferu z nośnika materialnego do postaci cyfrowej. Co więcej, nie są stosowane systemowe narzędzia usuwania błędów i poprawy jakości bazy²³⁰. Skutkuje to tym, że program stworzywszy regułę reakcji w oparciu o błąd w bazie danych, stosuje tą reakcję na równi z innymi.

Gdyby baza zawierała 1% błędnych reakcji, to można by się spodziewać, że te błędne reguły również będą stosowane proporcjonalnie w 1% przypadków. Tak się jednak nie dzieje. Podczas prac nad Chemicą podjęto próbę stosowania zbiorów reguł automatycznie ekstrahowanych z baz. Skończyło się to notorycznym stosowaniem reakcji stworzonych w oparciu o błędy w bazach. Dokonywały one znakomitych (choć przecież niemożliwych) transformacji chemicznych – były to przejścia na tyle atrakcyjne, że to właśnie one były proponowane, a nie przejścia stworzone w oparciu o wiedzę chemików i rzetelne źródła. Podejście to zarzucono, uodparniając się tym samym na pokusę automatycznego pozyskania wiedzy chemicznej bezpośrednio z baz danych. Jednak ta pokusa jest na tyle silna, że to właśnie ta droga jest obecnie najczęściej obierana przez twórców współczesnych programów^{231,235,236,261}.

Baza danych USPTO stanowiła podstawę dla automatycznie stworzonych reguł reakcji w podejściu Jensena²⁶¹. Jeżeli uczy się sieci neuronowe na przykładach reakcji chemicznych zawierających błędy, to czy możliwe jest w ten sposób stworzenie reguł poprawnych? Przyjrzyjmy się bliżej typom błędów występujących w USPTO (analizę rozkładu błędów w USPTO prowadziłem w latach 2017-2018 na potrzeby projektu nie wchodzącego w skład niniejszej rozprawy).

Pierwszy typ błędów (Ryc.21a) dotyczy pominięcia w zapisie części atomów. W tym przypadku jest to nieprawidłowy zapis grupy -OTBS, brakuje oczywiście w nim dwóch grup metylowych. Skutki są jednak dalej idące: interpreter zapisu zapełnia automatycznie puste walencyjności za pomocą protonów, dlatego reaktywność przedstawionej grupy jest diametralnie inna, a algorytm błędnie rozpozna wodorki krzemu jako grupę chemicznie inertną. Z kolei Ryc.21b przedstawia przykład zmiany atomu w zapisie – atom węgla z substratu staje się w produkcie atomem tlenu. Oczywiście nie odpowiada temu przekształceniu żadna reakcja chemiczna, niemniej algorytm stworzy regułę umożliwiającą niniejsze przejście. Na Ryc.21c przedstawiono reakcję Wittiga, jednak egzometylenowe wiązanie podwójne w produkcie nie koresponduje z produktem faktycznie powstającym w tej reakcji. „Reakcjom” przedstawionym na Ryc.21d-e nie odpowiadają znane transformacje chemiczne, niemniej lista pozbawionych sensu chemicznego zbiorów struktur po obu stronach strzałki stanie się podstawą dla stworzenia reguły reakcji. Te kilka przykładów błędnych zapisów zostanie potraktowane przez algorytm ekstrahujący reguły identycznie jak reakcje poprawne. Stanowi to jedno z poważniejszych ograniczeń zautomatyzowanej „drogi na skróty” w tworzeniu bazy reguł o reakcjach organicznych.

Kiedy w latach 60-tych XX-go wieku zaczęła rosnąć popularność wykorzystania komputerów, pomysł, by komputer wspomagał chemika wydał się oczywistą konsekwencją rozwoju tej technologii. Jednak na realizację tej idei przyszło czekać ponad pół wieku. Podobnie dzisiaj, obserwujemy rozwój narzędzi szeroko pojętej sztucznej inteligencji, sieci neuronowych i uczenia maszynowego. Jednak może się okazać, że na zastosowanie ich do chemii organicznej trzeba będzie jeszcze poczekać. Czy dzisiejsze programy komputerowe bazujące na paradygmacie poprawności automatycznej ekstrakcji reguł reakcji chemicznych, są zdolne generować poprawne i wartościowe plany syntetyczne?

Dziś odpowiedź na to pytanie brzmi: nie. Dopóki się to nie zmieni, uczenie maszynowe skazane będzie na powtarzanie tych samych błędów. Co więcej, chemia organiczna jest grą o liczbie reguł, idącej w dziesiątki tysięcy, a liczba tych reguł, cały czas rośnie, w miarę odkrywania nowych reakcji. Dodać należy do tego wyposażenie algorytmu do automatycznej ekstrakcji reguł w zdolność rozpoznawania kontekstu reakcji, relacji między elementami stereogenicznymi w strukturze oraz oceny jakie elementy strukturalne są dla zajścia reakcji kluczowe, a które stanowią inertny fragment szkieletu. Nie zapominajmy, że błędne przyswojenie choćby jednej automatycznie wyabstrahowanej reguły będzie eskalowało w drzewie retrosyntetycznym błędnymi rezultatami.

Wyzwanie to czeka w przyszłości na swe rozwiązanie.

2.3. Chematica / Syntaurus / Synthia™

2.3.1. Wstęp

Program Chematica zasługuje tu na szczególne potraktowanie, gdyż w trakcie studiów doktoranckich uczestniczyłem w pracach nad jego rozwojem.

Chematica była pomysłem prof. Bartosza Grzybowskiego, rozwijanym od wczesnych lat XXI wieku aż do dnia dzisiejszego, najpierw na Uniwersytecie Northwestern, a następnie w Instytucie Chemii Organicznej PAN w Warszawie. W stworzenie programu były zaangażowane dziesiątki osób. Projekt ten wymagał połączenia wiedzy oraz współpracy chemików, matematyków i informatyków, w stopniu niespotykanym w dotychczasowych podejściach. Jako, że uczestniczyłem w rozwoju tylko części programu, omówienie zasady działania całości lokalizuję we wstępie literaturowym, natomiast prowadzoną przeze mnie część prac nad tworzeniem eksperckiej bazy danych o regułach reakcji chemicznych, umieszczam jako badania własne.

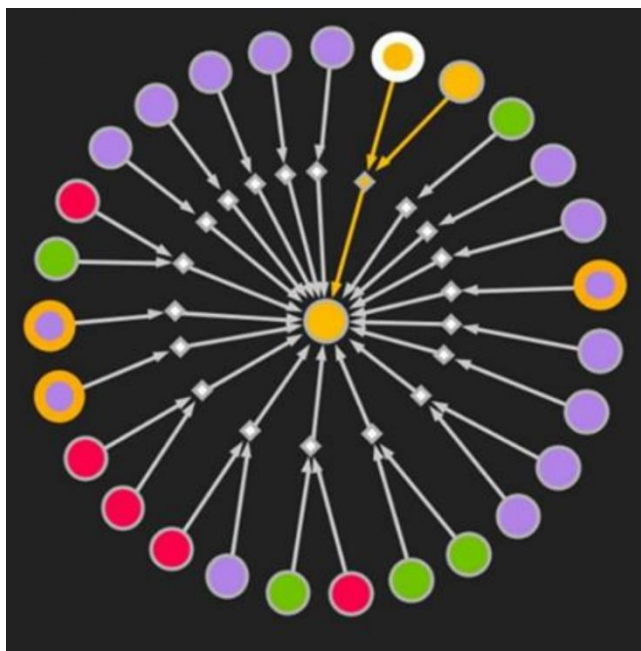
O czym mówimy, gdy mówimy o „Chematice” wymaga uściślenia. W pierwszej wersji programu, pod tą nazwą krył się moduł sieciowy, z którego rezultaty²²⁶⁻²²⁹ publikowano w latach 2005-2012. Moduł ten zawierał informacje z baz danych o reakcjach chemicznych, z których to danych utworzono sieć. Węzłami sieci były cząsteczki chemiczne, a łączyły je reakcje prowadzące z jednej do drugiej. Taka reprezentacja pozwalała nie tylko na inny rodzaj wizualizacji wiedzy chemicznej, lecz również pozwalała operować metadanymi chemicznymi w nowy sposób. Pozwalając znajdować plany syntetyczne optymalne według różnych zadanych kryteriów. Ważne jest tu słowo „znajdować”, bowiem moduł sieciowy mógł poruszać się tylko po ograniczonej liczbie węzłów budujących sieć. Jeżeli zapytalibyśmy moduł sieciowy o cząsteczkę, która nigdy nie została uzyskana, nie uzyskalibyśmy odpowiedzi. Czy to oznacza, że cząsteczka niefigurująca w sieci, jest niemożliwa do uzyskania? Bynajmniej, jednakże ówczesna wersja Chematici mogła tylko „znajdować” to co w sieci się znajdowało, nie mając możliwości „tworzenia” diametralnie nowych planów syntetycznych niebędących kombinacjami istniejących.

W tym miejscu na scenie pojawia się moduł retrosyntetyczny^{220,272-276}. Prace nad jego stworzeniem prof. Grzybowski rozpoczął na Uniwersytecie Northwestern, jednak po założeniu grupy badawczej w Instytucie Chemii Organicznej PAN, w 2013 roku, to w Warszawie prowadzono dalsze prace nad jego rozwojem.

Automatyczny moduł retrosyntetyczny nosi nazwę Syntaurus i wszystkie plany syntetyczne i przykłady walidacji omówione w tej pracy, uzyskano za jego pomocą. Ponadto po zakupie programu przez firmę Merck KGaA, na potrzeby komercyjne zmieniono nazwę z Chematica na Synthia™. Plany tu omawiane zostały wygenerowane zanim doszło do przejęcia własności przez firmę Merck KGaA, dlatego, dla uproszczenia, dalej będę używał jedynie nazwy Chematica.

2.3.2. Założenia i ogólna zasada działania

Najważniejszym wymogiem stawianym modułowi retrosyntetycznemu, było „wyjście poza sieć” w dalszym ciągu zachowując chemiczną poprawność. Wyjście to umożliwiłoby planowanie syntez z wykorzystaniem związków chemicznych dotychczas nigdy nie uzyskanych, acz teoretycznie możliwych. Dotychczas poprawność chemiczną gwarantowało ściśle odtwarzanie precedensów literaturowych – pozostawało więc „nauczenie” programu rozumienia treści precedensów. Rezultaty prób osiągnięcia tego celu z wykorzystaniem uczenia maszynowego omówiłem wcześniej. Zdecydowano się na stworzenie eksperckiej bazy reguł reakcji chemicznych – większej niż dotychczasowe i bardziej ukierunkowanej na percepcję kontekstu chemicznego.



Ryc.22. Przykład notacji retrosyntetycznej, przyjętej w programie Chemica. W centrum na żółto umieszczony jest cel syntetyczny. Otaczają go strzałki kończące się w szarych kwadratach – pod nimi kryje się informacja o zastosowanej regule reakcji chemicznej, o najpopularniejszych warunkach prowadzenia tej reakcji, odnośniki literaturowe do przypadków stosowania reakcji oraz, w przypadku konieczności zastosowania grup ochronnych, propozycja, które z nich zastosować. Do ów kwadracików/reguł reakcji prowadzi/prowadzą strzałki od kólek oznaczających substraty, tj. konkretne struktury związków chemicznych wygenerowanych na podstawie reguły reakcji. Kolory kólek oznaczają: czerwony – związek komercyjnie dostępny, zielony – związek znany w literaturze, fioletowy – związek nigdy nie opisany w literaturze, jednak teoretycznie mogący istnieć. Pomarańczowe obwódki oznaczają występowanie niekompatybilności chemicznej, natomiast niebieskie obwódki oznaczają występowanie konfliktu reaktywności wymagającego użycia grup zabezpieczających.

Proces tworzenia i kodowania reguł reakcji chemicznych opisuję w części badania własne. Chemica dysponując wyłącznie ręcznie zakodowanymi regułami reakcji chemicznych, nie mogła jeszcze spełniać stawianych jej zadań. Przy pomocy reguł, w oparciu o wprowadzoną strukturę celu syntetycznego, generowana jest automatycznie pierwsza generacja prekursorów G1. Zazwyczaj jest ich około 100 dla jednego związku. Możliwe jest na tym etapie obejrzenie ich struktur, sprawdzenie warunków reakcji, czy informacji o niekompatybilnościach chemicznych. Po wyborze przez użytkownika najbardziej przekonującej go transformacji może on kontynuować analizę retrosyntetyczną uzyskując dla wybranego G1 kolejne około 100 prekursorów G2 itd. Ten sposób interaktywnej analizy retrosyntetycznej, jaki znamy na przykład z LHASA, jest procesem długotrwałym i silnie zależącym od użytkownika, jego pamięci i wiedzy chemicznej. Tymczasem Chemica miała być narzędziem skierowanym również do osób potrzebujących planów syntetycznych, lecz nie będących ekspertami w dziedzinie planowania syntezy. Przesunięcie ciężaru odpowiedzialności za wybór transformacji do planu syntetycznego z użytkownika na program stało się kolejnym, acz rozwiązywanym równoległe, problemem. By ów problem rozwiązać należało opracować narzędzie różnicujące reakcje i transformacje w oparciu o zadane kryteria, należało zatem wykorzystać kryteria posiadające sens chemiczny. Dzięki temu 100 związków z generacji G1 przestawało być 100 równie dobrymi alternatywami – algorytm wyróżniał reakcje lepsze bądź gorsze, według wybranych kryteriów. Zmiana kryteriów oceny lub wag im przypisanych, zmieniała

jednocześnie listę z wytypowanymi reakcjami. Wybór kryteriów oceny pozostawiono użytkownikowi. Może on wybierać spośród szeregu zestandaryzowanych Scoring Functions. Wprowadzenie w 2016 roku tego pionierskiego narzędzia przeszukiwania drzew retrosyntetycznych, jest niewątpliwym osiągnięciem zespołu profesora Grzybowskiego²²⁰. Oprócz znacznej oszczędności czasu i ujednoczenia procesów wyszukiwania, zdejmuje to z użytkownika większość trudu związanego z interaktywnym tworzeniem planu, pozwalając poświęcić czas operatora na przegląd gotowych i posortowanych podług ocen planów syntetycznych. Wszystkie funkcje przypisano do jednej z dwóch opisanych poniżej kategorii: CSF i RSF. Funkcje zapisuje się w postaci wielomianu, gdzie zmiennymi są poszczególne parametry szczegółowe (np. MASS, SMILES_LEN), a przypisane im różne wagi są odzwierciedleniem preferencji i priorytetów użytkownika.

CSF (Chemicals' Scoring Function) – kategoria grupująca funkcje oceniające cząsteczki chemiczne zaproponowane jako substraty w oparciu o regułę reakcji. W ogólności pozwala poddać ocenie ilość i jakość cech strukturalnych wnoszonych przez transformację do struktury następnego związku w analizie retrosyntetycznej. W szczególności najużyteczniejszymi z tych funkcji są:

RINGS – funkcja mająca zastosowanie do cząsteczek posiadających pierścienie: jeżeli w reakcji tworzony jest pierścień, to liczba pierścieni w substratach i produktach jest różna, co jest wykrywane przez tę funkcję. Pomaga ona promować transformacje wnoszące występujące w celu syntetycznym cechy strukturalne, a jak najszybsze wprowadzenie tych kluczowych elementów wpływa na skrócenie całego planu syntetycznego. Dlatego promowanie transformacji wnoszących pożądane cechy strukturalne, względem transformacji tego nie robiących, stanowi przykład pozytywnej selekcji spośród wygenerowanych substratów.

STEREO – funkcja działająca analogicznie do funkcji RINGS, zliczająca centra stereogeniczne. W szybki sposób wykrywa reakcje stereospecyficzne i stereoselektywne, pozwalając budować plan syntetyczny na wprowadzaniu jak największej liczby centrów stereogenicznych w jak najmniejszej liczbie transformacji.

MASS – funkcja powiązana z masą cząsteczkową aktualnego celu syntetycznego i mas substratów dlań proponowanych (funkcja ta ściśle dotyczy par substrat/y – produkt, w każdej generacji). Pozwala na wyrażenie preferencji co do sposobu cięcia retrosyntetycznego. Czy ma ono prowadzić do substratów o zbliżonych masach, co odpowiada w przybliżeniu tworzeniu cząsteczki z substratów o zbliżonej masie, czy też preferujemy do dużego rdzenia dołączać mniejsze fragmenty, coraz bardziej go upodabniające do celu syntetycznego. Na podkreślenie w tym momencie zasługuje ważność dla planowania syntezy przyjętej strategii: liniowej lub zbieżnej. Syntezy liniowe dobrze sprawdzają się dla niewielkich związków i wykonywania syntezy w małym zespole. Taki kształt czyni z każdego etapu wąskie gardło, bez przejścia którego praca nad następnymi etapami jest niemożliwa i oprócz powtarzania pierwszych etapów w celu uzyskania większej ilości substratów do testu newralgicznego etapu, praca ulega zahamowaniu. Jest to szczególnie dotkliwe podczas pracy w dużym zespole, gdy problemy z jedną reakcją skutkują niemożnością optymalnego wykorzystania całego zespołu syntetyków. Syntezy zbieżne, są pod tym względem o wiele wdzięczniejsze, gdyż pozwalają rozdzielić pracę bardziej równomiernie, a w razie problemów z realizacją poszczególnej transformacji, całość prac nie jest paraliżowana. Dlatego rola funkcji MASS zasługuje na szczególne podkreślenie, ponieważ pozwala dopasować poszukiwany plan syntetyczny do możliwości kadrowych laboratorium wykonującego syntezę.

SMILES_LEN – to funkcja podobna do MASS. Jednak w większym stopniu opisuje ona złożoność strukturalną. Struktury w Chematica zapisane są za pomocą notacji SMILES/SMARTS²⁷⁷⁻²⁷⁸. W tym liniowym sposobie zapisu każdemu atomowi odpowiada jego skrót chemiczny. Nie są brane pod uwagę masy atomów, a jedynie ich ilość (z wyłączeniem atomów wodoru) budująca cząsteczkę, a co za tym idzie – jej złożoność. Oczywiście ilość atomów nie przekłada się w sposób

prosty na złożoność, gdyż 11 atomów węgla może budować zarówno niezłożony undekan, jak i dwupierścieniową metylodekalinę. Obecność zwiększających złożoność elementów, takich jak rozgałęzienia, pierścienie, centra stereogeniczne – wpływają na funkcję SMILES_LEN. Proste porównanie jodoetanu z kamforą dobrze ilustruje różnice między funkcjami MASS a SMILES_LEN. Dla funkcji MASS obie te cząsteczki mają zbliżoną masę, toteż nie ma z jej perspektywy między nimi różnicy, jednak duża masa jodu sprawia, że jednemu jego atomowi odpowiada 10 atomów węgla, a one mogą już tworzyć złożoną strukturę. Złożoność tą dostrzega właśnie funkcja SMILES_LEN.

KNOWN – jest to funkcja pozytywnie oceniająca wykorzystanie przez regułę reakcji substratu, który jest znany w literaturze (w notacji Chematici oznaczany jako zielone kółko, w odróżnieniu od komercyjnie dostępnych związków – oznaczanych czerwonym kółkiem, czy też związków nieznanymi w literaturze – oznaczonych kółkiem fioletowym).

BUY – funkcja analogiczna do KNOWN, lecz pozytywnie oceniająca wykorzystanie w analizie retrosyntetycznej, związków komercyjnie dostępnych.

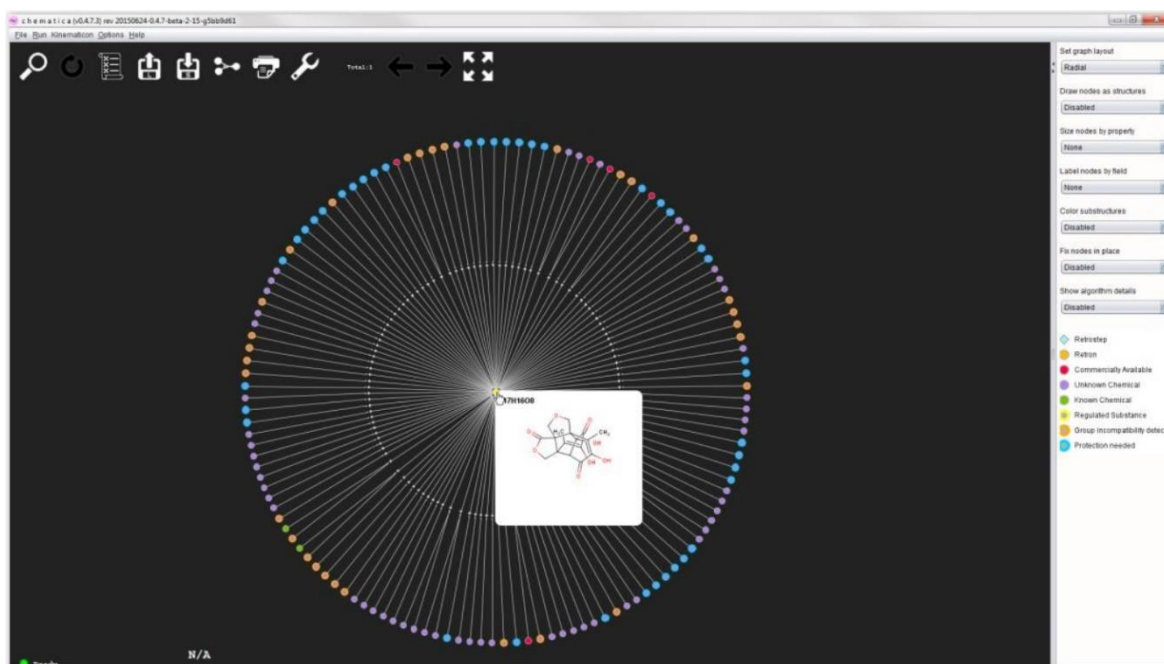
RSF (Reaction Scoring Function) – kategoria grupująca funkcje oceniające transformacje chemiczne zaproponowane w oparciu o regułę reakcji jako aplikowalne do celu syntetycznego. W ogólności pozwala wybrać plany najkrótsze i najmniej narażone na praktyczne utrudnienia podczas próby wykonywania planu.

CONFLICT – funkcja rozpoznająca nieusuwalny konflikt reaktywności pomiędzy grupami funkcyjnymi. Informacja o zaistnieniu konfliktu jest tworzona w oparciu o przypisaną do każdej reguły reakcji listę grup funkcyjnych, których obecność w cząsteczce może utrudniać zajście reakcji lub ją uniemożliwiać. Konfliktu reaktywności nie można znieść poprzez użycie grupy ochronnej, gdyż części grup funkcyjnych, na przykład grupy azydkowej czy sulfonowej, w ten sposób nie można zabezpieczyć. Funkcja oceniająca będzie promowała plany syntetyczne, w których nie występuje ryzyko problemów praktycznych wynikających z konfliktu reaktywności.

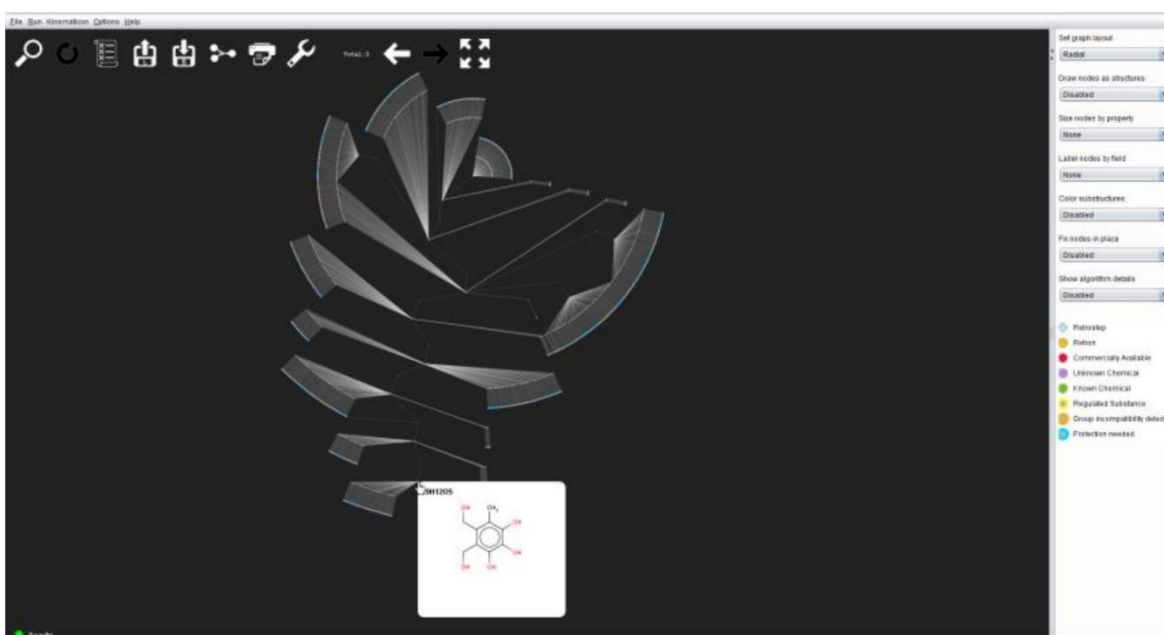
PROTECT – funkcja analogiczna do CONFLICT, jednak dokonująca wyróżnienia transformacji proponujących użycie substratów o grupach funkcyjnych wymagających grup zabezpieczających. Wprawdzie jest to identyfikacja problemu mniej dokuczliwego niż zasadniczy i nieomijalny konflikt reaktywności, niemniej z punktu widzenia realizacji planu syntetycznego, wybór tej transformacji wydłuża dodatkowo cały plan o 2 dodatkowe etapy: założenia i usunięcia grupy zabezpieczającej. Obniża to oczywiście wydajność całego procesu, dlatego wybór reakcji, gdzie można uniknąć stosowania grup zabezpieczających jest lepszym sposobem niż ich stosowanie.

YIELD – funkcja służąca rozpoznawaniu i faworyzowaniu reakcji o prawdopodobnie wysokiej wydajności i dyskryminująca reakcje prawdopodobnie mało wydajne. Ocena ta dokonywana jest o estymację wydajności, za pomocą metody opracowanej i opublikowanej uprzednio²⁷⁹.

Połączenie tych dwóch funkcji zbiorczych, stanowi narzędzie potrafiące rozpoznać parametry opisujące syntezę i dokonać ich ilościowego różnicowania. Stanowi to podstawę do algorytmicznej oceny poszczególnych generacji drzewa retrosyntetycznego i poruszania się po nim w najbardziej rokujących lokalnie kierunkach. Wracamy w tym miejscu do poruszonego na wstępie problemu najlepszej możliwej syntezy. Nigdy nie ma gwarancji, że w olbrzymim, teoretycznym drzewie retrosyntetycznym nie jest „ukryta” imponująca sekwencja reakcyjna. Dlatego to, co algorytmy posługujące się CSF i RSF znajdują to sekwencje transformacji, najlepiej ocenione przy użyciu zadanych kryteriów. Pozostaje zatem pytanie, czy to wystarczająco, by zaproponować wartościowe plany syntetyczne? Jest to precyzyjniej postawiona wersja hipotezy badawczej: po skodyfikowaniu reguł planowania syntezy, nie ma różnicy czy stosuje je człowiek czy komputer.



Ryc.23. Ilustracja pierwszej generacji prekursorów dla epikolaktonu²²⁰.



Ryc.24. Ilustracja manualnego wyboru sekwencji transformacji prowadzących do epikolaktonu²²⁰.

Skoro zatem nie ma różnicy, to nie ma przeszkody by program komputerowy mógł zaproponować plan porównywalny lub lepszy, bo używając tych samych reguł reakcji analizy będzie dokonywał bez porównania szybciej niż człowiek.

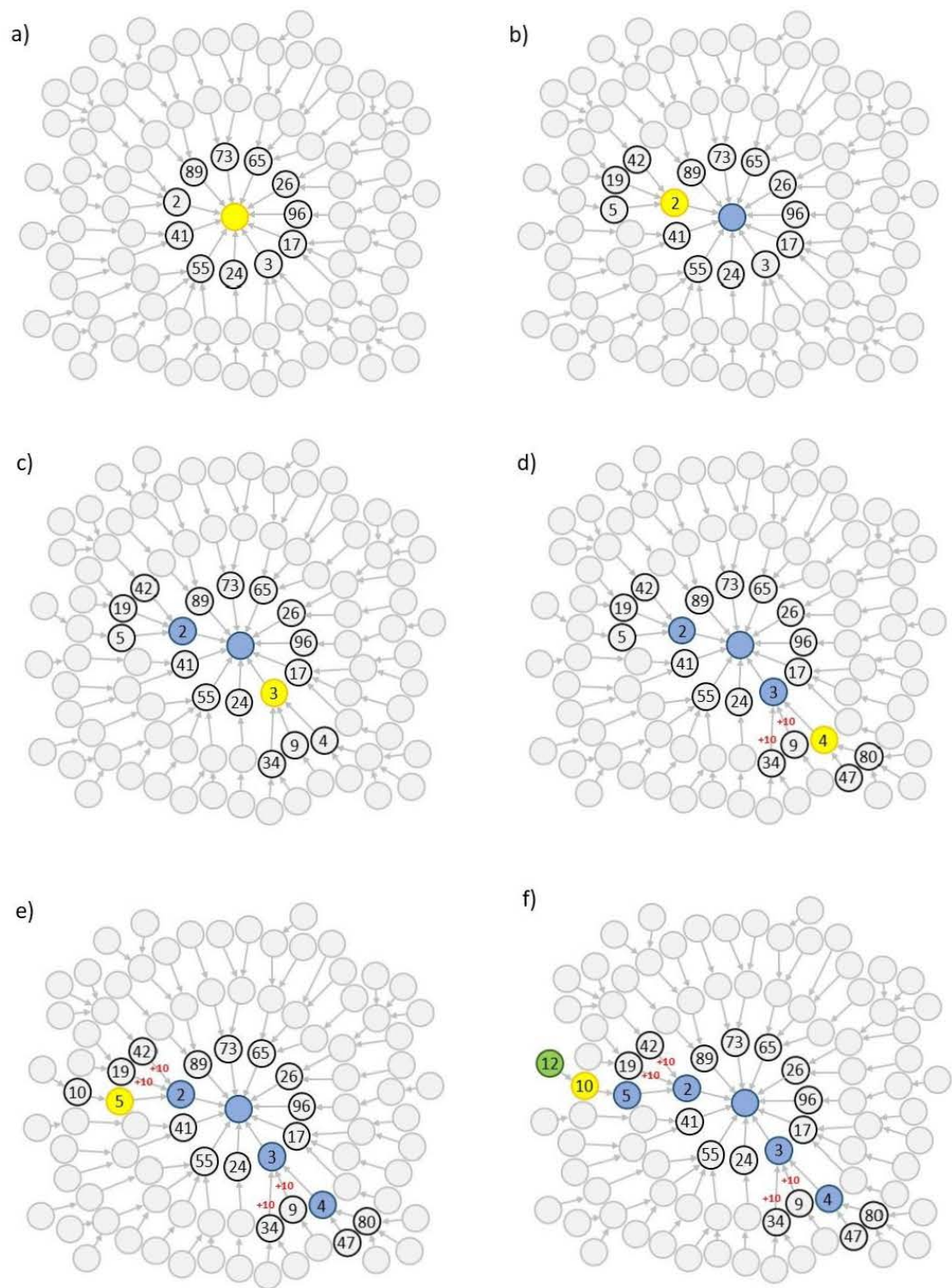
Różnicę między manualnym/interaktywnym przeszukiwaniem drzewa retrosyntetycznego a robieniem tego przy pomocy algorytmu, pomogą uzmysłowić dwie ilustracje Ryc.23. i Ryc.24. Na pierwszej ilustracji przedstawiono typowy wygląd pierwszej generacji substratów i transformacji zaproponowanych przez Chematicę dla zadanego celu syntetycznego (w tym przypadku dla epikolaktonu). Samo ich obejrzenie zajmuje długie minuty, a z każdą generacją należałoby postąpić

podobnie wnikliwie (Ryc.24.). Ten sposób pracy z programem zapewnia użytkownikowi poczucie kontroli, ale trudno przy jego użyciu przeprowadzać zaawansowane analizy. Użyto go jednak w celu walidacji programu²²⁰ poprzez poszukiwanie opublikowanych planów syntetycznych w drzewach retrosyntetycznych. Na powyższych ilustracjach przedstawiono przykład takiej pozytywnej walidacji dla epicolaktonu²²⁰.

Opisane powyżej trudności w manualnym przeprowadzaniu analizy retrosyntetycznej zostały rozwiązane przy użyciu algorytmu znajdującego sekwencje syntetyczne. Tradycyjna analiza retrosyntetyczna dokonywana jest przy użyciu ocen wyłącznie jakościowych odzwierciedlających pomysłowość i preferencje chemika. Podczas takiej analizy często wybór jest dokonywany na bazie ludzkiej intuicji podpowiadającej: „to będzie dobra transformacja”. Podejście to dobrze sprawdza się gdy analizie poddawanych jest kilka alternatyw. Jednak wobec większych zbiorów alternatywnych reakcji, nie zdaje ono egzaminu. Dlatego Chematica posługuje się liczbowymi kryteriami do porównywania możliwości ruchu na drzewie retrosyntetycznym.

Na Ryc.25. przedstawiono przykład wyszukiwania stosowanego przez Chematicę. Jest to oczywiście schemat uproszczony, bowiem, jak już wspomniano, każda generacja zawiera około 100 różnych transformacji. Tutaj pokazano znacznie mniejsze, symboliczne drzewo retrosyntetyczne. Wartości w kółkach to suma CSF i RSF omówionych wcześniej. Dzięki użytym funkcjom, jakościowe kryteria strukturalne są tu przekładane na konkretną liczbową ocenę – oczywiście inna funkcja oceniając to samo drzewo retrosyntetyczne przypisze związkom przejściowym i prowadzącym doń transformacjom inne oceny. Co więcej, może poruszać się po drzewie retrosyntetycznym w inny sposób i rozwijać gałęzie nieeksplorowane przez wcześniejsze funkcje, gdyż to co dla innych funkcji było ukryte za zbyt wysokimi wartościami ocen, po zmianie funkcji i obniżeniu ocen stawało się dostępne. Tu znów należy podkreślić, że program nie znajduje obiektywnie najlepszej ścieżki w całym drzewie, bo i całego drzewa nie poddaje wyczerpującej ocenie – a jedynie w oparciu o zastosowaną funkcję bada wycinek drzewa dający lokalnie najniższe wyniki (tu, najbardziej pożądane). Natomiast całe obszary drzewa retrosyntetycznego, będące rozwinięciem węzłów o ocenie nie rokującej chemicznego powodzenia są implicite oceniane jako złe. W tym sensie ocenie poddawane jest całe drzewo retrosyntetyczne.

Pierwsza ilustracja Ryc.25a przedstawia początek analizy. W centrum, żółtym kółkiem zaznaczono cel syntetyczny, a otaczająca go pierwsza generacja związków przejściowych i prowadzących doń transformacji (wyróżnione czarnymi obwódkami, od teraz nazywane zbiorczo wierzchołkami) poddane zostały ocenie (liczby wpisane w kółka). Po dokonaniu oceny, wierzchołki umieszczane są na liście ważności (priority queue, PQ). Na razie zawiera ona: 2, 3, 17, 24, 26, 41, 55, 65, 73, 89, 97. W następnym kroku (Ryc.25b) dotychczas rozwijany/analizowany wierzchołek zaznaczono na niebiesko, a wierzchołek o najniższym wyniku (czyli najlepiej oceniony przez funkcję) zostaje rozwinięty, zaznaczony na żółto i usunięty z PQ. Teraz PQ zawiera wierzchołki o ocenach: 3, 5, 17, 19, 24, 26, 41, 42, 55, 65, 73, 89, 97. Po rozwinięciu wierzchołka 2, następnym z najwyższą oceną jest wierzchołek 3 (Ryc.25c). Po jego rozwinięciu i analizie, do PQ dodane zostają: 4, 9, 34. Wierzchołek 3 zostaje usunięty z PQ. Następną najbardziej obiecującą opcją jest wierzchołek o wartości 4 (Ryc.25d). Poddany analizie dodaje do PQ prowadzące doń wierzchołki o ocenach: 47, 80, samemu będąc z PQ usunięty. Na tym etapie przypisane zostają kary do wierzchołków (czerwone +10) zbiegających do wierzchołka 3. Karze podlegają wierzchołki o tej samej historii rozwoju jak aktualnie analizowany wierzchołek. Służy to ograniczeniu analizy w jednej tylko części hipergrafu, wymuszając opuszczanie minimów lokalnych. Przekłada się to bezpośrednio na różnorodność uzyskiwanych sekwencji syntetycznych. Po tej operacji PQ zawiera: 5, 17, (9+10), 19, 24, 26, 41, 42, (34+10), 47, 55, 65, 73, 80, 89, 97, co oznacza, że następnie rozwijany będzie wierzchołek 5 (Ryc.25e). Zostaje on usunięty z PQ i rozwinięty. Warto tu zauważyć, że po rozwinięciu do PQ dodany zostaje wierzchołek 10. Gdyby nie kara (+10) nałożona



Ryc.25. Ilustracja działania algorytmu znajdującego sekwencje syntetyczne w drzewie retrosyntetycznym.

na wierzchołki zbiegające się w wierzchołku 3, to następnym analizowanym byłby wierzchołek 9. Do rozwinięcia PQ wskazało wierzchołek 10 (Ryc.25f). Po rozwinięciu znaleziono wierzchołek 12, który spełnia końcowe kryteria wyszukiwania (związek znany lub komercyjnie dostępny, poniżej zadanej ceny za gram i masy cząsteczkowej – wartości ustawiane przez użytkownika przed rozpoczęciem wyszukiwania, obok funkcji CSF i RSF oraz struktury celu syntetycznego). Ta ścieżka spełnia zadane funkcjami kryteria wyszukiwania, zostaje ona zapisana do przedstawienia użytkownikowi i dalszej analizy. Algorytm działa dalej generując następne ścieżki w sposób analogiczny. Na końcu wyszukiwania zostają one zaprezentowane w postaci listy, w zależności od przypisanych przez funkcję wag.

2.3.3. Szczególne zasady działania

Like Polish grammar, organic synthesis is full of exceptions.

B.A. Grzybowski, 2016²²⁰

W toku prac nad ekspercką bazą danych o reakcjach okazało się, że nie wszystkie instrukcje o przebiegu transformacji da się zawrzeć w postaci reguły reakcji. Przykładom takich problemów i sposobom ich rozwiązania, poświęcam niniejszy rozdział. Omówię sześć problemów wymagających specjalnego potraktowania: aromatyczną substytucję elektrofilową, reakcję Dielsa-Aldera, heurystyki konformacyjne, listę labilnych grup funkcyjnych, wybór substratu do cyklizacji oraz sekwencje strategiczne.

2.3.3.1. Aromatyczna substytucja elektrofilowa

Jedną z najbardziej użytecznych reakcji w repertuarze syntetycznej chemii organicznej, jest aromatyczna substytucja elektrofilowa (ASE). Najczęściej stosowana w postaci reakcji Friedla-Craftsa, nitrowaniu czy bromowaniu układów aromatycznych. Pozwala ona na precyzyjną funkcjonalizację i to właśnie czynniki ów precyzję (selektywność wyboru pozycji) determinujące, trudno poddają się zapisowi za pomocą reguł reakcji. Przyjrzyjmy się temu przypadkowi bliżej, w najprostszym wariantcie (bez specyfikowania podstawników przy węglu alkiilowym o hybrydyzacji sp^3 zapisywanemu jako [CX4:4]) acylowanie Friedla-Craftsa można zapisać jako: [c:1][C:2](=[O:3])[CX4,c:4].[Cl:-5]>>[c:1].[Cl:5][C:2](=[O:3])[*:4]. Zapis taki uwzględnia wszystkie węgle aromatyczne ([c:1]) i niemal całą reakcję mamy pomieszczoną w jednym zapisie. Szkopuł w tym, że zapis taki nic nie mówi o selektywności podstawienia orto/para-meta. Zapis ten dla niepodstawionego pierścienia benzenu jest poprawny, ale umożliwia też szersze zastosowanie reguły – pozwalając na przyłączenie grupy acylowej do dowolnego aromatycznego atomu węgla. Oczywiście nie ma to wiele wspólnego z chemiczną poprawnością, jak bowiem program miałby rozpoznawać sytuacje kiedy acylowanie zachodzi w pozycję meta, albo jak postulować zastosowanie tej reakcji, gdy w cząsteczce obecnych jest więcej pierścieni aromatycznych? Czy możliwe w tej sytuacji jest doprecyzowanie reguły reakcji i rozbitcie zapisu na więcej linii? Omówię to na przykładzie benzenu. W przypadku tego pierścienia, na kierunek podstawienia, wpływ będą miały podstawniki (rodzaj pierścienia czyni na wejściu wszystkie z nich równocennymi) – lista podstawników zawiera kilkadziesiąt różnych grup elektronodonorowych i elektronoakceptorowych. Następnie należałoby umieścić wszystkie ich kombinacje na wszystkich pozycjach benzenu, pozostawiając wolne pozycje tak by podstawienie zachodziło jednoznacznie. Jako, że pozycji dla kombinacji podstawników jest 5, to mogą one również występować kilkakrotnie. Jak widać dla samego benzenu liczba wariantów koniecznych do zakodowania staje się zbyt duża by poważnie traktować próbę rozwiązania problemu w ten sposób. A nie należy zapominać, że do wyczerpującego

zakodowania kombinacji pozostają jeszcze wszystkie klasy pierścieni heterocyklicznych oraz, że w jednej cząsteczce może wystąpić więcej niż jeden pierścień. Jak wówczas reguła porówna, który z nich jest bardziej reaktywny, skoro drugiego pierścienia, z jego schematem podstawienia, uprzednio nie zakodowano jako podstawnik przy pierwszym?

Aromatyczna substytucja elektrofilowa okazała się przykładem reakcji na tyle uniwersalnej i równocześnie na tyle wrażliwej na czynniki strukturalne, że próba wyabstrahowania jako reguły reakcji wspólnego rdzenia wszystkich substytucji okazała się zbyt trudna. Wobec tego zwrócono się ku innym metodom. Pytanie, na które miały odpowiedzieć, to która z pozycji aromatycznych w analizowanej cząsteczce jest najbardziej reaktywna w ASE. Dotychczasowe metody odpowiedzi na to pytanie nie mogły znaleźć bezpośredniego zastosowania w Chematica. Obliczenia kwantowomechaniczne gwarantowałyby poprawność i przy analizie pojedynczego przypadku byłyby rozsądnym wyborem, jednak podczas analizy retrosyntetycznej analizowane są miliony struktur, a ze względu na uniwersalność reakcji ASE, niebagatelna ich część musiałaby takiej analizie zostać poddana. Nawet jeżeli obliczenia zajmowałyby 1-10 sekund na strukturę, to przy 10 000 strukturach występujących w drzewie retrosyntetycznym, czas analizy wydłużyłby się do rzędu dziesiątek godzin. Taki czas pracy programu był nieakceptowalny. Zmusiło to twórców do opracowania modułu typującego centra reaktywne w oparciu o mieszane metody. Dzięki temu czas pojedynczej analizy jest rzędu milisekund, a poprawność predykcji przekracza 90%²⁷².

W tym celu Chematica wykorzystuje metody: stałe Hammetta dla podstawników, RAPA (ring average proton affinity), model Hückela i szereg dodatkowych reguł empirycznych. W zależności od rodzaju pierścienia aromatycznego, różne metody znajdują zastosowanie. I tak dla pierścienia benzenu najlepiej sprawdzają się stałe Hammetta wzbogacone o poprawki empiryczne, korygujące niedoszacowanie efektu grup silnie elektronodonorowych. Dla pierścieni heterocyklicznych, wybór najreaktywniejszej pozycji dyktowany jest przez typ pierścienia (np. tiofen, pirydyna, indol). Wytypowana przez heurystyki, najbardziej reaktywna pozycja, jest następnie korygowana wpływem podstawników. Wobec wielopierścieniowych węglowodorów aromatycznych nie można zastosować stałych Hammetta, dlatego zamiast tworzenie reguł dla wszystkich możliwych struktur zdecydowano się użyć modelu Hückela, który dla tej klasy związków daje dużą zgodność z predykcji z eksperymentami.

Mając narzędzie typowania najbardziej reaktywnej pozycji w poszczególnych typach pierścieni, można było przystąpić do rozszerzenia typowania na całe cząsteczki proponowane na drzewie retrosyntetycznym jako substraty. Rozszerzone typowanie odbywało się w 4-5 etapach. W pierwszym moduł dzieli ono cząsteczkę na indywidualne pierścienie i dla każdego z nich określa najbardziej reaktywną pozycję. W kolejnym etapie następuje sekwencyjne odrzucanie najmniej reaktywnego pierścienia/pozycji z listy. Jeżeli na końcu tego etapu zdarzy się, że analizowane pozycje mają zbliżoną reaktywność, wówczas zachowywane są do dalszej analizy. W trzecim etapie pozostałe pierścienie są poddawane są procedurze RAPA. Metoda oparta na samym PA (proton affinity) jest zbyt czasochłonna, dlatego używa się wartości wcześniej obliczonych dla pozycji w różnych pierścieniach i układach pierścieni. W oparciu o to porównuje się reaktywność pierścieni, a mniej reaktywne są odrzucane. Czwarty etap używa innej strategii, typując spośród pozostałych pierścieni najbardziej reaktywny w oparciu o heurystykę podobną jak w etapie drugim. Zazwyczaj po tym etapie zostaje tylko jeden pierścień, ale jeżeli substrat posiada dwa lub więcej równocennych pierścieni, które przetrwały eliminację na etapach 1-4, wówczas na etapie piątym decyzja podejmowana jest w oparciu o stałe Hammetta.

Powyższe typowanie dotyczy wyłącznie ASE. Podczas testowania poprawności typowania (na zbiorze testowym 18 000 reakcji wybranych z bazy Reaxys) konfrontowano predykcje z wynikami z bazy danych – nie były wówczas brane wówczas pod uwagę reakcje wykorzystujące metale przejściowe (Pd, Pt, Rh, Ru, Co itp.) lub mocne zasady (np. *n*BuLi, LDA). W ten sposób

odsiano inne reakcje mogące podszyc się pod ASE. Nie udało się osiągnąć 100% trafności – przypadki gdy typowanie nie dawało wyniku takiego jak w bazie danych podzielono na 4 kategorie. W pierwszej są reakcje, których rezultaty leżą w sprzeczności z powszechną wiedzą chemiczną. Są to przypadki stosowania nietypowych warunków dających inne niż zazwyczaj rezultaty, na przykład nitrowanie toluenu w pozycję meta, przy użyciu środowiska micelnego²⁸⁰. Do drugiej kategorii należą mechanizmy kilkietapowe, na przykład acylowanie *N*-acyloindolu za pomocą chlorku acetylu i AlCl₃. W reakcji tej postulowane jest tworzenie się między grupą *N*-acylową indolu, a AlCl₃ kompleksu, który jest grupą kierującą reakcję Friedla-Craftsa w pozycję 6²⁸¹. W trzeciej grupie wyjątków są reakcje, gdzie selekcji między pozycjami dokonuje się przy użyciu rodzaju reagenta – na przykład bromowanie 3,5-dihydroksytoluenu w pozycjach 2 lub 4²⁸²⁻²⁸³. Do czwartej kategorii należą błędy w bazie danych, opisujące precedensy, których faktycznie nie ma w publikacjach źródłowych lub które różnią się znacząco od opublikowanych.

Na podkreślenie zasługuje fakt, że ta pozornie prosta reakcja, jedynie sporadycznie występuje w planach innych programów do komputerowo wspomaganey retrosyntezy. Dla LHASA istnieje wyłącznie jeden taki precedens: plan z 1992 roku¹⁸ zawierający reakcję ASE, stosowaną dla toluenu. Dla SYNCHM brak przykładów tej reakcji oraz jakiegokolwiek podstawiania układów aromatycznych – wszystkie plany dla tego programu wychodzą już z odpowiednio podstawionych związków aromatycznych. Z kolei twórcy SYNSUP już w pierwszej pracy¹³⁷ prezentują reakcję ASE dla pierścienia benzenu, ale co ciekawe w przypadku selektywnego chlorowania przy użyciu Ph₂PCl₃, autorzy słowami: „*Experiment might show otherwise*”¹³⁷ stwierdzają, że trudno przewidzieć czy precedens literaturowy, na którym opierała się reguła, w przedstawianym przypadku również zapewni selektywność.

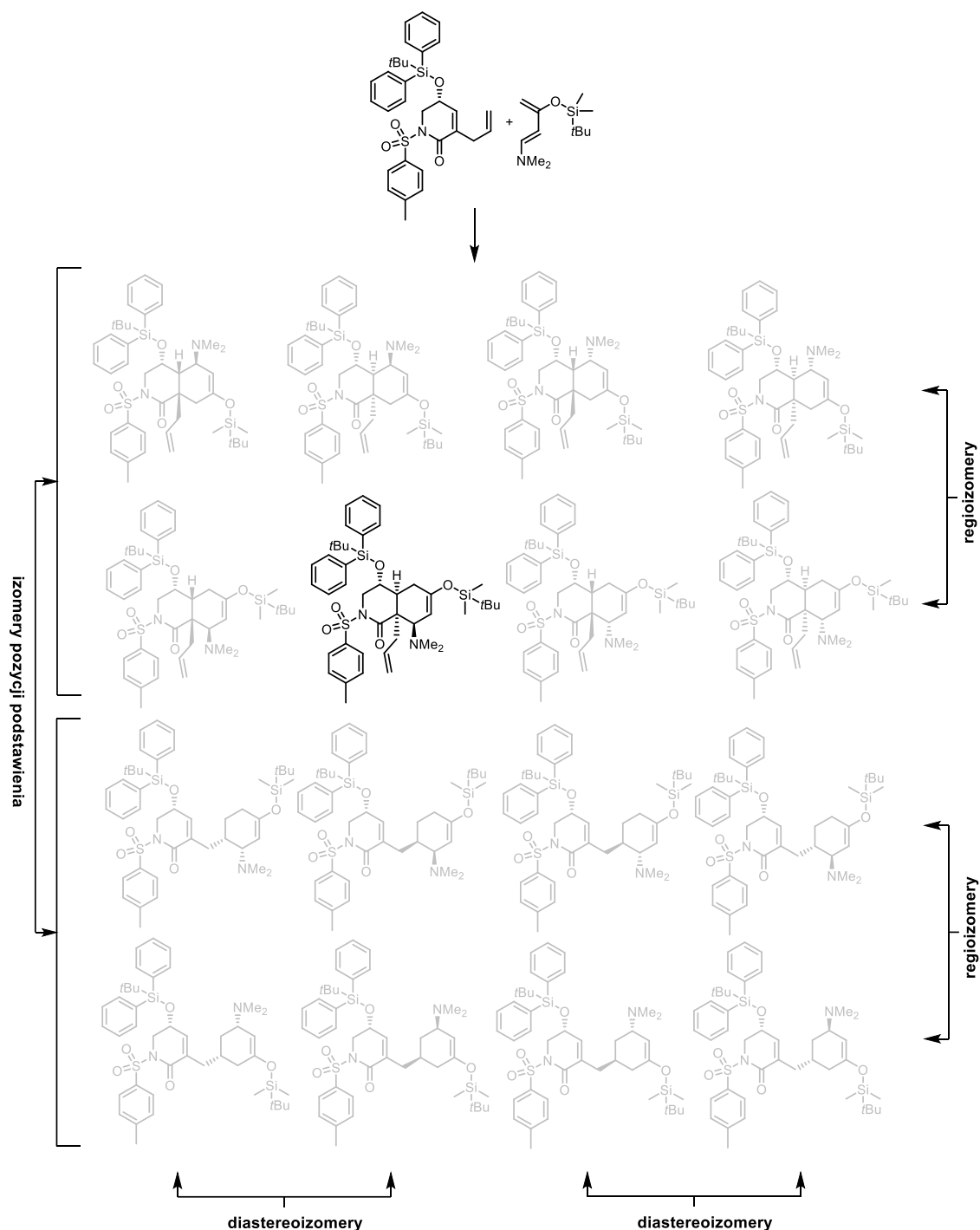
Na przykładzie powyższej reakcji, widać trudności jakie musi przezwyciężyć program wspomagający planowanie syntezy. Spełnienie wymogu jednoczesnej poprawności predykcji reaktywności chemicznej i szybkiego czasu działania, stanowi poważne wyzwanie. A mówimy tu tylko o jednym typie reakcji. Szczęśliwie większość reakcji chemicznych daje się wyrazić za pomocą reguł. Jednak tych, które się temu nie poddają nie można pominąć, zwłaszcza gdy są to wyjątkowo użyteczne transformacje takie jak reakcje ASE lub reakcja Dielsa-Aldera.

2.3.3.2. Reakcja Dielsa-Aldera

Ta najbardziej znana [4+2] cykloaddycja również może zachodzić na różne sposoby (reguła orto/para, produkt egzo/endo, izomeria pozycji podstawienia, gdy jest więcej niż jedno centrum reakcyjne, patrz Ryc.26.). Co więcej, w przeciwieństwie do ASE, w reakcji Dielsa-Aldera trudniej jest powiązać wpływ grup funkcyjnych i konfiguracji z produktem jaki powstaje. Dlatego, oprócz zakodowania najbardziej znanych przykładów reakcji, tworzenie reguł dla przypadków nieznanych w literaturze, opierając się wyłącznie na analogiach strukturalnych, nie daje gwarancji poprawności, a na pewno wyczerpania możliwych wariantów. Jak zatem inkorporować „rozumienie” tej transformacji inaczej niż za pomocą reguł reakcji?

Do testów przewidywania produktu reakcji Dielsa-Aldera użyto²⁸⁴ pochodzącego z bazy Reaxys zbioru kilku tysięcy par dian-dienofil. Nie testowano reakcji wewnątrzcząsteczkowych, gdyż w ich przypadku produkt jest ściśle zależny od kształtu substratu, wykluczając konformacyjnie część izomerów. We wstępnych testach porównano metody kwantowomechaniczne z metodami uczenia maszynowego (lasy losowe i sieci neuronowe). W porównaniu tym metody uczenia maszynowego okazały się zarówno poprawniejsze jak i szybsze (czasy rzędu ułamków sekund kontra czasy rzędu godzin). Po sprawdzeniu różnych wariantów uczenia maszynowego wykazano, że nie ma między nimi bardzo dużych różnic. Źródłem różnic były głównie stosowane deskryptory opisujące właściwości elektronowe podstawników elektronodonorowych i elektroakceptorowych jak

i właściwości steryczne dienu i dienofila. Ilość i różnorodność deskryptorów oraz zbioru uczącego była kluczowa dla „rozumienia” przez maszynę reakcji Dielsa-Aldera. Wybrana ostatecznie metoda lasów losowych wykazała się poprawnością 93,6% przewidywania regioizomeru, 91,3% poprawności przewidywania pozycji reagującej (na przykład, gdy w substracie występowały dwa dienofile) oraz 89,2% poprawności przewidywania diastereoizomeru. Należy zaznaczyć, że test poprawności predykcji wykonano na osobnym zbiorze, pochodzącym z syntez produktów



Ryc.26. Przykład reakcji Dielsa-Aldera i możliwych jej produktów (na szaro), za pomocą uczenia maszynowego typowany jest spośród nich faktycznie powstający izomer (na czarno).

naturalnych. Dlatego można zakładać, że jeżeli dla złożonych strukturalnie rdzeni, poprawności były tak wysokie to dla prostszych struktur byłyby jeszcze większe.

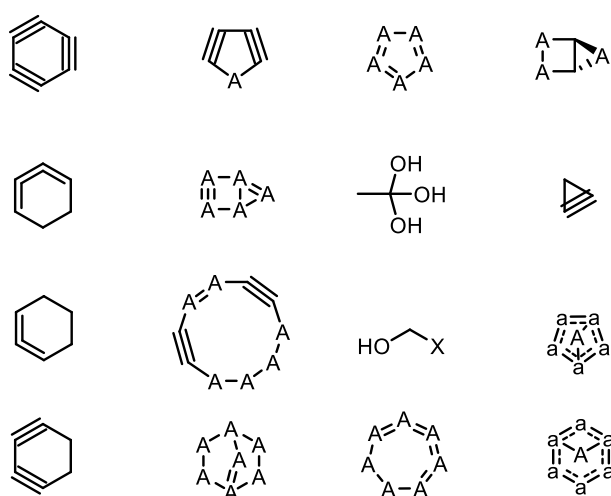
Jednak zastosowanie wyłącznie uczenia maszynowego nie jest wystarczające. W tym przypadku metodę wsparło ukierunkowaniem uczenia maszynowego za pomocą informacji o rdzeniu reakcyjnym. W reakcji Dielsa-Aldera jest on dobrze zdefiniowany.

Zastosowane podejście było możliwe dzięki obfитоści precedensów literaturowych dla tej reakcji. Większość znajdujących się w eksperckiej bazie danych reakcji ujętych w reguły nie posiada aż tak bogatego zbioru precedensów. Wyklucza to możliwość całkowitego zastąpienia zakodowanych reguł reakcji za pomocą uczenia maszynowego, jak wykazano²⁷⁶ w pracy z 2019.

Jak pamiętamy reakcja Dielsa-Aldera była jedną transformacji, na której opierała się LHASA. Również program SESAM wykorzystywał głównie tą reakcję do upraszczania struktury. Niestety wszystkie poprzednie podejścia polegały wyłącznie na znajdowaniu rdzenia reakcyjnego. Percepcja subtelnych efektów mogąca skutkować powstawaniem innego regioizomeru czy diastereoizomeru nie była nigdy inkorporowana do programów wspomagających planowanie syntezy.

2.3.3.3. Heurystyki konformacyjne

Tworząc reguły reakcji, nie zawsze można przewidzieć do jakiego rdzenia zostaną one zastosowane. Czasem można za pomocą warunków ograniczyć rozmiary pierścieni proponowanych jako związki wyjściowe, jednak w ten sposób nie da się wyeliminować wszystkich przypadków generowania dla substratów struktur chemicznie niestabilnych lub konformacyjnie zbyt naprężonych by były stabilne, a co za tym idzie, by mogły być traktowane jako substraty. Przykładem może być reakcja eliminacji prowadząca z pojedynczego wiązania między węglami do wiązania podwójnego. Gdy zostanie zastosowana wobec atomu mostkowego układu pierścieniowego, wówczas prowadzi to do zbyt naprężonych szkieletów łamiących regułę Bredta. Jak pamiętamy, w przykładzie prekursorów twistanonu proponowanych przez program SYNCHEM prekursor generowany wyłącznie w oparciu o regułę reakcji skutkowało propozycją użycia związku zbyt naprężonego. Program SYNCHEM nie miał narzędzi pozwalających rozpoznać że zaproponowany substrat może nie być możliwy do uzyskania. Chematica takim filtrem dysponuje.



Ryc.27. Kilka przykładów z zawartej w Chematicie listy „nonsensownych” motywów strukturalnych; X = halogenki, A = dowolny atom alifatyczny, a = dowolny atom aromatyczny.

W celu usunięcia takich prekursorów z wyszukiwania, stworzono bibliotekę około 600 ogólnie zapisanych „nonsensownych” motywów strukturalnych. Wybrane z nich zaprezentowano na Ryc.27. W przypadkach motywów budzących wątpliwości ich stabilność rozstrzygnięto przy pomocy obliczeń (w ten sposób 10 członowy pierścień zawierający wiązanie potrójne został dopuszczony jako możliwy). Heurystyka ta jest stosowana w momencie generowania substratów w oparciu o regułę reakcji. Jeżeli wygenerowany substrat zawiera choć jeden motyw znajdujący się na liście nie jest on nawet wyświetlany użytkownikowi, ani brany przez algorytm pod uwagę przy konstruowaniu ścieżki. Należy tu zaznaczyć, że nie oznacza to eliminacji reakcji przy użyciu benzynu. Reakcje takie w Chematicie się znajdują, jednak jako związek wyjściowy proponują prekursor benzynu (np. 1,2-dibromobenzen). Pomiędzy prekursorem a produktem cykloaddycji benzyn występuje tylko w kolbie reakcyjnej, ale nie jest kodowany jako osobny motyw. Podobne podejście zastosowano dla innych reakcji biegnących przez nietrwałe związki przejściowe. Dzięki temu Chematica konstruuje plany ze związków stabilnych, niestabilne (acz możliwe do wykorzystania w syntezie) sygnalizując „domyślnie” za pomocą warunków reakcji.

2.3.3.4. Dyskryminacja labilnych grup funkcyjnych

Chcąc tworzyć plany kilkuetapowe, Chematica musi zawierać informacje o tym, jakie podejścia w skali pojedynczego etapu poprawne – w skali sekwencji etapów, nie są praktyczne. Nie sposób oczywiście zawrzeć tej informacji w każdej poszczególniej regule reakcji. Pojedyncza reakcja generująca na przykład chlorek kwasowy, jest jak najbardziej poprawna, jednak propozycja następnego etapu będącego na przykład utlenieniem siarczku w innym miejscu cząsteczki, przy ciągłej obecności chlorku kwasowego, nie jest już dobrą propozycją. Problem nie leży w poszczególnych reakcjach – obie są poprawne – ale już ich połączenie w taką akurat sekwencję nie jest dobrym pomysłem. Jak zatem projektować wykonalne sekwencje unikając jednocześnie tych problematycznych?

W tym celu przeprowadzono²⁷² analizę statystyczną na sieci chemii organicznej (NOC), zawierającej około 7 milionów reakcji organicznych. Badanie miało na celu wytypowanie grup funkcyjnych, które w praktyce potwierdzonej na tych przykładach, po wprowadzeniu do cząsteczki nie pojawiały się już w następnym etapie. Innymi słowy, wprowadzano je tuż przed użyciem w następnym etapie, nie zachowując ich w strukturze dłużej niż to konieczne. W ten sposób usystematyzowano intuicyjną wiedzę praktykujących syntetyków i inkorporowano ją do programu w postaci listy 106 labilnych grup funkcyjnych. Znajdują się na niej na przykład: związki magnezoorganiczne, izocyjaniany, keteny, tioaldehydy, halogenki i azydki acylowe. Dzięki temu plany generowane przez Chematicę, nie zawierają sekwencji przejść praktycznie unikanych. Stanowi to przykład systematyzacji intuicji chemicznej i przekazywania ją programowi, by syntezę projektował jeszcze lepiej.

2.3.3.5. Ocena substratu do cyklizacji

Kolejnym przykładem spojrzenia szerzej niż na jeden etap, jest ocena substratów proponowanych do reakcji cyklizacji. Służy ona głównie wyeliminowaniu sytuacji, gdy najpierw tworzony jest pierścień makrocykliczny, by następnie dokonać jego kontrakcji i utworzyć mniejsze pierścienie. W takiej sytuacji samo uzyskanie makrocyklicznego prekursora może okazać się poważnym wyzwaniem. Sprostanie mu wcale nie gwarantuje, że kontrakcja, w postaci na przykład alkilowania aminy drugorzędowej atomem halogenku obecnego na tym samym pierścieniu, zajdzie w zadowalającym stopniu. Lepsze może się tu okazać zamykanie mniejszych pierścieni za pomocą bocznych łańcuchów alifatycznych o odpowiednich grupach.

Z punktu widzenia reguły reakcji oba powyższe scenariusze są równocenne. Z punktu widzenia praktyki, gdy nie jest niezbędnie konieczne wprowadzanie układu makrocyklicznego to lepiej posłużyć się innymi, prostszymi metodami. Chematica zawiera odpowiednią heurystykę, rozpoznaje ona sytuację gdy pierścień makrocykliczny jest wprowadzony niepotrzebnie i nie poddaje go dalszej analizie.

2.3.3.6. Sekwencje strategiczne

Sekwencje te są kolejnym sposobem rozszerzenia planowania poza horyzont jednego kroku syntetycznego. W perspektywie jednego kroku, gdy ocenie poddawana jest jedna generacja prekursorów, najlepsze oceny zyskują transformacje od razu spełniające kryteria funkcji CSF i RSF. Tym sposobem nie są rozwijane mniej rokujące reakcje. Tymczasem, niejednokrotnie w celu przeprowadzenia wartościowszej transformacji, trzeba się pogodzić z przedłużeniem planu syntetycznego o kilka dodatkowych etapów zmieniającymi strukturę tak, by przeprowadzenie tej zamierzonej transformacji było możliwe. Dobrym przykładem jest tu reakcja Dielsa-Aldera, wykonanie której istotnie zwiększa złożoność struktury. Motywem pozostałym po tej reakcji jest pierścień cykloheksenowy. I tak, poruszając się podczas analizy w kierunku retrosyntetycznym, gdy napotkamy pierścień sześcioczłonowy, możliwość zastosowania do jego stworzenia reakcji Dielsa-Aldera jest zazwyczaj jednym z pierwszych pomysłów. Jednak by rdzeń ten odpowiadał produktowi tej reakcji, musi on być jeszcze uzupełniony o wiązanie podwójne. Należy zatem w kierunku retrosyntetycznym wprowadzić dodatkowy motyw, a taki wzrost skomplikowania struktury nie zostanie wysoko oceniony przez CSF, skutkiem czego atrakcyjna transformacja jest tu izolowana przez etap mniej atrakcyjny. Sekwencje strategiczne służą pomocą w takiej sytuacji. Sumaryczny zysk funkcji, przy łącznym potraktowaniu reakcji Dielsa-Aldera i reakcji przekształcenia wiązania podwójnego (np. jego redukcji lub epoksydowania), uzasadnia włączenie do planu transformacji wcześniej przez funkcję oceniającą nie typowanych. W tym celu stworzono listę sekwencji strategicznych, zawierającą ponad 46 000 kombinacji klas reakcji, to znaczy usztywnionych sekwencji reguł reakcji wartościowych i przygotowujących doń transformacji pomocniczych. Wszystkie ze strategii można przypisać do jednej z trzech klas, w których wzrost złożoności struktury ma na celu: 1) wprowadzenia pierścienia, 2) wprowadzenie centrum stereogenicznego, 3) znaczne uproszczenie struktury.

Podczas wyszukiwania w panelu modułu Syntaurus można zdecydować czy strategię te mają być stosowane, czy też funkcje oceniające mają działać bez tej sugestii.

Przedstawione powyżej przykłady pokazują, że same reguły reakcji nie są wystarczające do tworzenia gotowych planów syntetycznych. Uzupełnianie i korygowanie rezultatów pracy programu Chematica za pomocą wiedzy chemicznej musi odbywać się na różnych poziomach. Dopiero wówczas program zacznie oferować plany syntetyczne warte wykonania.

2.3.4. Rezultaty

Powyżej opisano pokrótce najważniejsze zasady działania Chematici, niezbędne do zrozumienia jaka relacja łączy reguły reakcji chemicznych z planami syntetycznymi. Rezultaty pracy programu były poddawane ocenie od początku prac nad nim. Pierwszym kamieniem milowym było osiągnięcie progu 10 000 reguł reakcji w eksperckiej bazie danych – to niemal dwa razy więcej niż posiadał SYNSUP. Powyżej tego progu zaobserwowano pojawianie się analiz retrosyntetycznych nieoczywistych, lecz teoretycznie poprawnych. Upewniło to twórców, że przyjęte podejście jest poprawne i zwiększenie liczby reguł reakcji przekłada się na jakość i wartość analiz chemicznych

proponowanych przez Chematicę. Pierwsze rezultaty z tego okresu opublikowano w kwietniu 2016 roku²²⁰. Wówczas Chematica odtwarzała już plany syntetyczne znane w literaturze, oraz proponowała przekonujące analizy dla związków nieznanymi.

Zewnętrzna walidację programu rozpoczęto w październiku 2016 roku. Wówczas przedstawiciele firmy Merck KGaA/Sigma Aldrich postanowili przetestować Chematicę. Walidacja polegała na znalezieniu przez program planów syntetycznych dla 6 związków biologicznie aktywnych, których syntezą Merck KGaA był zainteresowany. Należy zaznaczyć, że w czasie gdy przeprowadzano testy, w eksperckiej bazie danych o regułach reakcji znajdowało się około 50 000 reguł (obecnie, w 2019 roku, jest to około 80 000). Związkami wybranymi do testów były: inhibitor białek BRD 7 i 9, α -hydroksyetizolam, inhibitor kinazy ATR, prototypowy lek na białaczkę, (S)-4-hydroksyduloksetyna, 5 β ,6 β -hydroksylurazidon. Za syntezę pierwszych 4 odpowiadali zawodowi chemicy firmy Merck KGaA, syntezę 2 ostatnich wykonali doktoranci IChO PAN, odpowiednio doktorantka Barbara Mikulak-Klucznik i autor rozprawy.

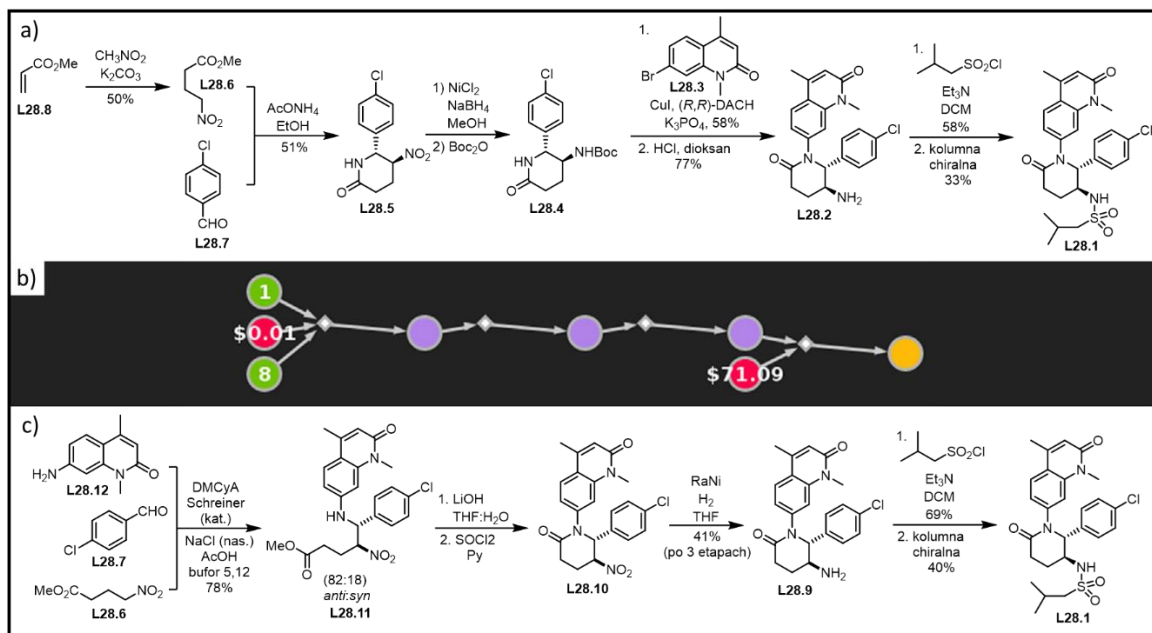
Wybór celów syntetycznych nie był przypadkowy. Wszystkie z nich są związkami o wysokiej wartości rynkowej (>100USD/mg), a dotychczasowe podejścia syntetyczne były albo mało wydajne, albo niewykonalne w większej skali, albo niemożliwe do powtórzenia w oparciu o opublikowane procedury. Ponadto na plany nałożono wymagania: musiały być one znacząco różne od istniejących podejść, substraty musiały być dostępne komercyjnie, planów nie należało zmieniać lecz jak najściślej trzymać się wytycznych zaproponowanych przez program (za wyjątkiem adaptacji warunków, takich jak temperatura, użyty katalizator, rozpuszczalnik), reakcje należało prowadzić od razu w skali setek miligramów lub gramów, celem było uzyskanie ponad 1 grama produktu, który musiał mieć czystość HPLC większą niż 98% oraz żadne jego pojedyncze zanieczyszczenie nie mogło być większe niż 0,5%.

Test ten wykonano na przełomie lat 2016/2017. Rezultaty były zaskakujące nawet dla twórców programu, bowiem udało się przeprowadzić syntezę wszystkich związków wybranych do testów, nawet tych, których syntezę nie były wcześniej znane, i których – pomimo prób – Merck KGaA nie był w stanie opracować. Co więcej kryteria przyjęte przez firmę zostały spełnione, a w porównaniu z istniejącymi syntezami, te pochodzące z Chematici, zostały przez zewnętrznych testerów jednoznacznie uznane za lepsze od tych wymyślonych i wykonanych wcześniej przez człowieka. Jest to pierwszy taki rezultat w historii prac nad komputerowo wspomaganą syntezą organiczną. Wprawdzie podejmowano wcześniej nieliczne próby walidacji eksperymentalnej, a przykłady te omówiono we wcześniejszych rozdziałach, lecz nigdy wcześniej plan „komputerowy” nie stanowił poważnej konkurencji dla chemika. Poniżej omówię przykłady planów i syntez przeprowadzonych w ramach testu przez chemików firmy Merck KGaA. Część tego testu wykonaną przeze mnie omawiam jako badania własne.

2.3.4.1. Synteza inhibitora białek BRD 7 i 9

Synteza związku **L28.1** była wcześniej znana w literaturze²⁸⁵ (Ryc.28a), jednak budziła zastrzeżenia chemików Merck KGaA ze względu na długość (8 etapów) i kilkukrotnie (5) stosowaną chromatografię kolumnową. Do tego całkowita wydajność wynosząca około 1% nie pozwalała nazwać tej metody opłacalną. Niska wydajność spowodowana była, między innymi, niską enancjoselektywnością i diastereoselektywnością.

Plan stworzony przez Chematicę²⁷², zakładał wykorzystanie trójkomponentowej reakcji aza-Henry'ego **L28.7+L28.7+L28.12**→**L28.11**. Ma ona kilka zalet w porównaniu z reakcją **L28.6+L28.7**→**L28.5** ze znanego planu. Znany plan zawierający kilka przemian w jednym kroku wydaje się atrakcyjny, bo tworzona imina po ataku karboanionu związku nitrowego spontanicznie cyklizuje do pierścienia laktamu. Po tym kroku wystarczy przyłączyć fragment chinolin-2(1H)-onu



Ryc.28. Ilustracja walidacji eksperymentalnej modułu retrosyntetycznego programu Chematica na przykładzie inhibitora białek BRD 7 i 9; a) synteza znana wcześniej w literaturze; b) plan wygenerowany przez program Chematica; c) pomyslna realizacja syntetyczna komputerowo wygenerowanego planu syntetycznego.

i sulfonamidu. Zastosowanie iminy nie zawierającej zawady sterycznej na atomie azotu ma zaletę (cyklizacja do laktamu zachodzi spontanicznie), ale też i wadę (brak czynnika sprzyjającego atakowi prowadzącemu do pozycji *anti*, co skutkuje niskimi nadmiarami diastereoizomerycznymi).

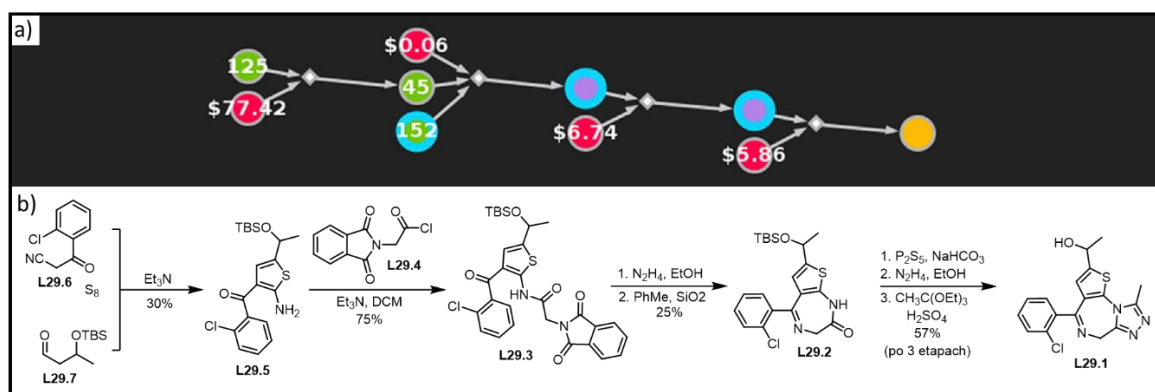
Plan Chematici inaczej podchodzi do tej reakcji – wprowadzie zamknięcie pierścienia laktamu jest teraz realizowane w 2 osobnych etapach (hydroliza estru i utworzenie chlorku kwasowego), jednak od razu do struktury wprowadzany jest fragment chinolin-2(1H)-onu. Dzięki jego łatwej do uzyskania pochodnej aminowej, wchodząc w skład iminy, skutecznie wpływa on na nadmiar diastereoizomeryczny dzięki wymuszonej sterycznie preferencji tworzenia *E*-izomeru iminy. Wprowadzenie tego fragmentu na wcześniejszym etapie pozwala uniknąć umiarkowanie wydajnego jego wprowadzenia do amidu **L28.4**+**L28.3**→**L28.2**. Pozostałe etapy mają już porównywalną wydajność i polegają głównie na zmianach grup funkcyjnych i przyłączeniu do rdzenia cząsteczki grup bocznych właściwych dla celu syntetycznego.

Podsumowując, dzięki zastosowaniu planu Chematici, skrócono syntezę do 6 etapów, a jej wydajność wzrosła 6-8 razy – jest to wyraźna poprawa w porównaniu do dotychczasowego planu.

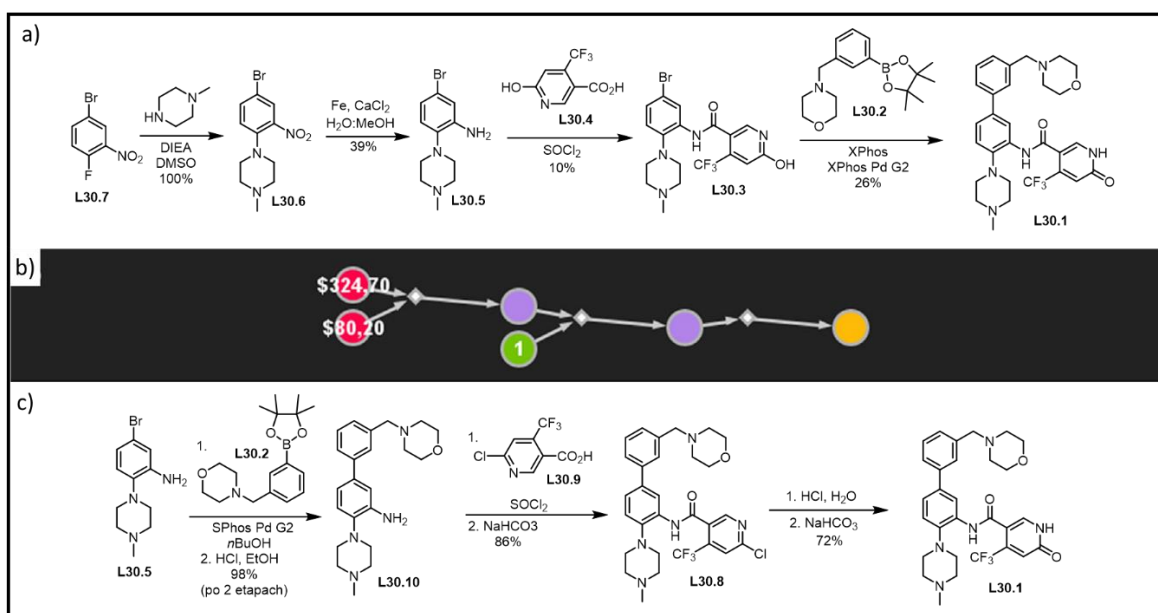
2.3.4.2. Synteza α -hydroksyetizolamu

Związek ten jest jednym z dwóch głównych metabolitów leku etizolam²⁸⁶ (benzodiazepina o właściwościach przeciwlękowych, przeciwdrgawkowych i nasennych). Wprawdzie syntezy szkieletów benzodiazepinowych, są już dobrze zbadane i opisane²⁸⁷⁻²⁸⁸, jednak brak było dotychczas metod syntezy ich hydroksylowych pochodnych. Dotychczasowe próby uzyskania związku **L29.1** przez chemików Merck KGaA, skończyły się niepowodzeniem. Główną przyczyną była tu łatwość eliminacji grupy hydroksylowej do wiązania podwójnego sprzężonego z układem aromatycznym.

Chematica i tym razem oparła plan (Ryc.29a) o reakcję multikomponentową, zidentyfikowaną jako najbardziej uproszczającą strukturę. Na pierwszym etapie zaproponowała reakcję Gewalda w celu syntezy pierścienia aminotiofenowego. Należy nadmienić, że nie są znane przykłady tej reakcji dla aldehydów posiadających w pozycji β zabezpieczoną grupę hydroksylową.



Ryc.29. Ilustracja walidacji eksperymentalnej modułu retrosyntetycznego programu Chematica na przykładzie α -hydroksyetizolamu; a) plan wygenerowany przez program Chematica; b) pomyślna realizacja syntetyczna, komputerowo wygenerowanego planu syntetycznego.



Ryc.30. Ilustracja walidacji eksperymentalnej modułu retrosyntetycznego programu Chematica na przykładzie prototypowego leku na białaczkę; a) synteza znana wcześniej w literaturze; b) plan wygenerowany przez program Chematica; c) pomyślna realizacja syntetyczna komputerowo wygenerowanego planu syntetycznego.

Tutaj udało się taką reakcję przeprowadzić. Jako substrat zidentyfikowano 3-hydroksybutyraldehyd, niebieską obwódką informując o konieczności zabezpieczenia grupy hydroksylowej. Uzyskany **L29.5** następnie acylowano i po usunięciu grupy ftalimidowej, następowała cyklizacja do rdzenia benzodiazepiny. Ostatni etap zamyka zarówno pierścień triazolowy, jak i zachodzi wówczas usunięcie grupy zabezpieczającej alkohol benzyłowy.

Podsumowując, zastosowanie trójkomponentowej reakcji Gewalda dla rdzenia aminotiofenu było wyborem trafnym, ponadto kompatybilność warunków reakcji z grupą zabezpieczającą alkohol, pozwoliło na pierwsze syntetyczne uzyskanie tego metabolitu z całkowitą wydajnością 3,2%.

2.3.4.3. Synteza prototypowego leku na białaczkę

Ostatnim z tu omówionych związków będących przedmiotem zainteresowania chemików firmy Merck KGaA, był prototyp leku na białaczkę²⁸⁹. Związek ten hamuje proliferację i podział komórek nowotworowych, ostrej białaczki szpikowej. Dotychczasowa metoda jego syntezy wynosiła około 1% i pomimo wysiłków włożonych w optymalizację syntezy nie udało się jej podnieść. Podobnie wysiłki zmierzające do zwiększenia skali syntezy, okazały się bezowocne. Niska wydajność trzech ostatnich etapów **L30.6**→**L30.5**→**L30.3**→**L30.1** dawała duże pole do poprawy.

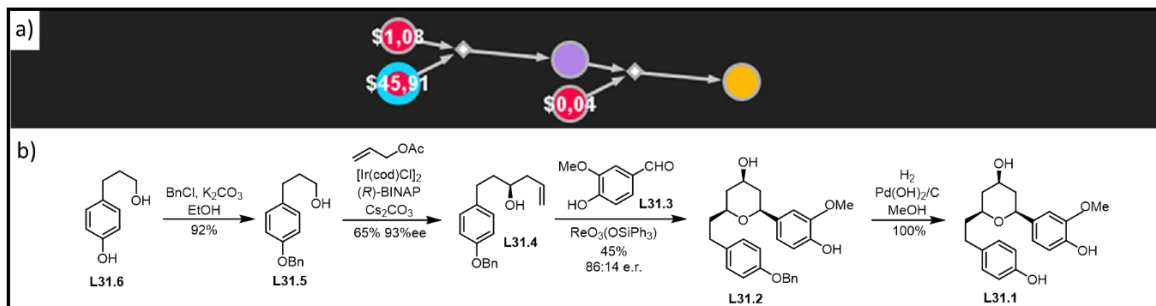
Co ciekawe, plan Chematici (Ryc.30a) używa tego samego, znanego w literaturze, dwupierścieniowego prekursora **L30.5**, co znany plan. Jednak między **L30.5**, a produktem **L30.1**, proponuje inne transformacje i w innej kolejności. Zamiast najpierw wprowadzić pierścień pirydynonu, a następnie przyłączyć fragment aryłomorfoliny. Chematica zaproponowała wprowadzenie aryłomorfolinowego fragmentu za pomocą reakcji Suzuki. Obecność grupy aminowej nie stanowi przeszkody w tej reakcji, a produkt oczyszczany jest dzięki krystalizacji. Następnie z chlorowodoru, w warunkach acylowania Schottena-Baumann, tworzony jest amid z pirydynowego prekursora pierścienia pirydynonowego, jednak innego (chlorek zamiast grupy hydroksylowej) niż w znanym planie. Ostatni etap to właśnie hydroliza ów chloropirydynowego prekursora **L30.8**.

Dzięki zastosowaniu kosmetycznych zdawałoby się zmian, całkowitą wydajność syntezy podniesiono z 1% do 60%. Co więcej, synteza ta odbywa się bez użycia oczyszczania za pomocą chromatografii kolumnowej, a kluczowe związki przejściowe, udaje się skutecznie krystalizować. Między innymi dzięki temu plan ten wykonano w skali gramowej.

2.3.4.4. Synteza engelheptanoksydu C

W pracy²⁷² oprócz opisu walidacji syntetycznej za pomocą związków będących obiektem zainteresowania firmy Merck KGaA, opublikowano również inne przykłady walidacji eksperymentalnej programu Chematica. Takim związkiem testowym był związek naturalny, engelheptanoksyd C²⁹⁰, wyizolowany z *Engelhardia roxburghiana*, nigdy wcześniej nie uzyskany syntetycznie. Jego syntezę (Ryc.31b) wykonała współpracująca z grupą prof. Grzybowski, grupa prof. Mrksicha z Uniwersytetu Northwestern.

Plan Chematici, zakłada wykorzystanie reakcji Prinsa do stworzenia pierścienia tetrahydropiranowego. Pierścień ten zawiera trzy centra stereogeniczne, z czego aż dwa z nich



Ryc.31. Ilustracja walidacji eksperymentalnej modułu retrosyntetycznego programu Chematica na przykładzie engelheptanoksydu C; a) plan wygenerowany przez program Chematica; b) pomyślna realizacja syntetyczna komputerowo wygenerowanego planu syntetycznego.

tworzone są w tej reakcji, pod kontrolą jednego centrum już istniejącego. To jedno pozostałe centrum stereogeniczne zostało wprowadzone za pomocą enancjoselektywnego allilowania **L31.5**→**L31.4**, według metodologii Kirche'go²⁹¹⁻²⁹².

Jest to pierwszy znany przykład syntezy engelheptanoksydu C. Jest to kolejny przykład, gdy Chematica zaproponowała plan syntezy dla nieznanego związku, owoconie zrealizowany w laboratorium.

Przedstawione powyżej plany syntetyczne, warto porównać z planami znanymi z innych programów do komputerowo wspomaganey analizy retrosyntetycznej. Jak przedstawiono wcześniej, planów walidowanych syntetycznie było zaledwie kilka. Dlatego porównywać będę różnorodność ich percepcji chemicznej.

Operowanie elementami stereogenicznymi, stanowi niezbędny wymóg dla programu mającego projektować plany syntetyczne. Z wcześniej przedstawionych programów, jedynie LHASA w ograniczonym zakresie (tylko jedna reakcja asymetryczna – kondensacja aldolowa dla prekursora erytronolidu B, Ryc.8.) i SYNSUP (chloramfenikol, Ryc.15.) operowały reakcjami asymetrycznymi. Jedyne dotychczasowy przykład wykonanej asymetrycznej syntezy pochodził z programu AIPHOS/KOSP (Ryc.19b), gdzie w syntezie DU-6859a, występuje asymetryczna redukcja ketonu, a w następnych etapach inwersja konfiguracji na tym centrum stereogenicznym. Na tym kończą się przykłady syntez stereoselektywnych znanych w literaturze przed Chematicą. Nawet współczesne podejścia (ARChem, ICSynth, sieć neuronowa Seglera-Wallera), ograniczają się do konstruowania rdzenia cząsteczki, zupełnie pomijając aspekt stereochemiczny.

Moduł retrosyntetyczny Chematici od początku dysponował percepcją stereochemii. Posiada w swej bazie danych reguły reakcji tworzące centra stereogeniczne, jak i reguły reakcji wykorzystujące indukcję pochodzącą od centrów stereogenicznych już obecnych w cząsteczce. Dobrym na to przykładem jest zaprezentowana synteza Engelheptanoksydu C (Ryc.31.), zawiera ona zarówno reakcję tworzenia centrum stereogenicznego (enancjoselektywne allilowanie), jak i reakcję Prinsa pod kontrolą istniejącego centrum stereogenicznego. Kolejnym przypadkiem jest omówiona w pracy²⁷² synteza (*S*)-4-hydroksyduloksetyny. W syntezie tej znajdujemy enancjoselektywną redukcję ketonu, a następnie inwersję konfiguracji w reakcji Mitsunobu. Syntezy bardziej zaawansowanych związków naturalnych, zawierających więcej centrów stereogenicznych i bardziej zaawansowane przykłady indukcji asymetrycznej, zostały już wykonane w oparciu o plany Chematici, a w momencie pisania tej rozprawy praca je opisująca jest recenzowana.

Można zatem przyjąć, że poprawna percepcja stereochemii oraz swobodne wykorzystywanie transformacji operujących na elementach stereochemicznych, stanowi dobre kryterium, pozwalające oddzielić programy zaawansowane od tych wymagających jeszcze uzupełnienia.

Innym godnym uwagi kryterium porównania, jest operowanie na związkach aromatycznych, to znaczy: selektywne wprowadzanie podstawników, wykonywanie funkcjonalizacji w określonej kolejności, czy tworzenie aromatycznych pierścieni heterocyklicznych. Dla LHASA w zakresie funkcjonalizacji związków aromatycznych istnieje wyłącznie jeden przykład planu zawierającego takie reakcje. Pochodzi on z późnego okresu rozwoju programu¹⁸ i zakłada sekwencję podstawień toulenu. Brak dla LHASA przykładów tworzenia aromatycznych pierścieni heterocyklicznych. Ciekawym przypadkiem jest tu SYNGEN, bowiem w opublikowanych dlań planach (eter slafradiolu Ryc.11., prekursor alkaloidu kory chinowca¹²⁵) występuje pierścień pirydyny, plany zaczynają się jednak od odpowiednio podstawionych prekursorów, a żaden nowy nie jest wprowadzany (z wyjątkiem jednej reakcji Grignarda). Jedyne SYNSUP już na wczesnym etapie prezentował szereg planów przedstawiających funkcjonalizację pierścieni aromatycznych. Chematica w przedstawionych planach zarówno tworzy różne pierścienie heterocykliczne, jak i przedstawia ich funkcjonalizację.

Reakcja Dielsa-Aldera, spośród bardziej zaawansowanych programów planujących syntezę z pewnością zawarta była w LHASA i w SYNCHEM (twistanon Ryc.11.). Dla pozostałych programów brak jest przykładów stosowania tej reakcji. W porównaniu jednak z Chematicą, LHASA wymagała by pierścień sześcioczłonowy tworzony w reakcji Dielsa-Aldera był obecny w strukturze końcowej. Ograniczenia tego wymogu omówiono na przykładach z Ryc.2. – Chematica takiego ograniczenia nie nakłada.

W przypadku klasycznych programów, rzadko spotykamy się z heurystykami korygującymi syntezę. Jednym z wyjątków może być SYNSUP z jego typowaniem wiązań strategicznych co przełożyło się na znaczne skrócenie czasu wyszukiwania. Również LHASA dla ograniczonej puli grup funkcyjnych potrafiła rozpoznać grupy wymagające ochrony. Jak dużą rolę w Chematicie odgrywają heurystyki, zostało wcześniej omówione. Brak większej liczby miejsc korygowania rezultatów w pozostałych programach, może tłumaczyć dlaczego pomimo dekad rozwoju plany przezeń proponowane nie znajdowały większego uznania.

2.3.5. Podsumowanie

After over a decade of laborious development, Chematica is finally capable of designing novel and experimentally efficient syntheses of medicinally and industrially relevant targets.

B.A. Grzybowski, 2018²⁷²

Jak dotąd Chematica jest pierwszym i jedynym programem komputerowym zdolnym generować chemicznie poprawne i użyteczne plany syntetyczne²²⁰. Dotychczas zostało to wykazane eksperymentalnie na zbiorze 10 związków chemicznych (leków, metabolitów leków, prototypów leków i związków naturalnych). Wszystkie podjęte próby syntetyczne zwieńczone zostały sukcesem. Związki, które wcześniej nigdy nie zostały wykonane, udało się uzyskać syntetycznie po raz pierwszy. Jest to mocne potwierdzenie skuteczności przyjętej zasady działania.

Oprócz zaprezentowanych związków o prostej strukturze, dotychczas zakończono już dwie syntezy totalne związków naturalnych o skomplikowanej strukturze (takamonidyna, daurycyna) w oparciu o plany syntetyczne wygenerowane przez program Chematica. Wyniki te jednak stanowią przedmiot rozpraw doktorskich innych członków zespołu. W skład niniejszej wchodzi fragment syntezy totalnej takamonidyny.

Połączenie bezprecedensowo obszernej bazy danych o regułach reakcji organicznych z heurystykami korygującymi generowane plany i algorytmami poszukującymi planów najlepiej spełniających funkcje – okazało się połączeniem gwarantującym uzyskanie założonego rezultatu.

Program po wykazaniu jego użyteczności, został nabyty przez firmę Merck KGaA, dalszy jego rozwój odbywa się we współpracy z nią.

3. Badania własne

3.1. Plan weryfikacji hipotezy badawczej

Hipoteza badawcza, stojąca u podstaw moich badań brzmi: „Zdolność projektowania wykonalnych planów syntetycznych, cząsteczek organicznych nie jest umiejętnością wyłącznie ludzką, lecz możliwe jest stworzenie programu komputerowego zdolność tą posiadającego.”

Weryfikacja tej hipotezy wymagała stworzenia programu komputerowego generującego plany syntetyczne. Oczywiście stworzenie takiego programu komputerowego jest zadaniem niewykonalnym dla jednej osoby. W rozdziale niniejszym, opiszę swój wkład w tworzenie programu Chematica, w szczególności w tworzenie jego eksperckiej bazy danych.

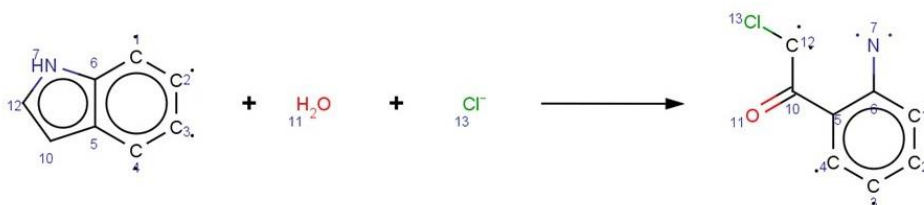
Następnym etapem weryfikacji hipotezy badawczej było wykonanie syntez według planów uzyskanych z Chematici. Jak opisano powyżej, w literaturze przedmiotu istnieją jedynie incydentalne przykłady syntetycznej walidacji komputerowo generowanych planów. Ponadto wszystkie one budzą zastrzeżenia i, jak wspomniano, nie można ich nazwać użytecznymi w porównaniu z istniejącymi syntezami. Nigdy też dla jednego programu nie wykonano więcej niż jednej walidacji syntetycznej. Dlatego systematyczna weryfikacja poprawności i wykonalności komputerowych planów syntetycznych była oczywistym i niezbędnym etapem weryfikacji postawionej hipotezy. Ta część walidacji była głównym zadaniem mojego doktoratu.

Oczywiście specyfika projektu umożliwiała nakładanie się pierwszego (tworzenie programu) i drugiego (syntetyczna weryfikacja) etapu prac. Dlatego pomimo dalszego rozwoju programu i powiększaniu bazy transformacji chemicznych, możliwe było równoczesne prowadzenie syntez.

3.2. Tworzenie bazy reguł reakcji chemicznych

W latach 2013-2016 zajmowałem się kodowaniem reguł transformacji chemicznych. Wszystkie kodowane reguły reakcji, były zapisywane w języku SMARTS, a każda reguła reakcji, zawarta w eksperckiej bazie danych Chematici, zawiera informacje o: 1) nazwie reakcji; 2) atomach rdzenia reakcji niezbędnych do zajścia transformacji; 3) skodyfikowanych ogólnych warunkach transformacji (np. LDA, Et₃N), których w momencie kodowania reguł było 108; 4) najpopularniejszych warunkach dla tej transformacji występujących w literaturze; 5) numerach DOI publikacji opisujących wykorzystanie transformacji (zazwyczaj umieszczano 1-4 odnośniki); 6) liście grup funkcyjnych niekompatybilnych z warunkami reakcji bądź grupami funkcyjnymi występującymi w rdzeniu reakcyjnym, o niekompatybilności możliwej do zniesienia za pomocą grup zabezpieczających; 7) liście grup funkcyjnych niekompatybilnych z warunkami reakcji bądź grupami funkcyjnymi występującymi w rdzeniu reakcyjnym, o niekompatybilności niemożliwej do zniesienia za pomocą grup zabezpieczających. Do tego reguła reakcji zawierała jeszcze między innymi kategorie: *Synthon count.*, *Explicit hydrogen*, *Corey 3 hetero*, *Useful rate*, *Diastereoselectivity*. *Useful rate*, był subiektywnie przypisywaną miarą użyteczności w skali 1-10. Pozostałe kategorie zazwyczaj nie miały kluczowego znaczenia przy tworzeniu reguły reakcji: gdy reakcja wykazywała diastereoselektywność, było to zaznaczane za pomocą odpowiedniej kategorii, lecz dla reakcji bez elementów stereogenicznych kategoria ta nie miała znaczenia. Również *Explicit hydrogen*, znajdował zastosowanie jedynie w przypadku szczególnie trudnych rdzeni reakcyjnych, gdzie nie wszystkie protony mogły być dopisywane automatycznie do wolnych walencyjności rdzenia reakcji, lecz musiały być szczególnie wyspecyfikowane. Poniżej omówię wybrane przykłady zakodowanych przeze mnie reguł reakcji.

Name: Sugasawa indole synthesis



Rxn SMARTS:

[c:1]1[c:2][c:3][c:4][c:5]2[cH:10][cH:12][nH:7][c:6]12.[OH2:11].[Cl-:13]

>>[c:1]1[c:2][c:3][c:4][c:5]([C:10](=[O:11])[C:12][Cl:13])[c:6]1[N:7]

Products: [c]1[c][c][c]2[cH][cH][nH][c]12.[OH2].[Cl-]

SMILES: Clc1ccc2[nH]ccc2c1.O.[Cl-]

Synthon count: 1

Explicit H: False

Reference: DOI: 10.1021/jo01318a021

Condition: NaBH4.dioxane.H2O.reflux

Corey 3 hetro: True

Groups to protect:

[#6][CH]=O.[#6]C([#6])=O.[#6][SX2H].[CX4,c][NH2].[CX4,c][NH][CX4,c]

Conditions number: 21, 47

Incompatibile groups:

[CX4][Cl,Br,I].[#6]O[OH].[#6]OO[#6].[#6]C(=[O])OC(=[O])[#6].

[#6]N=C([O,S].[#6]C#N.[#6][N+]#[C-].[#6]C(=O)[Cl,Br,I].

[#6][SX2][SX2][#6].[CX3]=[NX2][*!O].[CX3]=[NX2][O].

[#6]C(=[SX1])[#6].[#6][CH]=[SX1].[#6][SX3](=O)[OH].

[CX4]1[O,N][CX4]1.[#6]=[N+]=[N-].c[N+]#[N]

Useful rate: 8

Ryc.32. Reguła reakcji tworzenia indoli metodą Sugasawy, zakodowana przeze mnie 4.02.2014

Tworzenie pierścienia indolowego metodą Sugasawy²⁹³ jest przykładem reakcji o dobrze zdefiniowanym rdzeniu. Zapis Rxn SMARTS stanowi sedno reguły reakcji, a jego ręczny zapis był najbardziej czasochłonnym etapem tworzenia reguły. W liniowym zapisie SMARTS, liczba atomów po obu stronach transformacji musi być jednakowa. Dzięki temu reguła reakcji opisuje stan każdego atomu przed reakcją i po niej. W ten sposób reakcja chemiczna sprowadzana jest do zmiany wiązań chemicznych między atomami rdzenia. Zauważmy, że taki sposób opisu nie potrzebuje wykorzystywać pojęcia ruchu elektronów, a mechanizm reakcji zostaje sprowadzony do dwóch stanów: przed i po reakcji. Synton i retron oddzielone są za pomocą „>>”, co traktowane jest jako strzałka retrosyntetyczna. Na Ryc.32. wyświetla się ona jako pojedyncza strzałka, jest to skutek stosowanego wyświetlacza struktur. Nie należy reguły reakcji traktować jako przykładu reakcji chemicznej. Reguła reakcji jest ogólniejszym sposobem potraktowania transformacji chemicznej. Zawiera elementy niezbędne do zajścia reakcji, ale w ten sposób by pozostawić otwarte możliwości strukturalne w miejscach gdzie różne podstawniki mogą się znajdować, nie zakłócając przebiegu reakcji. Dzięki takiemu otwartemu zapisowi, reguła obejmuje zarówno znane w literaturze przypadki zastosowania reakcji, lecz również dopuszcza zajście reakcji dla substratów, dla których nigdy jej nie wykonano. Lecz nie wykonano nie dlatego, bo było to niemożliwe, lecz dlatego, że prawdopodobnie nikt wcześniej nie potrzebował danej cząsteczki na wybranej drodze uzyskać.

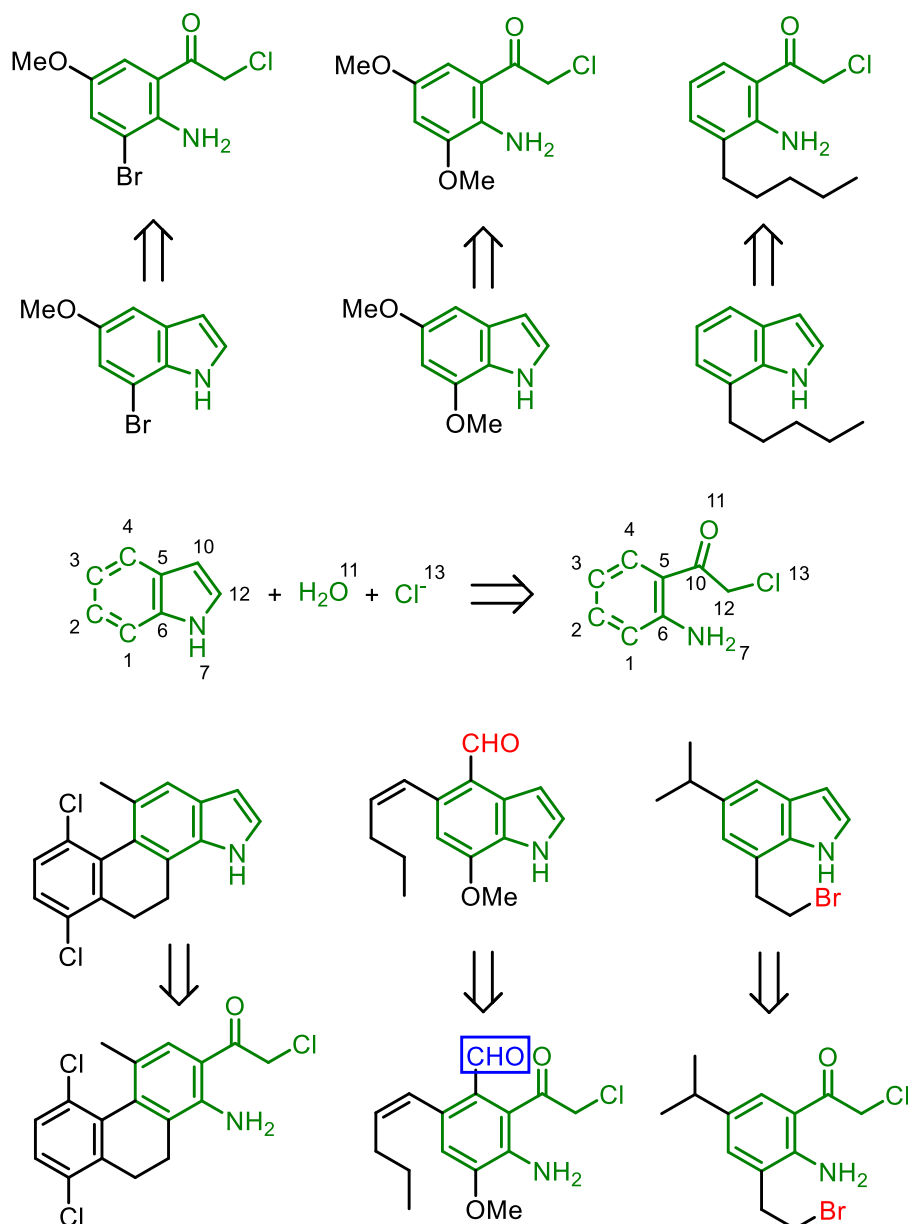
Zachowanie i balansowanie tej potencjalności jest najtrudniejszym etapem tworzenia reguły reakcji. Gdy znane są setki przykładów dla kodowanej reakcji, wówczas studia literaturowe pozwalają precyzyjnie określić jej granice stosowalności. Lecz gdy kodowałem reakcje nie posiadające wielu precedensów, często stawałem często przed dylematem czy zawęzić jej zakres stosowalności – wówczas zyskiwałem większą pewność co do poprawności zapisu, jednak tym samym wykluczałem potencjalne struktury retronów, dla których mogłaby być zastosowana. To balansowanie i ocena chemika stanowią unikalną wartość eksperckiej bazy danych Chematici, gdyż za każdą decyzją o dopuszczeniu szerszego bądź węższego zakresu stosowalności reguły, stał ostrożny ludzki namysł.

Do zakodowania niektórych reakcji, często potrzeba było stworzyć więcej niż jedną regułę reakcji. Dla niektórych przypadków wymagane były setki linii. Wszystko po to by nie pozostawić zapisu zbyt ogólnym. Jednak w dalszym ciągu każdy z tych setek zapisów nie odpowiadał wyłącznie jednej cząsteczce.

Zapis prostych reguł reakcji może przypominać zapis samych reakcji. Dla przypadku reakcji Sugasawy, można ulec takiemu złudzeniu. Jednak kierunek (i znaczenie) zapisanej strzałki oraz zapis niektórych atomów odróżnia je od siebie. W powyższym przykładzie atomy 1-4 retronu posiadają otwarte możliwości wiązania. To oznacza, że reguła ta będzie aplikowana za każdym razem, gdy podczas rozwijania drzewa retrosyntetycznego program napotka pierścień indolowy. Dotyczyć to będzie wszystkich pierścieni indolowych, bez względu na podstawnik/podstawniki znajdujące się na atomach 1-4. Ze względu na stosowany substrat, uzyskiwane indole na atomach 10 i 12 posiadają atomy wodoru. Tam nie występuje dowolność podstawnika. Zostało to wyrażone zapisem: [cH:10][cH:12] specyfikującym podstawniki (H), zamiast zapisu [c:10][c:12] pozwalającego na dowolny podstawnik. Oczywiście dowolny podstawnik na atomach 1-4 może oznaczać grupę aldehydową, niekompatybilną ze względu na warunki (warunki 21, 47 oznaczają odpowiednio grupę aminową i NaBH_4). Wówczas reguła ta nadal zostanie zastosowana dla takiego przypadku, lecz konflikt reaktywności zostanie rozpoznany. Zestawienie listy grup niekompatybilnych, warunków reakcji i zaaplikowanie ich do konkretnego badanego przypadku, da w rezultacie informację o konieczności ochrony niekompatybilnej grupy funkcyjnej. Pozwala to podczas kodowania pozostawiać otwarte walencyjności i szerszą pulę możliwości stosowania reguły, bez obawy o to, że reguła zostanie zastosowana do cząsteczki posiadającej grupy funkcyjne przeszkadzające w zajściu reakcji, lub ulegające niechcianej reakcji ubocznej.

Przykłady zastosowania tak skonstruowanej reguły reakcji zaprezentowano na Ryc.33. Reguła reakcji zapisana w środku definiuje rdzeń reakcji (oznaczony na zielono). Trzy górne przykłady pochodzą z doniesień literaturowych²⁹⁴⁻²⁹⁶ – są to przypadki indoli posiadających podstawniki (wykraczają one poza zapisany rdzeń reguły i oznaczone są na czarno), pozostawienie nieprzypisanych walencyjności dla atomów 1-4 pozwala zastosować regułę reakcji zarówno do przypadków znanych jak i nieznanymi w literaturze (dolne przykłady), acz chemicznie prawdopodobnych. Wprawdzie Chematica posiada bazę danych związków znanych w literaturze i komercyjnie dostępnych, jednak podczas stosowania reguły reakcji są one traktowane jednakowo. Zilustrowana na dolnych przykładach, identyfikacja niekompatybilnych grup funkcyjnych (dolny środkowy przypadek dla aldehydu, możliwego do zabezpieczenia i dolny prawy przypadek bromku niemożliwego do zabezpieczenia), zostały omówione powyżej. Znalezione przypadki możliwości zastosowania reguły dla cząsteczek posiadających grupy niekompatybilne w manualnym trybie retrosyntetycznym zostają użytkownikowi wyświetlane. Oznaczone są wówczas niebieskim kółkiem gdy zabezpieczenie grupy jest możliwe (jak dla aldehydu w przypadku powyżej) lub pomarańczowym kółkiem, gdy zabezpieczenie takiej grupy jest niemożliwe.

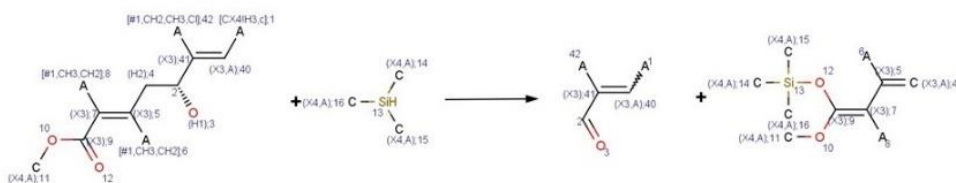
Przykładem bardziej zaawansowanym jest reguła dla organokatalitycznej winylogowej reakcji aldolowej Mukaiyamy (Ryc.34.). Pojawiają się w niej elementy strukturalne sprawiające problemy dotychczasowym programom: centrum stereogeniczne i wiązania podwójne o określonej



Ryc.33. Reakcja Sugasawy. Ilustracja reguły reakcji – na środku. Górne trzy przykłady to wybrane precedensy literaturowe zastosowania tej reakcji²⁹⁴⁻²⁹⁶. Dolne trzy przykłady reakcji nigdy nie zostały wykonane. Na zielono zaznaczono rdzeń reakcji, na czarno fragmenty strukturalne dopuszczalne w nieprzyzpisanych walencyjnościach, na czerwono zidentyfikowane w oparciu o regułę reakcji grupy wykazujące niekompatybilność, na niebiesko informacja, że grupę niekompatybilną można zabezpieczyć.

konfiguracji. Z achiralnych substratów, pod wpływem zewnętrznego źródła chiralności (katalizatora), generowane jest centrum stereogeniczne przy atomie [C@H:2]. Określoną konfigurację centrum stereogenicznego poznajemy po znaku „@”. Do zapisu konfiguracji stosowany jest zapis „@” lub „@@”, jednak nie istnieje ścisła korelacja między takim zapisem, a konfiguracjami R/S. Co więcej, w przypadku związków o kilku centrach stereogenicznych, centrum o konfiguracji S w substracie może być oznaczone „@”, a pomimo braku jego inwersji w produkcie może już być wymagany dla niego zapis „@@”. Fakt ten zwiększa trudność kodowania reguł reakcji stereospecyficznych i stereoselektywnych. Podobnie sprawa się ma z sekwencjami

Name: Organocatalysed vinylogous Mukaiyama aldol reaction



Rxn SMARTS:

```
[CX4!H3,c:1]/[CX3:40]=[CX3:41](\[#1,CH2,CH3,Cl:42])[C@H:2]([OH:3])[CH2:4][CX3:5](\[#1,CH3,CH2:6])=[CX3:7](/[#1,CH3,CH2:8])[CX3:9]([O:10][CX4:11])=[O:12].[Si:13]([CX4:14])([CX4:15])([CX4:16])>>[*:1][CX3:40]=[CX3:41]([*:42])[C:2]=[O:3].[CX3:4]=[CX3:5]([*:6])[CX3:7](/[*:8])=[CX3:9]([O:10][CX4:11])[O:12][Si:13]([CX4:14])([CX4:15])([CX4:16])
```

Products:

```
[CX4!H3,c]/[CX3]=[CX3](\[#1,CH2,CH3,Cl])[C@H]([OH])[CH2][CX3](\[#1,CH3,CH2])=[CX3](/[#1,CH3,CH2])[CX3]([O][CX4])=[O].[Si]([CX4])([CX4])([CX4])
```

SMILES:

```
COC(=O)\C=C\C[C@@H](O)C(\C)=C\c1cccc1.C[SiH](C)C(C)(C)C
```

Synthon count: 2

Explicit H: True

Reference: 10.1002/anie.201005954 and 10.1021/jm050422p

Condition: 1. organocatalyst.Et2O.-78C 2. py-HF.THF

Corey 3 hetro: True

Groups to protect:

```
[#6][CH2][OH].[c][OH].[OH][c][c][OH].[#6][CH]=O.[#6][SX2H].[OH][CX4][CX4][OH].[OH][CX4]C[CX4][OH].[CX4,c][NH2].[CX4,c][NH][CX4,c]
```

Conditions number: 106

Incompatibile groups: [#6][CH2][I].[#6][N+]#[C-

```
].[#6]N=C=[O,S].[#6]C(=O)[Cl,Br,I].[#6][CH]=[SX1].[#6]C(=[SX1])[#6].[#6]S(=O)[Cl,Br,I].[CX4]1[O,N][CX4]1.[Si][CX4,CX3].[#6][SX2][SX2][#6].[#6]C(=O)OC(=O)[#6].[#6]C(=O)[N]=[N+]=[N-].[#6]O[OH].[#6]OO[#6].[CX3]=[NX2][*!O].[CX3]=[NX2H].[CX3]=[NX2][O].Cl=C.N.[CX4,c][NX3][NH2].c[N+]#[N].[#6][Li].[#6][Mg][*].[#6][Zn][*].[#6][SX2,O]C#N.[#6][NX2]=O.[CX3]=[CX2]=[CX3,CX2].[CX3]=[CX2]=O.[#6][NX3][OH].[CX3]=[CX3][OH].[OH][CX4][O]
```

Useful rate: 6

Ryc.34. Reguła organokatalitycznej, winylogowej reakcji aldolowej Mukaiyamy, zakodowana przeze mnie w tygodniu 19-25.10.2015

wiązań podwójnych. Oznaczone są one za pomocą znaków „/” (konfiguracja E) lub „\” (konfiguracja Z) umieszczanych przed i po atomach tworzących wiązanie podwójne. Mając jednak do czynienia z kilkoma następującymi po sobie, sprzężonymi wiązaniami podwójnymi, zapis się komplikuje i prosta zależność „/” dla konfiguracji E i „\” dla Z – przestaje obowiązywać. Ponadto, nie istnieje też ścisła zależność między oznaczeniem „/” i „\” a rodzajem podstawników, to znaczy

ich priorytetacją w oparciu o reguły Cahn-Ingolda-Preloga. Mając do czynienia z wiązaniem podwójnym o czterech podstawnikach decydujące jest, który z nich zostanie włączony do łańcucha głównego, a który wytypowany jako podstawnik boczny zapisany w nawiasie „(,)”. Rozstrzygające jest, jak stworzony ręcznie zapis reguły będzie interpretowany przez odpowiedni moduł Chematici. Za każdym razem poprawność zapisu testowana jest na zbiorach konkretnych struktur syntonów, czy reguła generuje poprawnie ich struktury, oraz czy reguła nie jest stosowana dla struktur, których reakcji nie opisuje. Mając na uwadze brak ścisłych zależności między zapisem konwencji SMARTS a regułami S/R czy Cahn-Ingolda-Preloga, dokonuje się dopasowania poprawności interpretacji reguły reakcji bezpośrednio przez interpreter strukturalny Chematici. Jeżeli w cząsteczce jednocześnie występują centra stereogeniczne i wiązania podwójne (lub jeszcze dodatkowo układ kilku pierścieni), to kodowanie reguł dla takich rdzeni było dla mnie niemałym wyzwaniem.

W okresie od 10.2013 do 5.2016 zakodowałem kilka tysięcy reguł reakcji. Na początku doktoratu, pierwszym zadaniem postawionym grupie kodującej było stworzenie reguł reakcji tworzenia aromatycznych pierścieni heterocyklicznych. Podzieliliśmy między siebie klasy pierścieni i przystąpiliśmy do kodowania. W tym czasie stworzyłem reguły reakcji tworzenia pierścieni indolowych, azaindolowych, furanów, benzofuranów, oksadiazoli, tiadiazoli, tetrazoli i pirydyn. Podczas prac zakodowałem nie tylko wszystkie reakcje imienne tworzenia tych pierścieni, lecz także wiele reakcji nie wyróżnionych imieniem ich odkrywcy/popularyzatora. Pierścienie furanowe potraktowałem wówczas szerzej, kodując między innymi reakcje ich utleniania, fotoutleniania, reakcji Achmatowicza, Paterno-Büchiego, czy reakcji 2-(trimetylosiloksy)-furanów z nukleofilami.

W następnym okresie rozszerzyłem zakres kodowanych reguł reakcji o niearomatyczne pierścienie heterocykliczne. Stworzyłem wówczas reguły dla epoksydów, azyrydyn, chromanów, soli pyroniowych, organicznych związków boru. Rozpocząłem wówczas kodowanie reguł reakcji enancjoselektywnych i diastereoselektywnych.

Gdy baza danych o reakcjach zawierała reakcje uznane przez nas za „podstawowe”, następnym krokiem było dodanie reakcji bardziej zaawansowanych. W tym celu zająłem się tworzeniem reguł dla reakcji opracowanych lub wykorzystywanych w syntezach totalnych przez: Eliasa J. Corey’ a, Jamesa Morkena, Michela Kirche, Varindera Aggarwala, Erica Carreirę i Benjamina Lista.

W 2015 roku, ekspercka baza danych przekroczyła próg 20 000 reguł reakcji. Zaobserwowano wówczas, że plany syntetyczne generowane przez Chematicę zaczęły zawierać nieoczywiste transformacje. Sprawdzając propozycje Chematici dla wybranych cząsteczek leków, okazało się, że program stosuje reguły reakcji w sposób nieintuicyjny. Dotyczyło to zwłaszcza wykorzystania reakcji wielokomponentowych. Rozpoznanie możliwości ich zastosowania nie jest dla programu komputerowego wyzwaniem większym niż rozpoznanie możliwości zastosowania reakcji S_N2 . Tymczasem dla człowieka rozpoznanie takich motywów strukturalnych nie jest jednakowo łatwe. Ów różnica w zdolności percepcji może tłumaczyć, dlaczego w zrealizowanych planach syntetycznych Chematici jest aż tyle reakcji trójkomponentowych.

Pierwszą metodą walidacji Chematici, było sprawdzenie, czy jest ona w stanie generować plany syntetyczne identyczne jak już wykonane. Walidacja ta, przeprowadzona w pierwszej połowie 2015 roku, wypadła pomyślnie, po czym jej wyniki opublikowano²²⁰. Chcąc potwierdzić użyteczność i wykonalność planów syntetycznych Chematici, w dalszym ciągu należało jednak przeprowadzić walidację syntetyczną. Zadanie to rozpocząłem od połowy 2015 roku.

3.3. Weryfikacja syntetyczna planów programu Chematica

Głównym celem badań w trakcie moich studiów doktoranckich była walidacja planów syntetycznych generowanych przez program Chematica. Prace syntetyczne rozpocząłem w 2015 roku i im będzie poświęcony ten rozdział.

Jednym z celów tego rozdziału jest opisanie warunków współpracy człowieka-chemika z komputerem, ze szczególnym uwzględnieniem tego, jakie pomysły i inspiracje pochodziły od komputera, a ile kreatywności i pomysłowości chemika, wymagało ich wykonanie.

3.3.1. Synteza Dronedaronu

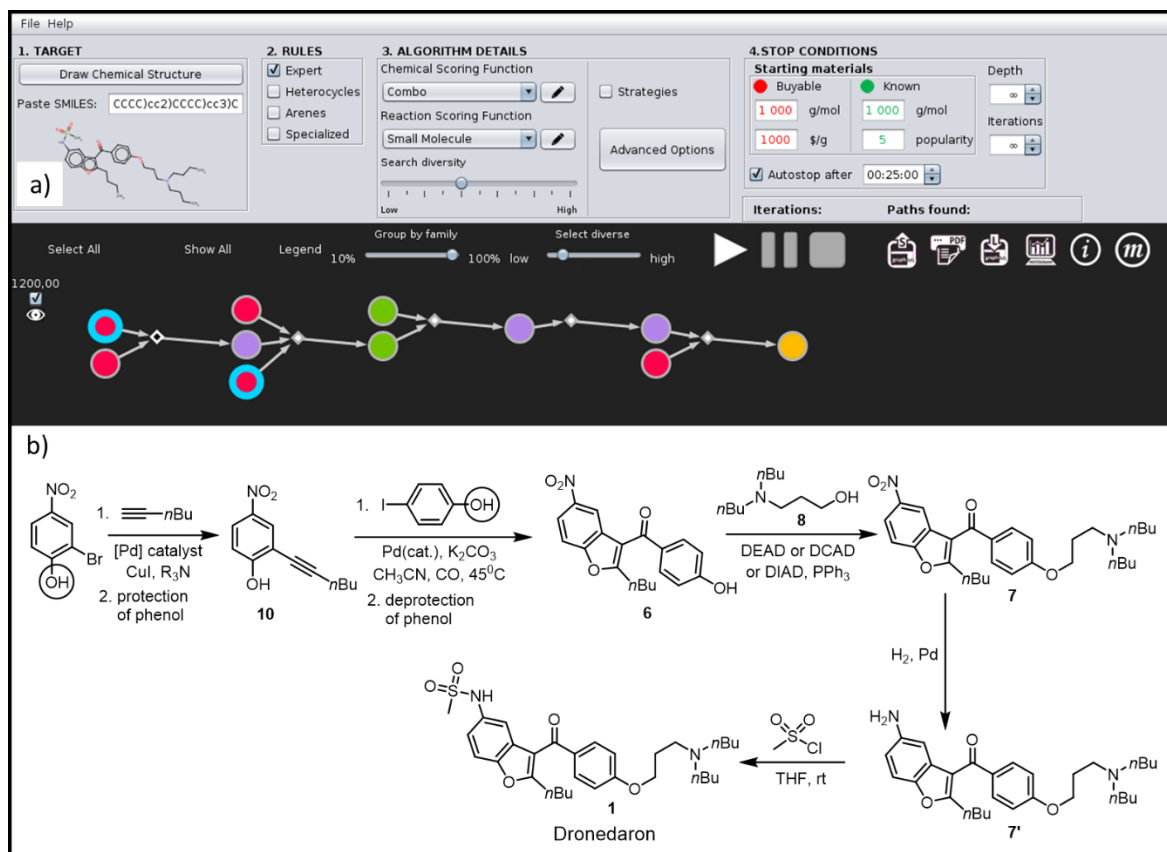
Pierwszym celem syntetycznym, którego syntezy się podjąłem był Dronedaron (**1**) - lek zapobiegający arytmii serca²⁹⁷. Cząsteczka ta wydała się interesującym celem syntetycznym ze względu na 46²⁷² patentów zawierających różne warianty jego syntezy. Jeżeli komputer znalazłby ścieżkę syntetyczną lub choćby sekwencję syntezy rdzenia Dronedaronu, która nie została już wcześniej opatentowana, byłby to interesujący precedens, natomiast jeżeli udałoby się ów teoretyczny plan wykonać syntetycznie – byłaby to pierwsza syntetyczna walidacja Chematici.

3.3.1.1. Dyskusja planu z programu Chematica dla (1)

Na Ryc.35. przedstawiono ścieżkę syntetyczną wygenerowaną automatycznie przez Chematicę. Zwraca uwagę dwukrotne pojawienie się niebieskiej obwódki dla substratów (Ryc.35a). Oznacza to konieczność zabezpieczenia grupy funkcyjnej (tu, w obu przypadkach grupy fenolowej) przed jej użyciem w reakcji. Oznacza to również, że po głównej reakcji grupa zabezpieczająca zostanie usunięta. Chematica proponowała kilka grup zabezpieczających do wyboru, jednak warunki ich wprowadzenia i usunięcia z cząsteczki pozostawały po stronie użytkownika.

Ryc.35b przedstawia struktury proponowane przez program wraz z proponowanymi warunkami zapisanymi w regułach reakcji. Pewna niejednorodność notacji warunków, na przykład oznaczenie konieczności użycia katalizatora palladowego raz oznaczona jako „[Pd] catalyst”, raz jako „Pd(cat)” – wynika z tworzenia reguł reakcji przez różne osoby. Konieczność zastosowania grupy zabezpieczającej zaznaczono kółkiem.

Jako związek wyjściowy został wybrany 2-bromo-4-nitrofenol (związek znany w literaturze, oznaczony czerwonym kółkiem). Domyślnie zaznaczoną, pierwszą reakcją powinno być zabezpieczenie grupy fenolowej – jest to uzasadnione ryzykiem cyklizacji 2-alkinofenoli do benzofuranów w obecności katalizatorów palladowych. Ryzyko to zostało odzwierciedlone w regule reakcji przez umieszczenie grupy fenolowej na liście grup niekompatybilnych. Po zabezpieczeniu fenolu, należy przeprowadzić reakcję Sonogashiry z dowolnie wybranym katalizatorem palladowym (program tego nie specyfikuje), a następnie usunąć grupę zabezpieczającą. Tak uzyskany 2-alkinofenol (związek nieznany wówczas w literaturze, co symbolizuje fioletowe kółko), należy poddać katalizowanej kompleksem palladu, karbonylatywnej annulacji. Ponownie zaznaczono, że jodek arylowy, będący partnerem w tej reakcji, nie może posiadać wolnej grupy fenolowej, więc przed reakcją należy ją zabezpieczyć. Uzyskany w wyniku annulacji keton nadal będzie posiadał grupę zabezpieczającą fenol, więc po reakcji należy ją usunąć. W kolejnym kroku, do syntezy eteru alkilowo-arylowego została wybrana reakcja Mitsunobu. Partnerami w tej reakcji są: fenol i aminoalkohol. Warto zwrócić uwagę że oba substraty do tej reakcji oznaczono zielonymi kółkami, to znaczy, że w literaturze pojawiły się te związki i znane są metody ich syntezy. Jednak jeżeli w przypadku fenolu substrat oznaczono fioletowym kółkiem, to znaczy, że nikt go jeszcze nie



Ryc.35. a) Okno programu Syntaurus (automatyczny moduł retrosyntetyczny programu Chematica) z planem dla Dronedaronu (**1**); b) ilustracja planu syntetycznego dla Dronedaronu (**1**) zawierająca warunki reakcji pochodzące z reguł reakcji – grupy fenolowe w kółkach oznaczają konieczność zastosowania wobec nich grup zabezpieczających.

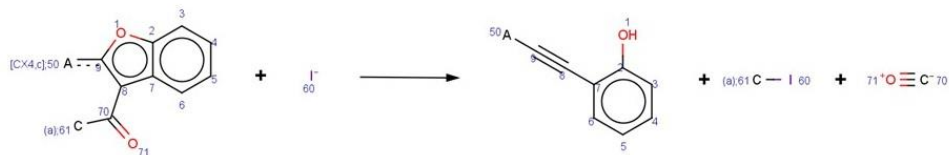
uzyskał. Po reakcji Mitsunobu, pozostaje jedynie zredukować grupę nitrową do aminowej, po czym przeprowadzić ją w metylosulfonamid, uzyskując Dronedaron (**1**).

Takie instrukcje i wskazówki pochodziły ze strony programu – to na ich bazie miała zostać wykonana synteza. Oczywiście najbardziej interesujące było wykorzystanie trójkomponentowej annulacji, budującej w jednej reakcji cały rdzeń Dronedaronu. Oznaczenie substratu do tej reakcji fioletowym kółkiem, stanowiło przesłankę, że reakcji tej nikt jeszcze nie wykonał. Zapoznałem się wówczas z patentami chroniącymi różne warianty uzyskiwania Dronedaronu – o ile opisane w patentach ostatnie etapy: redukcji grupy nitrowej i mesyloowania trudno było zastąpić atrakcyjniejszą alternatywą to brak doniesień o wykorzystaniu nieintuicyjnej annulacji, proponowanej przez Chematicę, stanowił ciekawą alternatywę dla opatentowanych syntez i w konsekwencji był główną pobudką do rozpoczęcia prac syntetycznych nad Dronedaronem (**1**).

Zanim przejdę do opisu wykonanych syntez, pragnę zwrócić uwagę na interesujący zbieg okoliczności. Otóż jak wspominałem wcześniej, pierwszymi klasami związków, których reguły reakcji kodowałem, były pierścienie heterocykliczne, wśród nich benzofurany. Ponad rok przed wygenerowaniem przez Chematicę dla Dronedaronu analizy retrosyntetycznej zawierającej trójkomponentową annulację, 26.03.2014 to ja zakodowałem regułę tej właśnie reakcji (Ryc.36.).

Ujawnia to ciekawy aspekt pracy nad bazą danych Chematici. Kodując ów regułę reakcji, nie wiedziałem o istnieniu leku Dronedaron. Kodowałem ją tak by zgadzała się z precedensem literaturowym²⁹⁸, pozwalając na różne podstawniki na pierścieniu 2-alkinofenolu i jeszcze więcej na jodku aryłowym. To że Chematica rok później rozpoznała i dopasowała jeden z tysięcy

Name: Synthesis of 3-aryl-benzo[b]furans



Rxn SMARTS:

```
[o:1]2[c:2]1[c:3][c:4][c:5][c:6][c:7]1[c:8]([C:70](=[O:71])[c:61])[c:9]2[CX4,c:50].[I-:60]
```

```
>>[O:1][c:2]1[c:3][c:4][c:5][c:6][c:7]1[C:8]#[C:9][*:50].[I:60][c:61].[C:70]#[O+:71]
```

Products: [o]2[c]1[c][c][c][c]1[c]([C](=[O])[c])[c]2[CX4,c].[I-]

SMILES: COC(=O)\C=C\c1ccc2oc(c(C(=O)c3cccs3)c2c1)-c1ccccc1.[I-]

Synthon count: 3

Explicit H: False

Reference: DOI: 10.1021/jo010839c

Condition: Pd(cat.).K2CO3.CH3CN.CO.45C

Corey 3 hetro: True

Groups to protect:

```
[c][OH].[#6][CH]=O.[#6][SX2H].[CX4,c][NH2].[CX4,c][NH][CX4,c]
```

Conditions number: 8

Incompatibile groups:

```
[#6][CH]([#6])[Cl,Br,I].[#6]C([#6])([#6])[Cl,Br,I].c[Br,Cl,I].[#6]O[OH].[#6]
```

```
OO[#6].[#6]C(=[O])OC(=[O])[#6].[#6]N=C=[O,S].[#6][N+]#[C-
```

```
].[#6]C(=O)[Cl,Br,I].[#6]C(=[SX1])[#6].[#6][CH]=[SX1].[#6][SX3](=O)[OH]
```

```
.C#[CH].[#6]=[N+]=[N-].c[N+]#[N]
```

Useful rate: 6

Ryc.36. Reguła trójkomponentowej reakcji syntezy benzofuranów, zakodowana przeze mnie 26.03.2014.

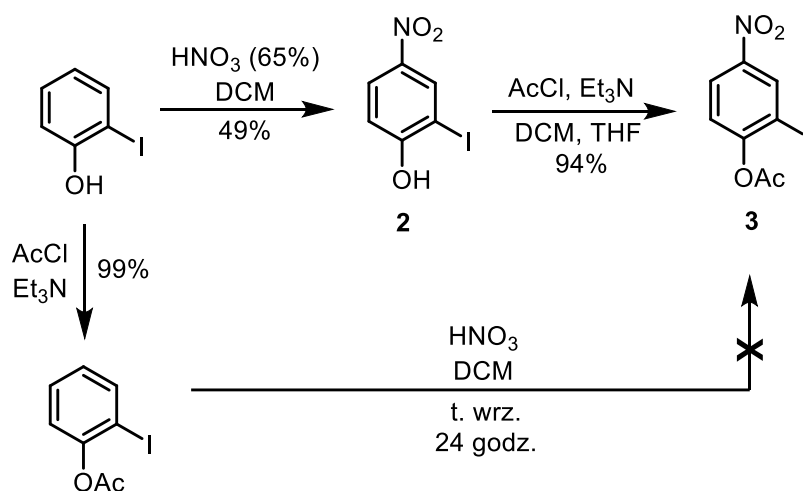
zakodowanych przeze mnie rdzeni reguł reakcji do szkieletu Dronedaronu, stanowiło koincydencję. Zdażyłem już wówczas zapomnieć o tej reakcji, pozostawała ona jednak w bazie danych, czekając na zastosowanie. Stanowi to przykład tego, że program komputerowy, w przeciwieństwie do człowieka nie zapomina i wraz ze wzrostem bazy danych może być tylko lepszy, podczas gdy człowiek, który go uczył już zapomniał co było przedmiotem zeszłorocznej „lekcji”. Przykład ten ponadto był dobrą okazją do sprawdzenia, czy pozwalając na różne podstawniki na pierścieniu nie zakodowałem reguły zbyt ogólnie.

3.3.1.2. Adaptacja reakcji z planu

Adaptację planu (Ryc.35b) rozpocząłem od wyboru substratu. Wytypowany przez Chemicę 2-bromo-4-nitrofenol, jest co prawda związkiem komercyjnie dostępnym, lecz cena niemal 1000 PLN/5 g kazała mi znaleźć tańszą alternatywę, mając w perspektywie kilkietapową syntezę. Obecność atomu bromu służyła wyłącznie możliwości przeprowadzenia reakcji Sonogashiry, dlatego zainteresowałem się również 2-jodofenolami. Odpowiedni 2-jodo-4-nitrofenol,

również był zbyt drogi (ponad 1000 PLN/1 g). Spróbowałem zatem wyjść z 15 razy tańszego 2-jodofenolu i wprowadzić doń grupę nitrową. Istniała już odpowiednia procedura w literaturze²⁹⁹, wykorzystująca stężony (65%) HNO₃ w DCM. Zaletą tej metody jest tani substrat i odczynniki, oraz łatwość skalowania reakcji, co jest szczególnie ważne przy pierwszym substracie na kilkietapowej ścieżce. Niewątpliwym ograniczeniem była osiągnięta wydajność, literatura, raportowała o uzyskaniu wydajności 43%, co wynika oczywiście z braku kontroli nad nitrowaniem w pozycję *orto* i *para*. W pierwszych podejściach uzyskiwałem wydajności 32%-41%, ostatecznie udało mi się podnieść wydajność nitrowania do 49%. Produkty nitrowania obu pozycji były możliwe do rozdzielenia na kolumnie chromatograficznej. W ten sposób uzyskałem dostęp do fenolu (**2**).

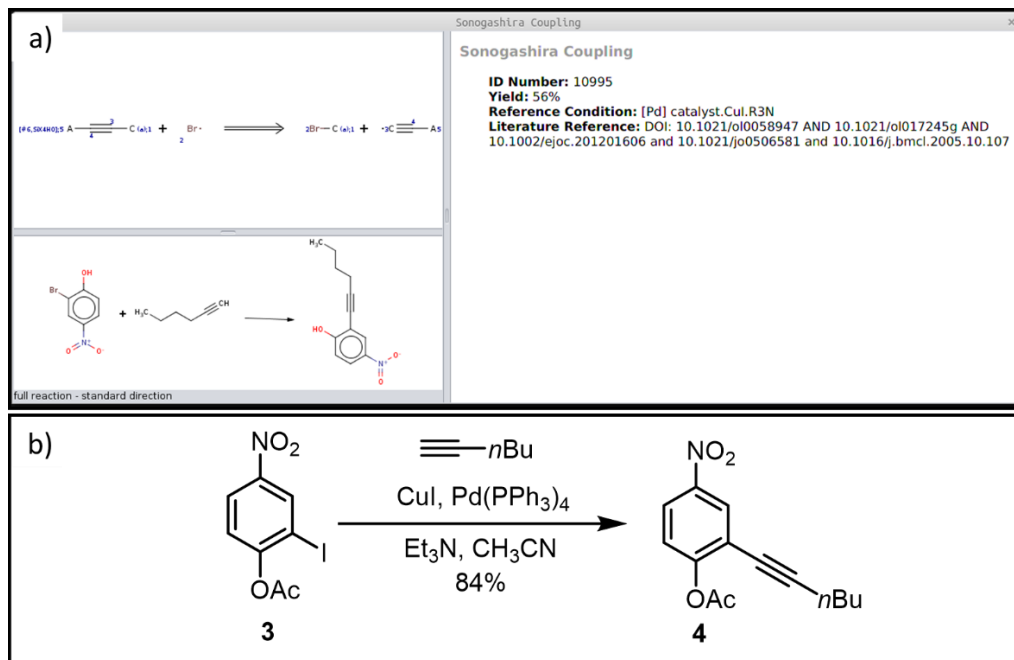
W celu uniknięcia niepożądanego i nieodwracalnego cyklizacji do benzofuranu, fenol (**2**) przed reakcją Sonogashiry należało zabezpieczyć. W tym celu wybrałem grupę acylową. Ester (**3**) uzyskałem w acylowaniu przy użyciu chlorku acetylu – początkowo uzyskiwałem wydajności rzędu 50%, później udało mi się podnieść wydajność do satysfakcjonujących 94%. W tym momencie pomyślałem, że skoro i tak muszę zabezpieczyć grupę fenolową w (**2**), to równie dobrze mogę acylować 2-jodofenol. Wówczas ze względu na steryczny wpływ grupy acylowej, pozycja 6 pierścienia będzie bardziej przesłaniana, co powinno pozwolić przeprowadzić nitrowanie w sposób bardziej selektywny. W tym celu w reakcji z chlorkiem acetylu uzyskałem octan 2-jodofenolu (Ryc.37.). Jednak ester ten, nie ulegał już reakcji nitrowania w stosowanych uprzednio warunkach. Pomimo ogrzewania we wrzeniu pod chłodnicą zwrotną, nie zaobserwowałem jakichkolwiek śladów produktu. Na tym poprzestałem i ester (**3**), uzyskiwałem w opracowany pierwotnie sposób.



Ryc.37. Alternatywna próba uzyskania **3**, wykorzystująca steryczny wpływ grupy acylowej.

Reakcja Sonogashiry, była pierwszą reakcją zaproponowaną przez program. Dysponując zabezpieczonym substratem (**3**), wykonałem ją wykorzystując jako rozpuszczalnik CH₃CN, jako odczynniki 1-heksyn (1,3eq.), Et₃N (10 eq.), jako katalizatory użyłem PdCl₂(PPh₃)₂ (2% mol) i CuI (4% mol). W pierwszej próbie uzyskałem wydajność 31%, w drugiej 46%. Postanowiłem wówczas zmienić katalizator palladowy na Pd(PPh₃)₄, używając (1,5% mol) i odpowiednio (3,1% mol) CuI. Pozwoliło to uzyskać wydajność 34%. Używając tego samego zestawu katalizatorów, reakcję przeprowadziłem w THF. Poskutkowało to wzrostem wydajności do 66%, a zwiększenie ilości 1-heksynu z 1,3 eq. do 2,0 eq., spowodowało dalszy wzrost do 82%. Reakcja kilkakrotnie przeprowadzana w tych warunkach (Ryc.38b) zachodziła z wydajnością 82-84%, co uznano za zadowalający wynik, pozwalający podjąć prace nad następnym etapem i zapewniający ilość substratu wystarczającą do dalszych prac.

Plan Chematici (Ryc.38a) nie precyzował rodzaju katalizatora palladowego, toteż w pierwszej kolejności testowałem te najczęściej stosowane w reakcji Sonogashiry³⁰⁰⁻³⁰¹. Chematica sugerowała też użycie jako zasady aminy trzeciorzędowej i taka zasada sprawdziła się w tej reakcji. Z braku sugestii dotyczącej rozpuszczalnika sprawdzono CH₃CN i THF, rodzaj rozpuszczalnika okazał się mieć niebagatelny wpływ na wydajność reakcji.

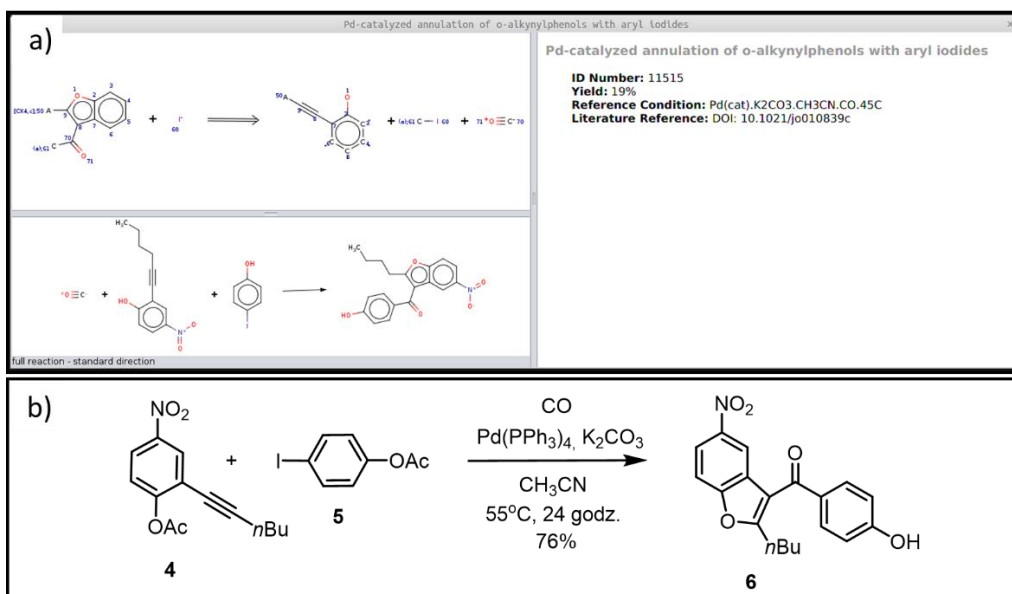


Ryc.38. a) reakcja Sonogashiry zaproponowana przez program Chematica; b) zrealizowana syntetycznie reakcja Sonogashiry dla estru (**3**).

Kluczową reakcją w zaplanowanej syntezie Dronedaronu była trójkomponentowa, karbonylatywna annulacja, budująca w jednym kroku główny szkielet cząsteczki (**1**). Z tego powodu, poświęciłem jej najwięcej uwagi, na przestrzeni kilkudziesięciu eksperymentów opracowując warunki jej adaptacji. Publikacja stanowiąca podstawę do stworzenia reguły reakcji, jedynie na niewielkiej liczbie przykładów pokazuje wachlarz substratów ulegających tej annulacji. Ważną informacją jest, że reakcji ulegają zarówno alkiiny aromatyczne jak i aromatyczno-alifatyczne. Również charakter podstawników na pierścieniu fenolu pozwala zarówno na grupy elektronodonorowe jak i elektroakceptorowe. Podczas kodowania reguły reakcji (Ryc.36.) stanowiło to dla mnie przesłankę dla dopuszczenia szerokiego spektrum podstawników.

Reguła reakcji dostarczała informacji o wstępnych warunkach, od których rozpocząłem próby jej realizacji. Na Ryc.39a przedstawiono plan proponowany przez Chematicę, warunki pochodzące z publikacji niewiele różnią się od tych ostatecznie zastosowanych (Ryc.39b), lecz zanim do tego doszło przetestowałem szereg parametrów dowiadując się więcej o tej transformacji.

W pierwszej kolejności należało uzyskać odpowiedni substrat, fenol (**10**), usuwając grupę acylową z (**4**) (Ryc.40a). Hydrolizę tego estru przeprowadzałem przy użyciu wodzianu hydrazyny w THF, który w ciągu 20 minut niemal ilościowo przeprowadzał ester (**4**) w fenol (**10**). Co ważne, przed każdą próbą reakcji trójkomponentowej, musiałem najpierw przeprowadzać deacylowanie, gdyż fenol (**4**) okazał się związkiem niestabilnym, spontanicznie przechodzącym w benzofuran (**11**).



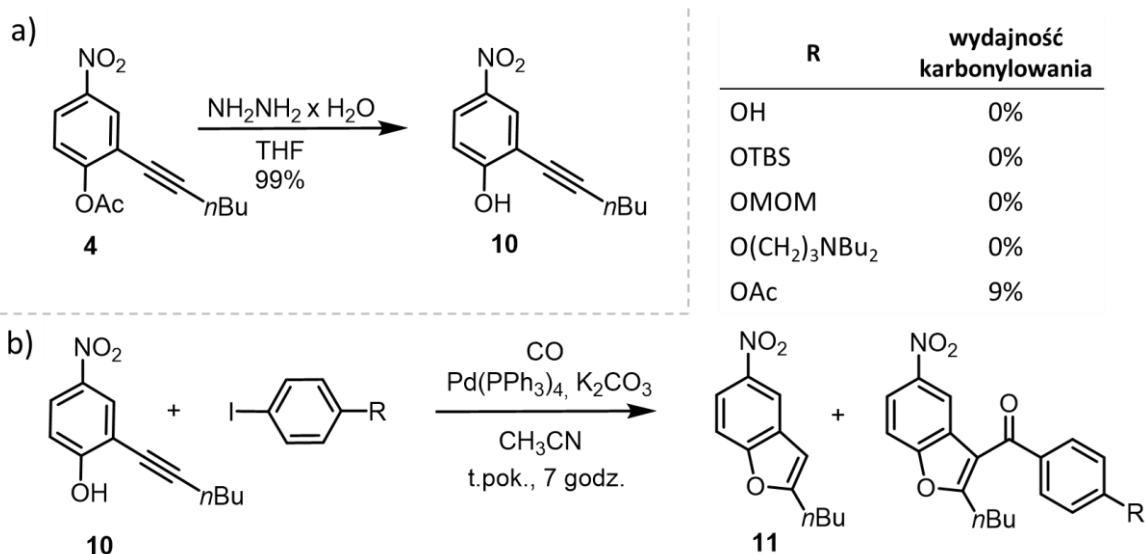
Ryc.39. a) reakcja karbonylacyjnej, trójkomponentowej annulacji zaproponowana przez program Chematica; b) zrealizowana syntetycznie karbonylacyjna trójkomponentowa annulacja dla estru (**4**).

Wyizolowany za pomocą kolumny chromatograficznej, czysty (**10**), po 24 godz. w temperaturze 4°C i w atmosferze argonu, zawierał już około 50% postaci cyklicznej (**11**).

Wstępnymi warunkami reakcji trójkomponentowej były: CH₃CN, K₂CO₃ (4eq.), Pd(PPh₃)₄ (5%mol) i atmosfera CO. Prowadząc pierwsze testy reakcji w temperaturze pokojowej, po 7 godzinach często jedynym produktem był (**11**). Reakcja jego tworzenia była niekorzystną reakcją uboczną, zachodzi ona szybciej niż tworzenie się kompleksu palladu z tlenkiem węgla(II) i jodoarene. Zwiększenie temperatury reakcji tylko przyspieszało powstawanie (**11**), dlatego postanowiłem pierwsze próby prowadzić (wbrew sugestii Chematici) w niższej temperaturze, dopóki szybkość karbonylowania nie będzie większa od szybkości wewnątrzcząsteczkowej cyklizacji do (**11**) i dopiero wówczas przeprowadzić reakcję w wyższej temperaturze. Szybkość tworzenia się kompleksu zależała między innymi od użytego jodoarenu, dlatego w pierwszej kolejności postanowiłem sprawdzić, jak grupy w pozycji para w jodoarenie wpłyną na zajście reakcji. Zgodnie z sugestią programu, postanowiłem wyjść z 4-jodofenolu. Sam 4-jodofenol nie uległ reakcji w przyjętych warunkach, podobnie fenol zabezpieczony w postaci eteru *tert*-butylo-dimetylosililowego (TBS) czy acetalu metoksymetylowego (MOM). Również wcześniejsze przyłączenie grupy -(CH₂)₃NBu₂ nie spowodowało zajścia reakcji.

Dopiero użycie grupy acylowej jako grupy ochronnej, pozwoliło po raz pierwszy na uzyskanie produktu reakcji z wydajnością 9% (Ryc.40b). Wynik ten stanowił punkt wyjściowy do dalszych prób mających na celu podniesienia wydajności.

Każdą z powyższych reakcji powtarzałem 2-4 razy, zmieniając stosowaną objętość rozpuszczalnika. Prowadzenie reakcji w większym stężeniu ma znaczenie w przypadku konkurencji między reakcją wewnątrzcząsteczkową i międzycząsteczkową. Dla reakcji wewnątrzcząsteczkowej stężenie mieszaniny reakcyjnej nie wpływa istotnie na jej przebieg, natomiast dla kilku składników większe stężenie będzie sprzyjać powstawaniu kompleksu palladu do reakcji karbonylowania. Ten sposób różnicowania szybkości reakcji okazał się skuteczny. Reakcję, w której uzyskano wydajność 9%, prowadzono w stężeniu 0,14M (w przeliczeniu na fenol, jodku arylowego używano 2 eq.). Zwiększenie stężenia do 0,38M spowodowało wzrost wydajności do 17%. Przeprowadzanie reakcji w większych stężeniach z powodów technicznych było już trudne, gdyż K₂CO₃, utrudniał efektywne mieszanie. Dlatego postanowiono zwiększyć ilość kompleksu przez zwiększenie ilości katalizatora

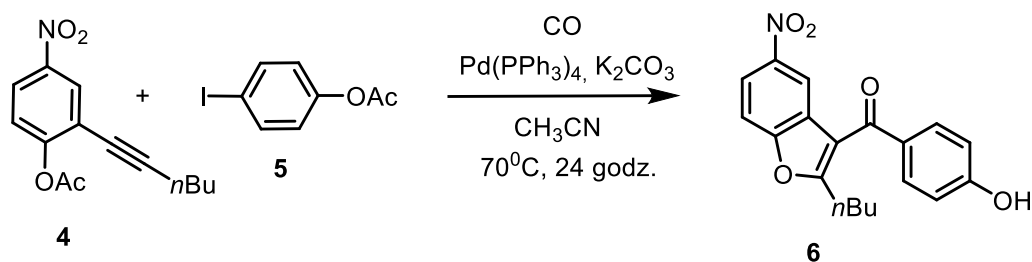


Ryc.40. a) usunięcie grupy acylowej z fenolu, przeprowadzane przed reakcją trójkomponentową; b) pierwsze próby wykonania reakcji trójkomponentowej, karbonylatywnej annulacji.

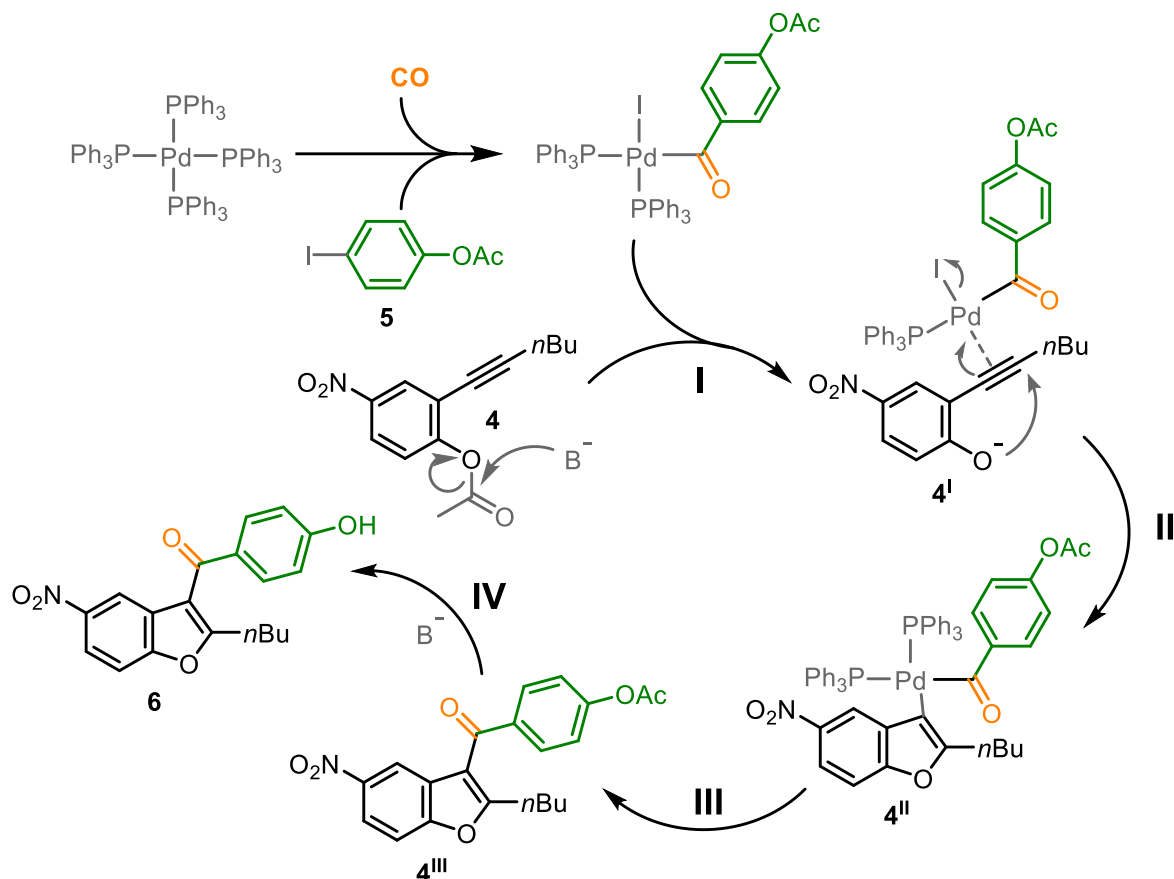
z 5 (%mol) do 10 (%mol), przy stężeniu 0,40M. Uzyskałem wówczas wydajność 30%. Postanowiłem jeszcze spróbować wygenerować kompleks palladu w mieszaninie reakcyjnej i dopiero po godzinie podać do niej fenol. Nie spowodowało to wzrostu wydajności, uzyskałem wówczas wydajność 10%. Podobnie próba dalszego spowolnienia obu reakcji: prowadząc je w 0°C nie obserwowałem wówczas konwersji fenolu, takie spowolnienie okazało się hamować obie reakcje.

Opracowawszy te warunki, postanowiłem sprawdzić wpływ na reakcję innych czynników. Zmiana rozpuszczalnika z CH₃CN na THF czy DMF nie spowodowała powstania produktu karbonylowania. Podobnie zmiana zasady na imidazol czy DBU lub dodatek AgBF₄, nie prowadziły do pożądanego produktu.

Zaobserwowałem wówczas, że w reakcji powstają śladowe ilości jeszcze jednego produktu. Nie zwróciłem na niego wcześniej uwagi, gdyż miał wyższą polarność i na płytkach TLC eluował tuż nad miejscem nałożenia. Po jego wyizolowaniu, okazało się, że jest to fenol (**6**). Oznaczało to, że w warunkach reakcji zachodzi również deacylowanie produktu reakcji karbonylowania. Zatem deprotekcja produktu karbonylowania, którą i tak miałem wykonać jako następny etap, zachodziła w mieszaninie reakcyjnej na wcześniejszym etapie. To naprowadziło mnie na pomysł, by jako substratu do reakcji użyć nie fenolu (**10**) proponowanego przez Chematicę, lecz jego postaci acylowanej (**4**), którą bezpośrednio uzyskiwałem w reakcji Sonogashiry i którą każdorazowo przed reakcją musiałem małymi porcjami odbezpieczać. Możliwość uniknięcia uciążliwego etapu odbezpieczenia z wodzianem hydrazyny okazała się dodatkową zaletą tego rozwiązania. Zwróciłem też uwagę na nierównocенność grup acylowych. Skoro deacylowaniu ulegał fenol z ketonem w pozycji 4., to octan fenolu z grupą nitrową w pozycji 4., powinien być jeszcze lepszą grupą opuszczającą i tracić grupę acylową jako pierwszy. Innymi słowy, deprotekcja, którą przeprowadzałem jako osobny, wstępny etap, mogła okazać się możliwa w jednej mieszaninie reakcyjnej, podobnie jak realizacja etapu następnego. Tak zaplanowaną reakcję (Ryc.41.) przeprowadziłem w wyższej temperaturze i prowadząc reakcję przez 24 godz., co miało ułatwić generowanie fenolu *in situ*.



Ryc.41. Włączenie deacylowania fenoli *in situ*, do reakcji karbonylacyjnej annulacji.

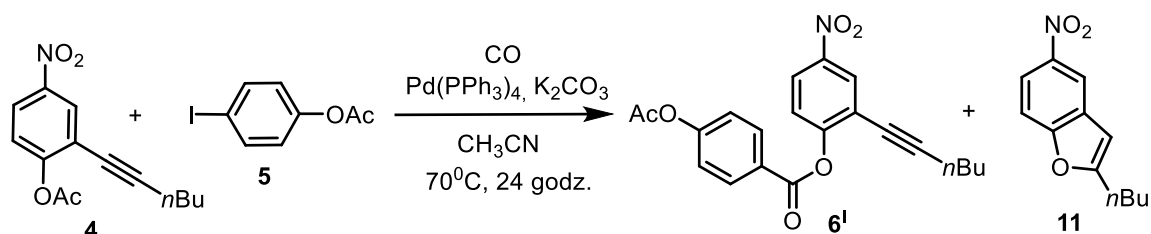


Ryc.42. Propozycja mechanizmu transformacji prowadzących do powstania fenolu (6).

Ku mojemu zaskoczeniu, już za pierwszym razem reakcja zaszła, z wydajnością 21% fenolu (6). Skróciłem tym samym całą syntezę o dwa etapy. Ani publikacja, na której się opierałem prowadząc reakcję, ani tym samym reguła z Chematici, nie uwzględniały generowania fenolu *in situ*. W badanym przypadku okazało się to możliwe, dzięki obecności grupy nitrowej i jej elektronoakceptorowemu charakterowi. W zaproponowanym mechanizmie (Ryc.42.) etap [I] deacylowania octanu 4-nitrofenolu (4) jest limitujący dla reakcji, spowalnia ją i zapobiega przed niepożądaną cyklizacją prowadzącą do benzofuranu (11). Niezależnie od deacylowania (4), następuje formowanie się kompleksu z tlenku węgla(II) i jodoarenu (5) w sferze koordynacyjnej atomu palladu. Kompleks skoordynowany z wiązaniem potrójnym alkinu (4^I), skutkować będzie podczas cyklizacji 2-alkinofenolu do benzofuranu (etap [II]), następczym acylowaniem (4^{II}) do (4^{III}) w etapie [III]. Prace nad zróżnicowaniem szybkości reakcji wewnątrzcząsteczkowej prowadzącej do (11) i międzycząsteczkowej, prowadzącej do (4^{III}) opisałem wcześniej. Zgodnie z obserwacją

poprzedzającą eksperyment (Ryc.41.), dotychczas izolowany produkt (**4^{III}**), jeżeli pozwolimy reakcji zachodzić dalej, może ulegać drugiemu deacylowaniu w etapie [IV], prowadząc do pożądanego fenolu (**6**).

Po około kilku tygodniach od przeprowadzenia karbonylowania połączonego z podwójnym, sekwencyjnym deacylowaniem, zaobserwowałem, że reakcja przeprowadzona w opracowanych wcześniej warunkach przestała zachodzić. W mieszaninie reakcyjnej, zamiast ketonu, zacząłem znajdować, niepowstający wcześniej produkt: ester (**6^I**) (Ryc.43.).

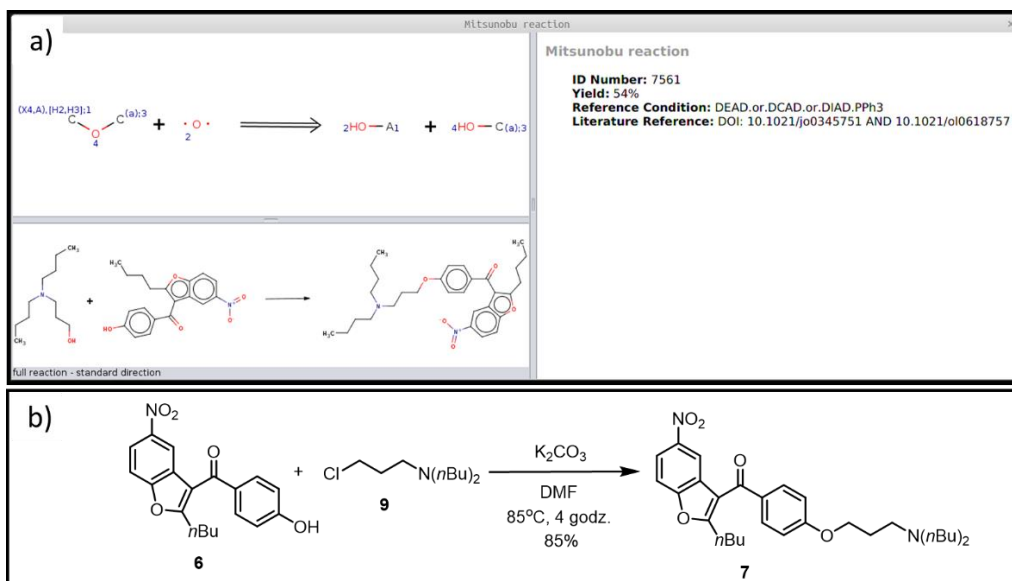


Ryc.43. Uboczna reakcja, trójkomponentowego karbonylowania fenolu, prowadząca do estru (**6^I**).

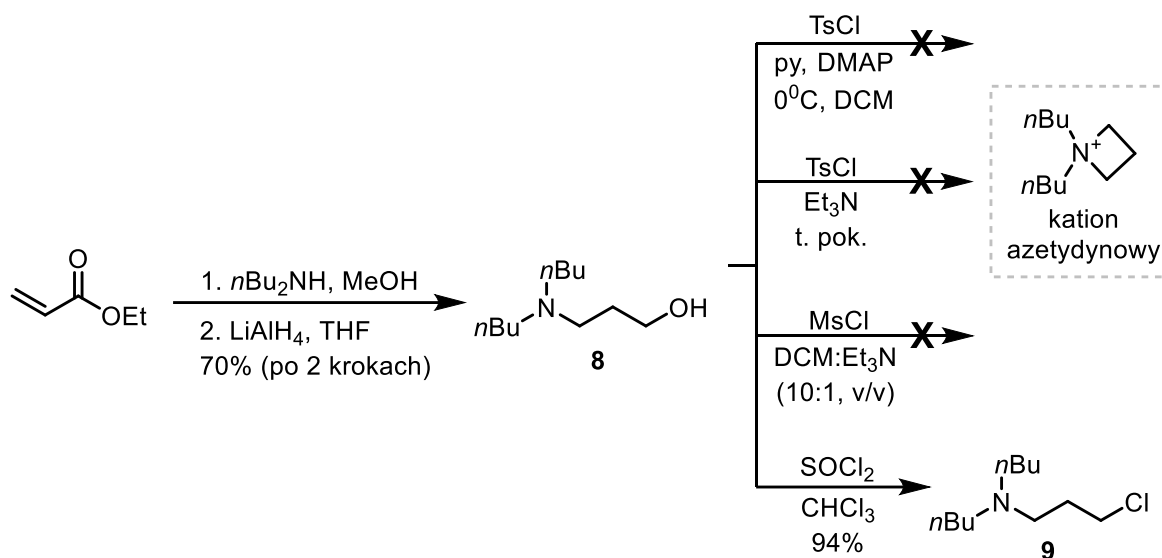
Odróżnienie pożądanego ketonu (**4^{III}**) od estru (**6^I**) nie było łatwe, oba posiadały na tyle zbliżoną polarność, że ich rozróżnienie za pomocą chromatografii cienkowarstwowej nie było możliwe. Podobnie ich widma ¹H NMR, nie wykazywały różnicy pozwalającej na ich jednoznaczną identyfikację, wszak posiadały tę samą ilość protonów, zarówno alifatycznych jak i aromatycznych. Dopiero zastosowanie ¹³C NMR pozwoliło mi jednoznacznie odróżnić przesunięcie węgla karbonylowego estru (150 ppm) od węgla karbonylowego ketonu (190 ppm). Dlatego próbując odtworzyć warunki, w których reakcja pierwotnie zachodziła, za każdym razem, ekstrahowaną, wysuszoną i odparowaną mieszaninę poreakcyjną, badałem za pomocą ¹³C NMR.

Reakcję przeprowadzałem w świeżo destylowanym CH₃CN, ponadto odgazowywałem rozpuszczalnik i nasycałem go CO – bez rezultatu. Użyłem świeżego Pd(PPh₃)₄ – bez rezultatu. Reakcję powtórzyłem w temperaturach zbliżonych do pokojowej, ale kontrolując ją za pomocą łaźni olejowej – bez rezultatu. Kiedy wydawało mi się, że sprawdziłem już wszystkie parametry, użyłem K₂CO₃ (nabytego w Sigma-Aldrich) ze świeżo otwartego pojemnika, zamiast dotychczas stosowanego K₂CO₃ z Chempur. Rezultatem było ponowne uzyskanie, wyłącznie ketonu (**6**) z nieobserwowaną wcześniej wydajnością 47% i odzyskanie nieprzereagowanego substratu. Wytlumaczeniem faktu zmiany regioselektywności reakcji z ketonu (**6**) na ester (**6^I**) i z powrotem na keton (**6**) przy zmianie stosowanego K₂CO₃, mogą być śladowe ilości wilgoci wiążące się przy każdorazowym otwarciu ze stałym K₂CO₃, jednak dalszych badań nad tym zjawiskiem nie prowadziłem.

W celu podniesienia wydajności sprawdziłem jeszcze szereg warunków. Wybrane z nich zebrałem w **Tabeli 1**, umieszczonej w części eksperymentalnej. Dalsze reakcje prowadziłem już w wariacie z podwójnym deacylowaniem (Ryc.41.). W wyniku optymalizacji warunków wybrałem: temperaturę: 55°C, czas: 24 godz., zasadę: K₂CO₃ (4,8 eq.), katalizator Pd(PPh₃)₄ (6% mol), stosowano 2,4 eq. octanu 4-jodofenolu, a CH₃CN był wcześniej odgazowany i nasycony CO. Dzięki temu udało mi się osiągnąć wydajność 73%. Dalsze zwiększenie ilości octanu 4-jodofenolu do 3,2 eq., skutkowało wzrostem wydajności do 76%. Była to też najwyższa wydajność uzyskana w tej reakcji. Testowane inne zasady, Na₂CO₃ i Cs₂CO₃ dały wydajności odpowiednio 12% i 9%, po zastosowaniu Li₂CO₃ lub K₃PO₄, reakcja nie zaszła. Inny katalizator palladowy: Pd₂(dba)₃, skutkowało uzyskaniem 16% wydajności. Kluczowa okazała się ilość stosowanej zasady,



Ryc.44. a) reakcja Mitsunobu zaproponowana przez program Chematica; b) zrealizowana syntetycznie reakcja S_N2 dla fenolu (**6**) i chlorku (**9**).



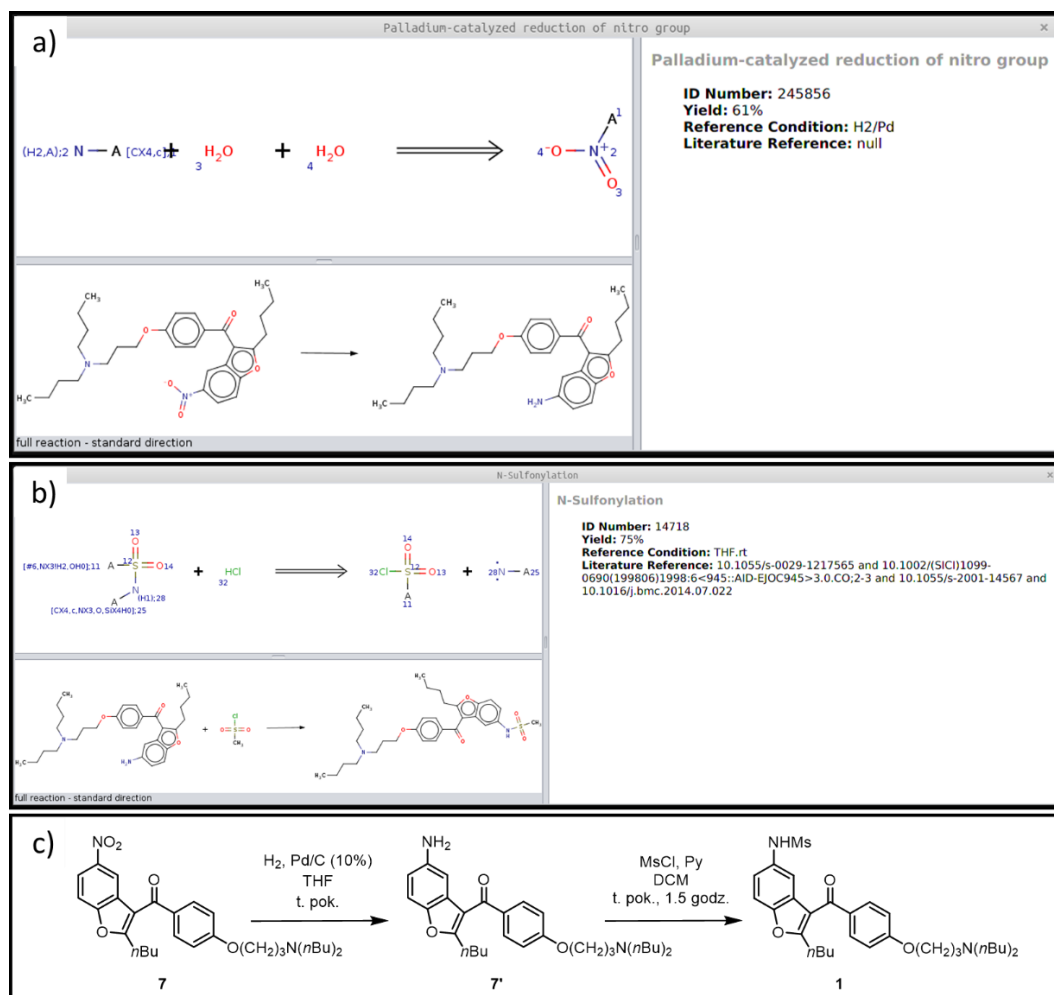
Ryc.45. Próby uzyskania elektrofilowego partnera (**9**) do syntezy (**7**).

jakiegokolwiek próby zmniejszenia ilości stosowanego K_2CO_3 , kończyły się znacznym spadkiem wydajności.

Po opracowaniu reakcji karbonylowania, pozostawało przyłączyć alkiлоaminę i zamienić grupę nitrową na sulfonamid metylowy. Zgodnie z planem Chematici (Ryc.44a), jako pierwszą należało przyłączyć aminę. Proponowano w tym celu reakcję Mitsunobu tworzenia eteru aryloalkilowego z fenolu (**6**) i alkoholu (**8**).

Alkohol (**10**) w planie syntetycznym oznaczono czerwonym kółkiem, co oznaczało, że jest to substancja komercyjnie dostępna. Postanowiłem jednak samemu przeprowadzić jego syntezę (Ryc.45.). Wychodząc z akrylanu etylu i dibutyloaminy, w reakcji Michaela uzyskałem odpowiednią

aminę. Izolacja produktu tego etapu polegała jedynie na odparowaniu metanolu, w którym prowadziłem reakcję. Następnie bez dalszego oczyszczania poddałem ester redukcji do alkoholu (**8**), używając LiAlH_4 . Po dwóch etapach uzyskując wydajność 70%. Zamiast założonej reakcji Mitsunobu, postanowiłem zastosować prostszą reakcję $\text{S}_{\text{N}}2$. W tym celu alkohol (**8**) musiałem przeprowadzić w postać elektrofilową, na przykład w ester tosyłowy. Z niemałym zdziwieniem przyjąłem niepowodzenie pierwszych prób tej reakcji. Spróbowałem również uzyskać ester



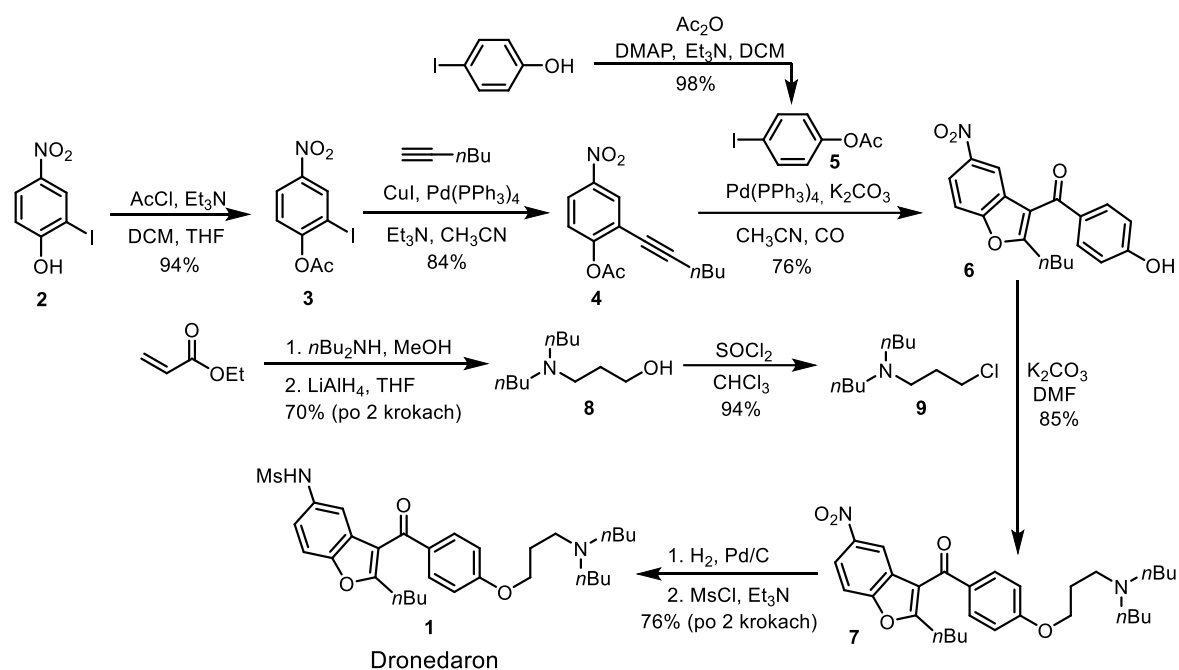
Ryc.46. a) reakcja redukcji grupy nitrowej, zaproponowana przez program Chematica; b) reakcja sulfonylowania aminy zaproponowana przez program Chematica; c) zrealizowana syntetyczna sekwencja reakcji: redukcji nitro związku (**7**) i następująca po niej reakcja sulfonylowania aminy (**7'**) prowadząca do uzyskania Dronedaronu (**1**).

mesylowy – z podobnym skutkiem. Uważniejsze sprawdzenie literatury³⁰²⁻³⁰³ wykazało, że podobne problemy, w uzyskaniu takich elektrofilowych odczynników były już odnotowywane. Jeżeli alkohol przeprowadzano w zbyt dobrą grupę opuszczającą jak bromek lub mesylan, to następowała dalsza, wewnątrzcząsteczkowa reakcja tworzenia kationu azetydynowego. Wprowadzie kation azetydynowy również wykazuje właściwości elektrofilowe, jednak nie udało mi się go wyizolować w czystej postaci, umożliwiającej następną reakcję. Zamiast tego przeprowadziłem alkohol (**8**) w chlorek (**9**), który okazał się stabilny w temperaturze pokojowej, nie wykazując tendencji do wewnątrzcząsteczkowej cyklizacji.

Dysponując zarówno fenolem (**6**), jak i chlorkiem (**9**), przeprowadziłem syntezę eteru (**7**), metodą Williamsona. Ze względu na to, że chlorek jest słabszą grupą odchodzącą, musiałem użyć DMF sprzyjający reakcji S_N2. Do tego podniesienie temperatury do 85°C pozwoliło uzyskać pożądany eter (**7**) z wydajnością 85%.

Zakończenie syntezy, zgodnie z planem Chematici, wymagało już tylko redukcji grupy nitrowej (Ryc.46a) i mesylowanie tak uzyskanej aminy (**7'**) (Ryc.46b). Pierwotnie próbowano zredukować grupę nitrową w związku (**7**) za pomocą SnCl₂ x 2H₂O w AcOEt. Jednak pomimo obserwowania pełnej konwersji (**7**), przerabianie mieszaniny reakcyjnej i izolacja produktu w tak małej skali skutkowałą wydajnościami poniżej 50%. Dlatego, by uniknąć problemów następczych przez tlenek cyny, zdecydowano się na redukcję za pomocą H₂/Pd-C, zgodnie z sugestią Chematici. Uzyskiwana amina nie była stabilna, toteż po stwierdzeniu pełnej konwersji (**7**), odsączano Pd-C, a przesącz po zatężeniu, natychmiast poddawano mesylowaniu (Ryc.46c), stosując tylko 1 eq. chlorku mesylu, pamiętając, że w przypadku amin istnieje ryzyko podwójnego ich mesylowania³⁰⁴. Powtórzywszy kilkakrotnie sekwencję redukcji i mesylowania, oraz będąc ograniczonym przez niewielką ilość substratu (**7**) dostępną do końcowych etapów, uzyskałem najwyższą łączną wydajność 76% dla dwóch ostatnich etapów, uzyskując w ten sposób Dronedaron (**1**).

3.3.1.3. Podsumowanie



Ryc.47. Schemat syntezy leku Dronedaron (**1**), wykonanej w oparciu o plan wygenerowany przez program Chematica.

Synteza Dronedaronu (**1**) była pierwszym przypadkiem syntetycznej walidacji planu syntezy opracowanego przez Chematicę. Plan (Ryc.35b) zawierał 5 etapów i dwie domyślne protekcje/deprotekcje, ostatecznie zrealizowany plan (Ryc.47.) składał się z 10 etapów. Różnica ilości etapów wynika z konieczności uzyskania substratów zaproponowanych przez program i realizacji domyślnie przyjętych etapów protekcji/deprotekcji. Skutecznie udało się wykonać trójkomponentową, karbonylatywną annulację – jej adaptacja, zawierająca dwa deacylowania, komponuje się z planem zaproponowanym przez program. Przykład ten ukazuje, że inspiracje

czerpane z Chematici w konfrontacji z praktyką syntetyczną, okazują się wykonalne i podatne na, nieplanowane wcześniej, modyfikacje opracowywane w trakcie badań.

3.3.2. Synteza 5 β /6 β -hydroksylurazydonu (12,12')

Po wykazaniu na przykładzie Dronedaronu (**1**), gotowości do walidacji syntetycznej, Chematica została poddana dalszej walidacji, przeprowadzonej we współpracy z firmą Merck KGaA. Współpraca ta była o tyle istotna, że w przypadku walidacji zewnętrznej, wybór celów syntetycznych, był całkowicie niezależny od twórców programu. Ponadto różnorodność strukturalna celów syntetycznych, umożliwiała ewentualne preferencyjne próby testowania Chematici na przykładach cząsteczek zawierających elementy strukturalne nadreprezentowane w bazie danych. Na tak liczny zbiór testowy błąd preferencyjnego wyboru został wykluczony. Jesienią 2016 roku, we współpracy z chemikami Merck KGaA, wygenerowano i wybrano do walidacji syntetycznej ścieżki dla wszystkich badanych struktur cząsteczek.

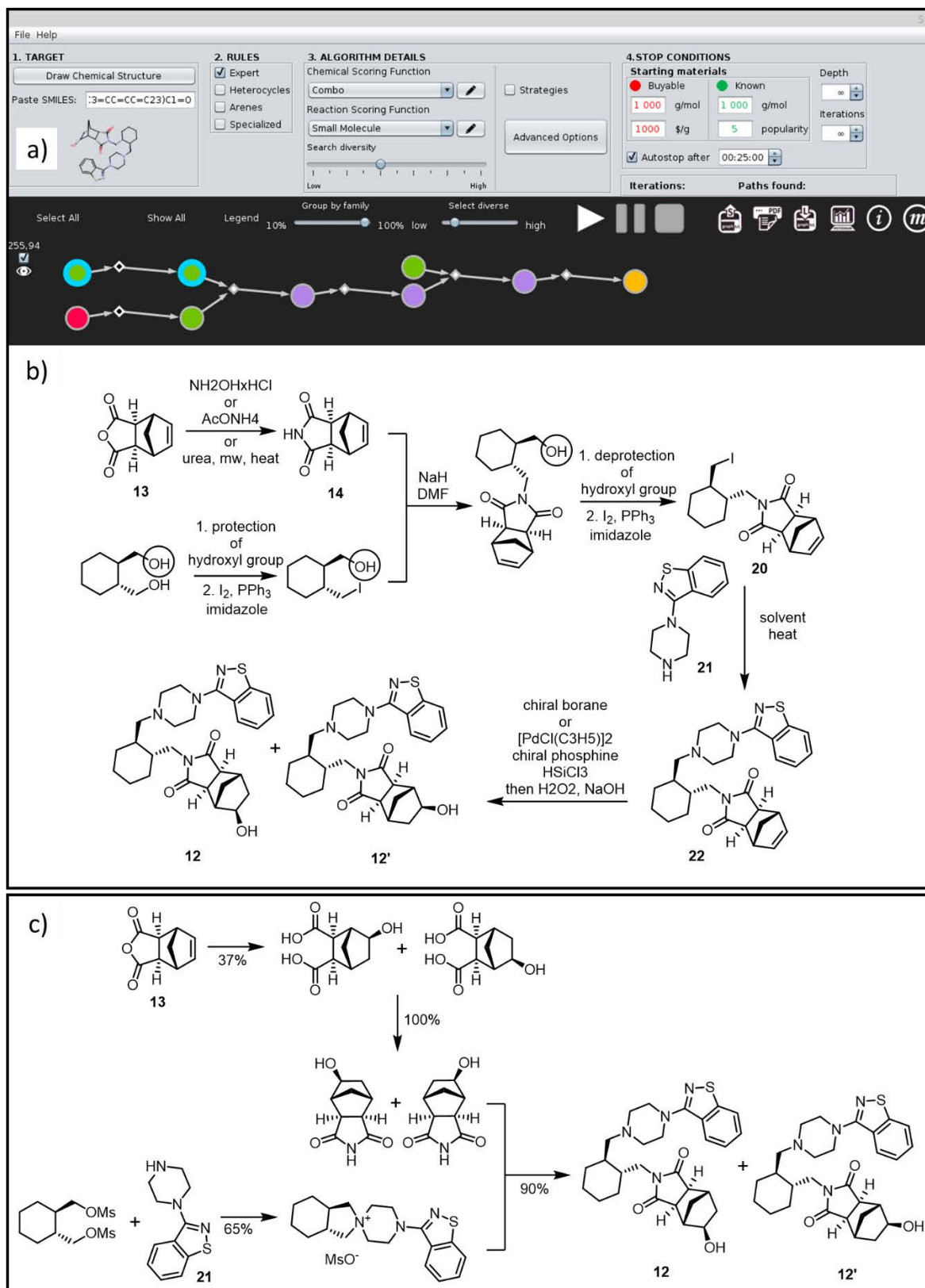
Spośród wszystkich testowanych cząsteczek, najdłuższa ścieżka syntezy została zaproponowana dla 5 β /6 β -hydroksylurazydonu (**12,12'**) (Ryc.48a). Przeprowadzenie tej syntezy było moim zadaniem, które wykonałem na przełomie 2016/2017 roku. Znana była dotychczas jedna³⁰⁵ metoda prowadząca do (**12,12'**), lecz nie udało się jej powtórzyć. Stąd zapotrzebowanie na nowy sposób uzyskania (**12,12'**).

3.3.2.1. Dyskusja planu z programu Chematica dla (12,12')

Pożądanym celem syntetycznym jest mieszanina 5 β -hydroksylurazydonu (**12**) i 6 β -hydroksylurazydonu (**12'**). Mieszanina (**12,12'**) odpowiada głównemu produktowi metabolizmu leku lurazydon³⁰⁶⁻³⁰⁷ w organizmie, znanemu w literaturze również pod nazwą ID-14283. W tym procesie pozycje 5 i 6 nie są różnicowane, a w obu przypadkach grupa hydroksylowa wprowadzana jest w pozycję β . Synteza mieszaniny metabolitów miała na celu uzyskanie wzorca do badań metabolizmu leku lurazydon. Dlatego skoro organizm nie różnicuje pozycji 5 i 6, to również i w założonej syntezie, akceptowanym produktem była mieszanina obu metabolitów.

Plan zaproponowany przez Chematicę (Ryc.48b), w odróżnieniu od znanego podejścia (Ryc.48c), wprowadzenie grupy hydroksylowej do cząsteczki pozostawia na samym końcu. Jest to ryzykowne podejście, ponieważ najbardziej ryzykowny etap planuje się przeprowadzić na samym końcu. Co więcej, reakcja ta musi zostać przeprowadzona stereospecyficznie, by uzyskać jedynie produkt o konfiguracji β . Znana dotychczas synteza wychodzi z komercyjnie dostępnego adduktu Dielsa-Aldera (**13**). Jednak w odróżnieniu od planu Chematici, od razu dokonywana jest jego hydroliza, wprowadzająca do szkieletu grupę hydroksylową. Dalsze przekształcenie w imid i wykorzystanie go jako nukleofila przyłączonego do podjednostki cykloheksylowej, jest wspólne dla obu planów.

Niska wydajność całkowita (22%) znanego w literaturze planu (Ryc.48c) stanowiła kolejną pobudkę do opracowania nowej syntezy 5 β /6 β -hydroksylurazydonu. Oprócz tego, plan syntetyczny Chematici zakładał (zasygnalizowaną niebieskim kółkiem) domyślną protekcję diolu cykloheksylowego (**15**). Protekcja ta jest konieczna, aby zamiana grupy hydroksylowej na pierwszorzędowy jodek alkilowy w reakcji Garegga-Samuelssona zaszła wyłącznie na jednej grupie hydroksylowej. Równocześnie bezwodnik (**13**) należy przeprowadzić w odpowiedni imid (**14**), proponowane są tu aż trzy różne zestawy warunków. Następnie imid (**14**) i wciąż zabezpieczony (powtórzone niebieskie kółko) jodek należy połączyć za pomocą reakcji S_N2. Sugerowane jest w tym celu wygenerowanie silnego nukleofila azotowego przy użyciu NaH. Po sfunkcjonalizowaniu jednej

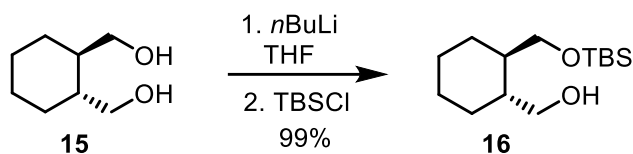


Ryc.48. a) Okno programu Syntaurus (automatyczny moduł retrosyntyczny programu Chematica) z planem dla 5β/6β-hydroksylurazydonu (**12,12'**); b) ilustracja planu syntetycznego dla 5β/6β-hydroksylurazydonu (**12,12'**), zawiera warunki reakcji pochodzące z reguł reakcji, grupy fenolowe w kółkach oznaczają konieczność zastosowania wobec nich grup zabezpieczających.

grupy hydroksylowej, należy usunąć wybraną przez użytkownika grupę zabezpieczającą i pozostałą wolną grupę hydroksylową również zamienić na pierwszorzędowy jodek (**20**) w reakcji Garegga-Samuelssona. Następnie wykorzystuje się ów jodek (**20**) do alkilowania aminy drugorzędowej do aminy trzeciorzędowej. Warunki tej reakcji S_N2 opisane zostały bardzo ogólnie. Ostatnim etapem pozostaje stereospecyficzne (lub choćby stereoselektywne) hydroksylowanie wiązania podwójnego. Do przeprowadzenia tej transformacji, Chematica proponuje dwa zestawy warunków: albo wykorzystanie chiralnego boranu w asymetrycznym hydroborowaniu³⁰⁸, albo katalizowane kompleksem palladu asymetryczne hydrosililowanie³⁰⁹⁻³¹⁰. Co ważne, reakcja ta została zakodowana na tyle szczegółowo, że dotyczy właśnie rdzenia norbornenu i specyficznych dłań warunków, dlatego szansa, że zaproponowane reakcje okażą się adekwatne w tym przypadku jest większa.

3.3.2.2. Adaptacja reakcji z planu, opis syntezy (12,12')

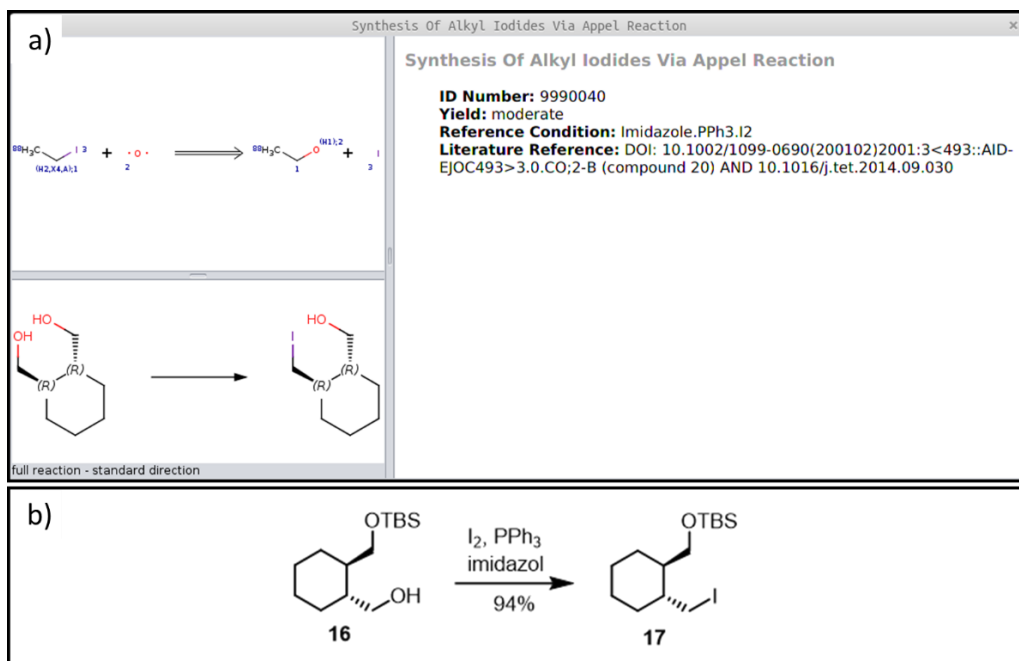
Realizację planu należało zacząć od wyboru grupy ochronnej dla diolu (**15**) (Ryc.49.). Grupa ta musiała spełniać kilka kryteriów. Po pierwsze, być stabilna w środowisku silnie zasadowym i nie ulegać reakcji z jodem. Po drugie, musi być możliwe jej wprowadzenie na tylko jedną z grup hydroksylowych. Szczęśliwie diol (**15**) wykazuje symetrię osiową, dlatego nie ma znaczenia, do której z grup hydroksylowych przyłączy się grupa ochronna. Znane w literaturze przypadki³¹¹⁻³¹² wykorzystania eteru sililowego w takiej selektywnej monoprotekcji, skłoniły mnie do wyboru grupy TBS. Procedury literaturowe zakładają konieczność prowadzenia reakcji przez ponad 12 godzin. W moich eksperymentach, wydajność 99% uzyskiwałem już po 4 godzinach. Zanim przystąpiłem do wykonania reakcji, liczyłem się z powstaniem mieszaniny: pojedynczo i podwójnie podstawionych dioli, oraz co za tym idzie, konieczności odzyskiwania nieprzereagowanego substratu. Co ciekawe, ani razu nie obserwowałem produktu podwójnego sililowania. Fakt ten nigdy nie został w literaturze dokładnie zbadany. Przypuszczam, że źródłem obserwowanej wysokiej selektywności pojedynczego podstawienia, jest różnica energii pomiędzy podwójnym a pojedynczym alkoholem litu. Pojedynczy alkohol litu może wykorzystać drugą grupę hydroksylową do koordynacji kationu litu, czego podwójny alkohol litu nie wykazuje.



Ryc.49. Schemat reakcji zabezpieczenia jednej grupy hydroksylowej diolu (**15**) w postaci eteru sililowego (**16**).

Reakcję prowadziłem w skalach z zakresu 1,0-5,3 g, uzyskując wydajności 95-99%. W przypadku prowadzenia reakcji w zbyt dużym stężeniu, mieszanina reakcyjna zestalała się. Dodanie stałego TBSCl na jej powierzchnię skutkowało roztworzeniem osadu i uzyskaniem wyłącznie pojedynczo podstawionego produktu z wydajnością powyżej 95%. Stanowi to kolejną przesłankę za wysoką preferencją tworzenia pojedynczo sililowanego alkoholu (**16**).

Uzyskawszy pojedynczo zabezpieczoną postać diolu (**15**), mogłem przystąpić do pierwszej reakcji założonej w planie Chematici. Korzystając ze standardowych warunków reakcji Garegga-Samuelssona³¹³, reakcję prowadziłem kilkakrotnie w skalach 0,4-7,6 g, uzyskując wydajności w przedziale 64-94%. Wpływ na wydajność ma chłodzenie reakcji, zwłaszcza w skali kilku gramów, gdyż reakcja po dodaniu jodu do roztworu alkoholu, imidazolu i PPh₃ jest egzotermiczna, co



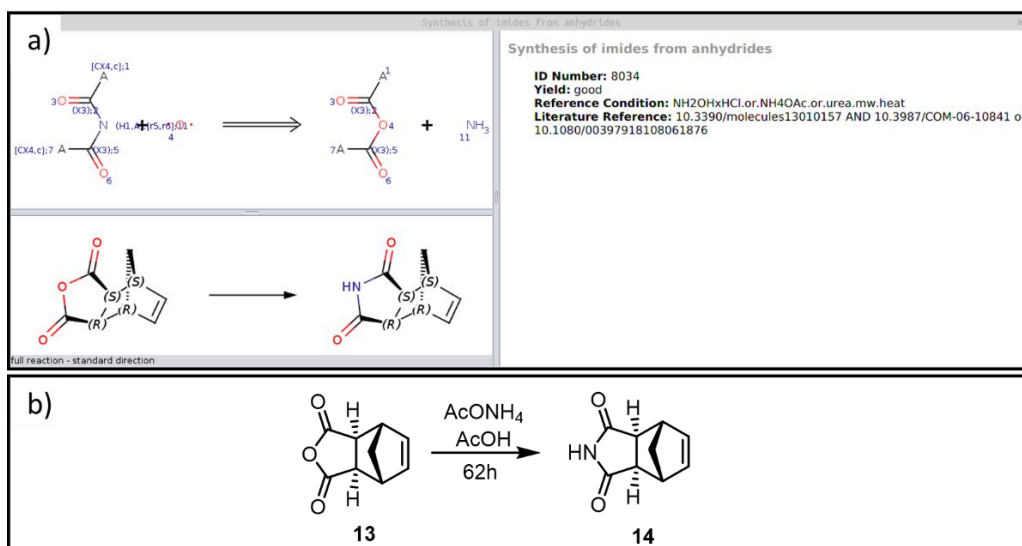
Ryc.50. a) reakcja Garegga-Samuelssona (jodowy wariant reakcji Appela) zaproponowana przez program Chemica; b) zrealizowana syntetyczna reakcja Garegga-Samuelssona dla alkoholu (**16**).

jednokrotnie doprowadziło mieszaninę reakcyjną (reakcję prowadzono w DCM) – do wrzenia. W tej sytuacji, wydajność reakcji była najniższa.

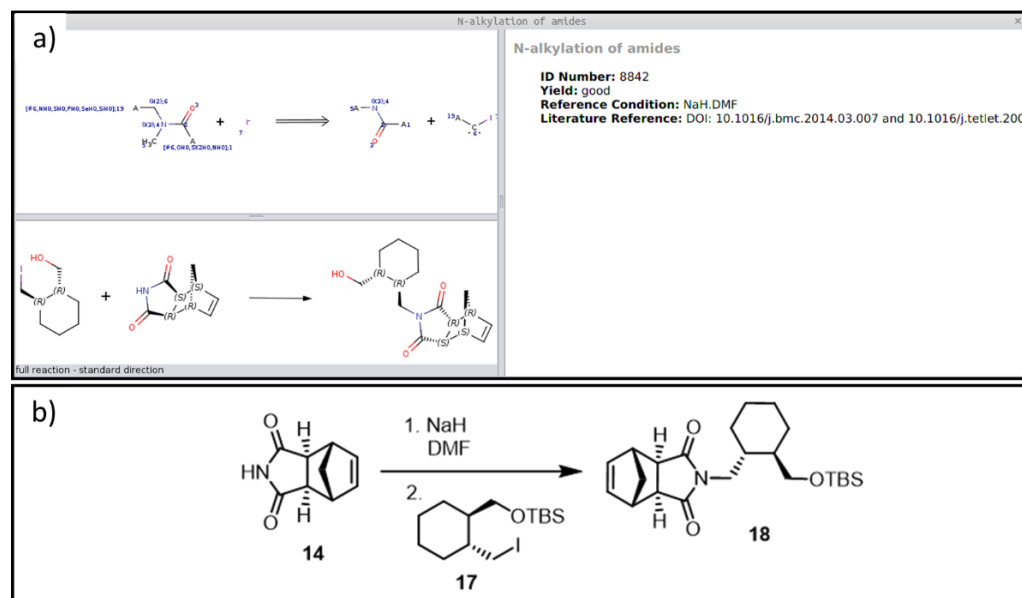
Strukturę uzyskanego związku wstępnie potwierdzono za pomocą 1H NMR i ^{13}C NMR, jednak pomiar masy za pomocą różnych metod spektrometrii masowej, nie udało się. Jodek (**17**) jonizował się z trudnością, a główna obserwowana masa pochodziła od produktu rozpadu, pozbawionego grupy *tert*-butylowej. Wyjątkowo w tym wypadku skorzystano z analizy elementarnej. Oznaczenie %C, %H i %I, pozwoliło ostatecznie potwierdzić uzyskaną strukturę (**17**).

Następnie musiałem uzyskać imid (**14**). W tym celu wykorzystałem, zaproponowany przez Chemicę bezwodnik (**13**). Stosowany dalej bezwodnik (**13**) pochodził zarówno ze źródeł komercyjnych, jak i był uzyskiwany przeze mnie. Jego synteza składała się z dwóch etapów – w pierwszym, świeżo destylowany cyklopentadien był rozpuszczany w schłodzonej do 0°C mieszaninie heksanu i octanu etylu, po dodaniu bezwodnika maleinowego uzyskiwano kryształy produktu. Reakcję prowadzono w skalach 14,4-98,3 g, uzyskując wydajności 68-77%. Produkt tej reakcji posiada jednak konfigurację *endo*, tymczasem pożądaną bezwodnik (**13**) ma konfigurację *egzo*. Przeprowadzenie bezwodnika o konfiguracji *endo* w (**13**) o konfiguracji *egzo*, wymagało przeprowadzenia reakcji retro-Dielsa Aldera, a następnie w wysokiej temperaturze ponownej reakcji Dielsa-Aldera. Wykonanie tej reakcji jest najłatwiejsze w skalach dekagramowych, tłumaczy to skalę prowadzenia reakcji Dielsa-Aldera we wcześniejszym etapie. Kryształy bezwodnika stapia się bez użycia rozpuszczalnika, ogrzewając do temperatury 195-200°C i następnie krystalizując z AcOEt. Przekroczenie 200°C skutkuje intensywnym zbrązowieniem mieszaniny, uzyskuje się wówczas mniej kryształów po krystalizacji z AcOEt. Pierwsza krystalizacja pozwala uzyskać (**13**) o *dr* (10:1) (*endo:egzo*).

Imid (**14**) uzyskiwano za pomocą standardowej procedury, wykorzystującej wielogodzinne ogrzewanie do wrzenia bezwodnika (**13**) w kwasie octowym z dodatkiem AcONH₄. Reakcję powtórzono kilkakrotnie w skali 3,0-10,0 g uzyskując wydajności od 78% do ilościowej.



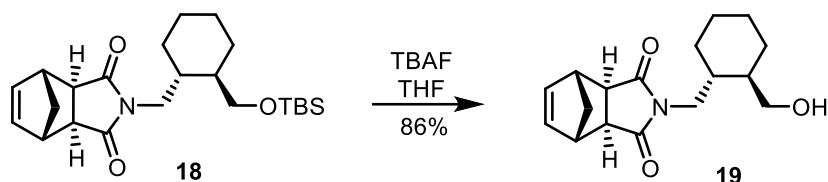
Ryc.51. a) reakcja tworzenia imidu z cyklicznego bezwodnika zaproponowana przez program Chemica; b) zrealizowana syntetyczna reakcja tworzenie imidu (**14**) z bezwodnika (**13**).



Ryc.52. a) reakcja alkilowania imidu za pomocą pierwszorzędowego jodku alkilowego zaproponowana przez program Chemica; b) zrealizowana syntetyczna reakcja alkilowania imidu (**14**) za pomocą jodku (**17**).

W kolejnym kroku, musiałem przyłączyć jodek (**17**) do imidu (**14**). W tym celu generowałem najpierw z imidu (**14**) odpowiedni anion. Rozpuszczałem (**14**) w DMF wsypując ostrożnie NaH (95%). Po pół godzinie deprotonowania, dodawałem roztwór jodku (**17**) w DMF. Reakcję w tych warunkach prowadziłem w skalach od 63 mg do 7,6g. Uzyskiwałem wydajności od 56% do 93%. Warunku NaH/DMF, zaproponowane przez Chemicę, sprawdziły się tutaj znakomicie.

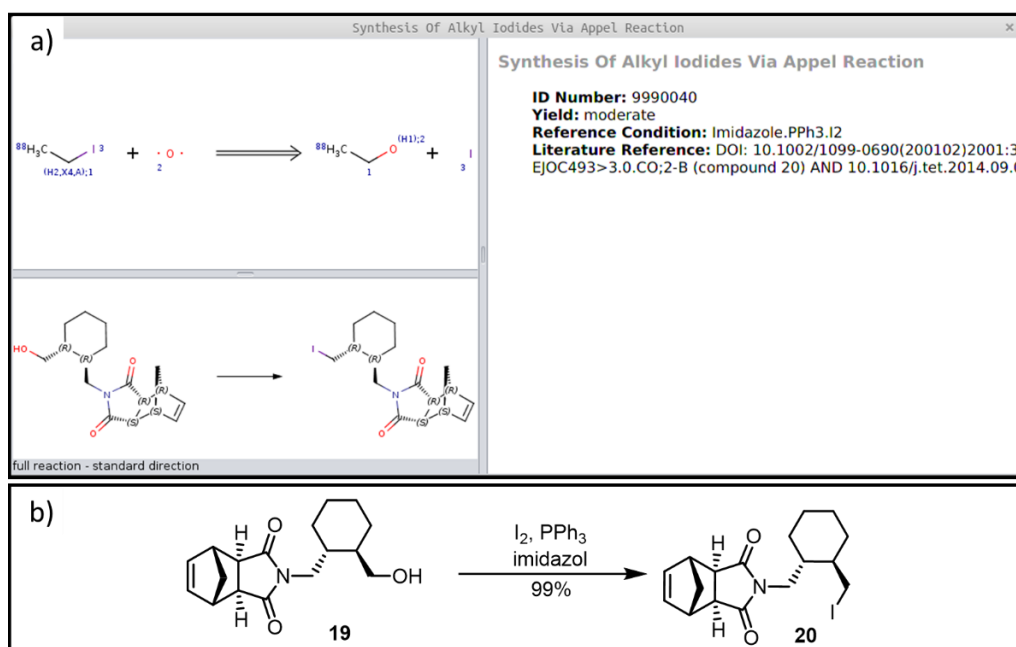
Wstępnie próbowałem uzyskać alkilowany imid (**18**) na inne sposoby. Zastosowałem najpierw najpopularniejsze warunki do tego alkilowania: K_2CO_3 /aceton³¹⁴⁻³¹⁵, nie obserwowałem jednak wówczas żadnej konwersji substratu. Ponadto próbowałem wykorzystać reakcję Mitsunobu, również popularną metodę alkilowania imidów³¹⁶. Alkohol (**16**) będący substratem do reakcji z imidem (**14**) uzyskałem już wcześniej. Niestety i w tym wypadku nie obserwowałem konwersji



Ryc.53. Reakcja usunięcia ochronnego eteru siliowego TBS z (**18**), usunięcie grup zabezpieczających jest w Chematicie domyślne, dlatego nie zostało przez program zaproponowane jako kolejny etap.

żadnego z substratów. Dopiero wykorzystanie NaH pozwoliło skutecznie przeprowadzić alkirowanie imidu (**14**).

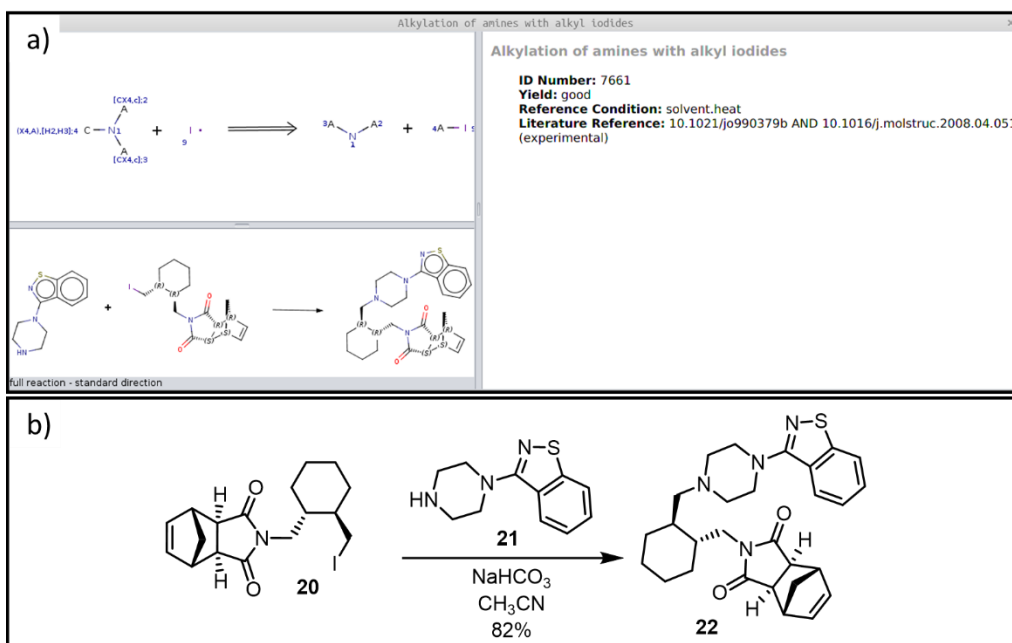
Zalkilowawszy imid (**14**) jodkiem (**17**), należało usunąć eter siliowy TBS, zabezpieczający pierwszorzędową grupę hydroksylową. W tym celu użyłem źródła anionów fluorkowych w postaci roztworu TBAF w THF. Pierwszy eksperyment wykonałem w małej skali 70,9 mg, uzyskując wydajność 66%. Kolejne powtórzenia pozwoliły zarówno na podniesienie skali (5,87g) jak i wydajności do 86%.



Ryc.54. a) reakcja Garegga-Samuelszona (jodowy wariant reakcji Appela) zaproponowana przez program Chematica; b) zrealizowana syntetycznie reakcja Garegga-Samuelszona dla alkoholu (**19**).

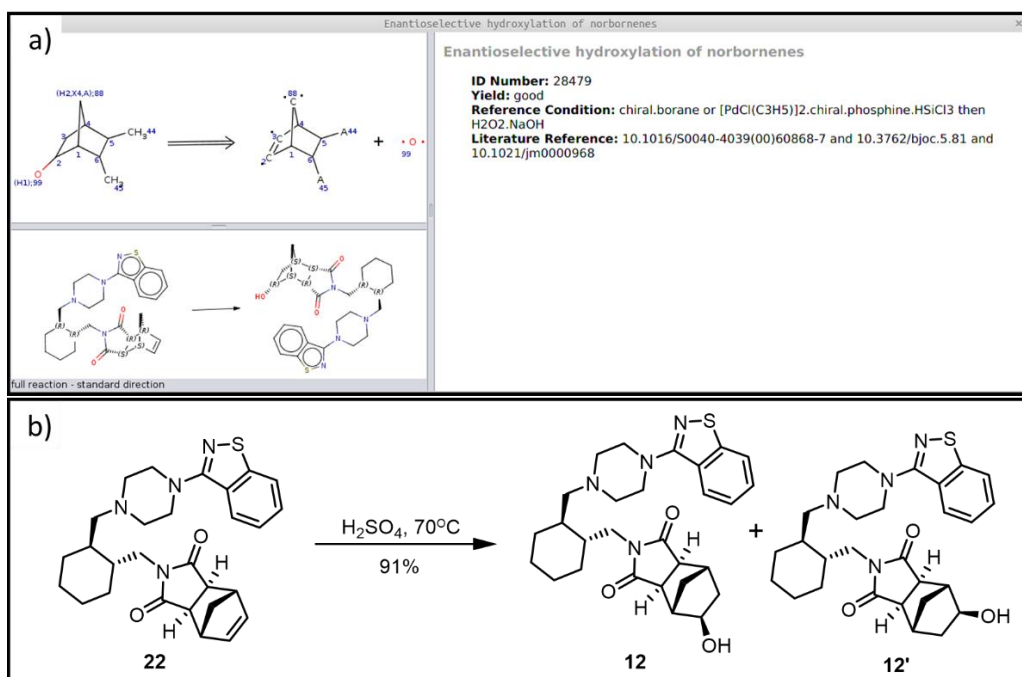
Uzyskany alkohol (**19**) wykorzystałem w następnej reakcji zaplanowanej przez Chematicę (Ryc.54a). Ponownie należało wykonać reakcję Garegga-Samuelszona, zastosowałem warunki poprzednio zastosowane dla alkoholu (**16**). I tym razem reakcja zaszła pomyślnie. Już w pierwszej próbie, w skali 348,9 mg, uzyskałem wydajność 94%. Reakcję prowadziłem w największej skali 3,49 g, uzyskując wówczas najwyższą wydajność 99%. W żadnym z moich podejść do tej reakcji, jej wydajność nie była niższa niż 92%.

Dysponując jodkiem (**20**), mogłem podjąć próbę przeprowadzenia reakcji jodku (**20**) z aminą (**21**) (Ryc.55.). Chematica dla tego etapu posiada wyłącznie bardzo ogólną sugestię, dlatego warunki musiałem opracować samemu. W pierwszej kolejności rozważałem rodzaj stosowanej zasady,



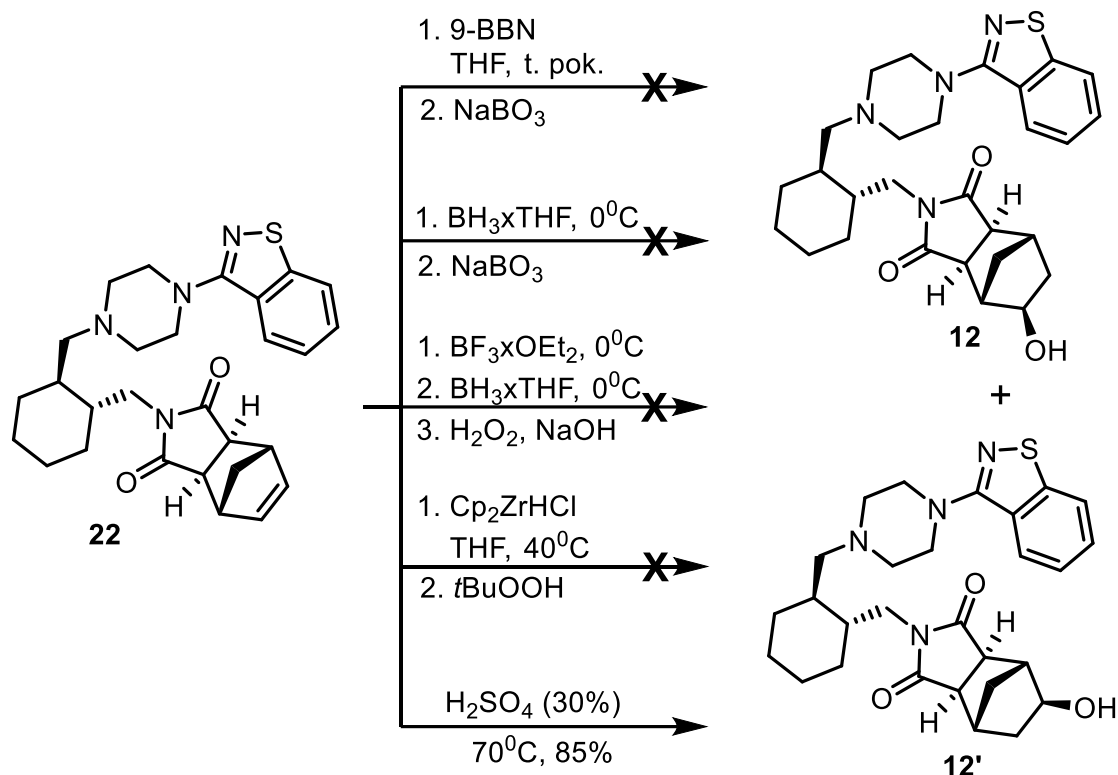
Ryc.55. a) reakcja alkilowania aminy drugorzędowej za pomocą jodku alkilowego zaproponowana przez program Chemica; b) zrealizowana syntetycznie reakcja alkilowania drugorzędowej aminy (**21**) za pomocą pierwszorzędowego jodku (**20**).

obawiając się ryzyka konkurencyjnej eliminacji pierwszorzędowego jodku. Spróbowałem przeprowadzić reakcję w dwóch różnych warunkach: K₂CO₃/DMF w temperaturze pokojowej oraz przy użyciu NaHCO₃/CH₃CN w 70°C, w obu przypadkach stosując nadmiar (**20**). Reakcja w DMF nie zaszła, natomiast w CH₃CN udało mi się wyizolować produkt z wydajnością 37%. Po przerobieniu reakcji, podczas nakładania surowej mieszaniny reakcyjnej na kolumnę chromatograficzną zaobserwowałem pojawienie się znacznej ilości białego osadu. Wyciągnąłem z tej obserwacji wniosek, że jeden ze składników mieszaniny reakcyjnej rozkłada się w trakcie reakcji. Zachowania takiego spodziewałem się raczej po jodku (**20**), ale chemicy z Merck KGaA ostrzegli mnie, że źródłem niestabilności może być pierścień 1,2-benzotiazolowy wchodzący w skład (**21**). Związek ten w postaci wolnej aminy nie jest stabilny w temperaturze pokojowej. Postanowiłem zatem przeprowadzić reakcję w temperaturze 0°C, by uniknąć ryzyka rozkładu termicznego (**21**), uzyskałem jednak wówczas jedynie śladowe ilości produktu. Wnioskowałem z tego, że temperatura jest kluczowa dla zajścia tej reakcji, dlatego postanowiłem użyć nadmiaru nie jodku (**20**) lecz aminy (**21**) i przeprowadzić reakcję w jedynych warunkach, które pozwoliły na jej zajście. Tym razem uzyskałem wydajność 73%, a po odzyskaniu nieprzereagowanego substratu (**20**) wydajność w przeliczeniu na przereagowany jodek wyniosła 97%. Utwierdziło mnie to w przekonaniu, że przyczyną spadku wydajności jest rozkład aminy (**21**). Zatem temperatura jednocześnie warunkowała reakcję, jak i powodowała rozkład jednego z substratów. Nie miałem jednak pewności, czy to jedynie wpływ temperatury jest tak destrukcyjny. Następną reakcję powtórzyłem w temperaturze pokojowej zamiast w 70°C, używając większego nadmiaru aminy (1,3 eq) oraz izolując kolbę reakcyjną przed dostępem światła za pomocą folii aluminiowej. Wszystkie te środki pozwoliły uzyskać najwyższą dotychczasową wydajność – 82% dla alkenu (**22**).



Ryc.56. a) reakcja stereoselektywnego hydroksylowania wiązania podwójnego w rdzeniu norbornenu zaproponowana przez program Chemica; b) zrealizowana syntetycznie reakcja stereosyficznego hydroksylowania alkenu (**22**), prowadząca do 5 β /6 β -hydroksylurazydonu.

Ostatnią reakcją zaplanowaną przez Chemicę było stereokontrolowane wprowadzenie grupy hydroksylowej (Ryc.56a). Program proponował w tym celu dwóch metod: 1) przeprowadzenia hydroborowania wiązania podwójnego za pomocą optycznie czynnego, dialkilboranu, z następczym utlenieniem do estru kwasu borowego i hydrolizą, lub 2) katalizowane kompleksem palladu hydrosililowanie z następczą hydrolizą. W pierwszej kolejności postanowiłem spróbować hydroborowania. Analiza zatłoczenia sterycznego obu stron norbornenu wskazuje na ich nierównocenneść, która może stanowić podstawę utrudnionego podejścia boranu od strony *endo* i wymuszać hydroborowanie do pozycji *egzo* – a w takiej właśnie występuje grupa hydroksylowa w (**12**,**12'**). Selektywność podejścia może być dodatkowo zwiększona przez zwiększenie zatłoczenie sterycznego samego dialkilboranu. 9-BBN wydawał się znakomitym kandydatem spełniającym to kryterium³¹⁷⁻³¹⁸. Przeprowadzono(Ryc.57.) kilka eksperymentów stosując różne stosunki ilościowe alkenu (**22**) i 9-BBN, jak również różne temperatury prowadzenia reakcji. Niestety w żadnym przypadku nie obserwowano nawet śladów konwersji alkenu (**22**). Detekcję za pomocą ESI-MS prowadziłem zarówno po pierwszym etapie hydroborowania, jak i po drugim etapie utlenienia. Próbowałem wówczas dwóch sposobów na utlenienie alkilboranu: użycia klasycznych warunków H₂O₂ i następnie NaOH, lub użycia jako utleniacza NaBO₃³¹⁹⁻³²⁰. Niepowodzenia pierwszych prób przywiodły mnie to do przypuszczenia, że 9-BBN mógł okazać się zbyt zatłoczony sterycznie by zbliżyć się do wiązania podwójnego alkenu (**22**). Spróbowałem użyć kompleksu BH₃xTHF³²¹⁻³²³ (Ryc.57.), który nie powinien wykazywać zawady sterycznej takiej jak 9-BBN, co miało ułatwić zajście hydroborowania. W pierwszym eksperymencie ponownie nie zaobserwowałem by alken (**22**) ulegał reakcji. Teoretycznie jedna cząsteczka boranu powinna móc przyłączyć się do trzech cząsteczek alkenu, ale być może otoczenie atomu boru trzema rdzeniami norbornenowymi zachodziłoby zbyt trudno, dlatego w następnych eksperymentach zwiększyłem ilość BH₃xTHF do (0,66 eq) i (1,0 eq) – ponownie bez rezultatu.



Ryc.57. Przykłady podejmowanych prób hydroksylowania alkenu (**22**).

Wówczas zwróciłem uwagę na fakt, że występujące w strukturze (**22**) dwie aminy trzeciorzędowe mogą konkurencyjnie wiązać borany będące kwasami Lewisa. Tłumaczyłoby to brak konwersji w poprzednich eksperymentach. Zanim doszłoby do hydroborowania, szybciej zachodziła reakcja kwas-zasada, boranu z aminą. Chcąc wyeliminować wpływ amin, przed dodaniem kompleksu BH₃ postanowiłem dodać BF₃xOEt₂ (2,0 eq). Jako silniejszy kwas miał on związać się z aminą, następnie dodałem kompleks BH₃. I tym razem (**22**) nie uległ reakcji. Zwiększyłem więc ilość BH₃xTHF, skutek był podobny. Postanowiłem spróbować innych kompleksów boranu, w różnych ekwiwalentach. Wypróbowałem BH₃xSMe₂, BH₃x*t*BuNH₂, BH₃x4-*N*-metylomorfolina, BH₃xDIPEA, BH₃xNMe₃ – bez rezultatu. Wobec serii niepowodzeń z boranem, postanowiłem spróbować zastąpić hydroborowanie hydrocyrkonowaniem³²⁴. Po przyłączeniu odczynnika Schwarza, wiązanie C-Zr można utlenić za pomocą *t*BuOOH. Wszystkie próby wykonania tej reakcji zakończyły się podobnie – brakiem reakcji (**22**).

Mając w pamięci wrażliwość pierścienia 1,2-benzizotiazolowego, która ujawniła się w poprzednim etapie, odczuwałem opór przed stosowaniem bardziej agresywnych warunków reakcji. Jednak skoro alken (**22**) nie wchodził w reakcję w stosunkowo łagodnych warunkach, postanowiłem spróbować mniej wyrafinowanej metody. Przygotowałem 30% wodny roztwór H₂SO₄ i rozpuściłem w nim alken (**22**), który dzięki posiadany dwóm grupom aminowym rozpuścił się w roztworze. Już w pierwszym eksperymencie postanowiłem zastosować grzanie, reakcję prowadziłem przez noc w 70^oC. Nazajutrz ku mojemu zdziwieniu, próbka mieszaniny reakcyjnej po zalkalizowaniu i mikroekstrakcji do Et₂O wykazała całkowitą konwersję alkenu (**22**) i pojawienie się produktu i niższym R_f, którego spodziewałem się po alkoholu. Po przerobieniu mieszaniny reakcyjnej i oczyszczeniu produktu za pomocą kolumny chromatograficznej, masa produktu zaobserwowana na ESI-MS okazała się zgodna z masą pożądanego produktu. ¹H NMR potwierdził

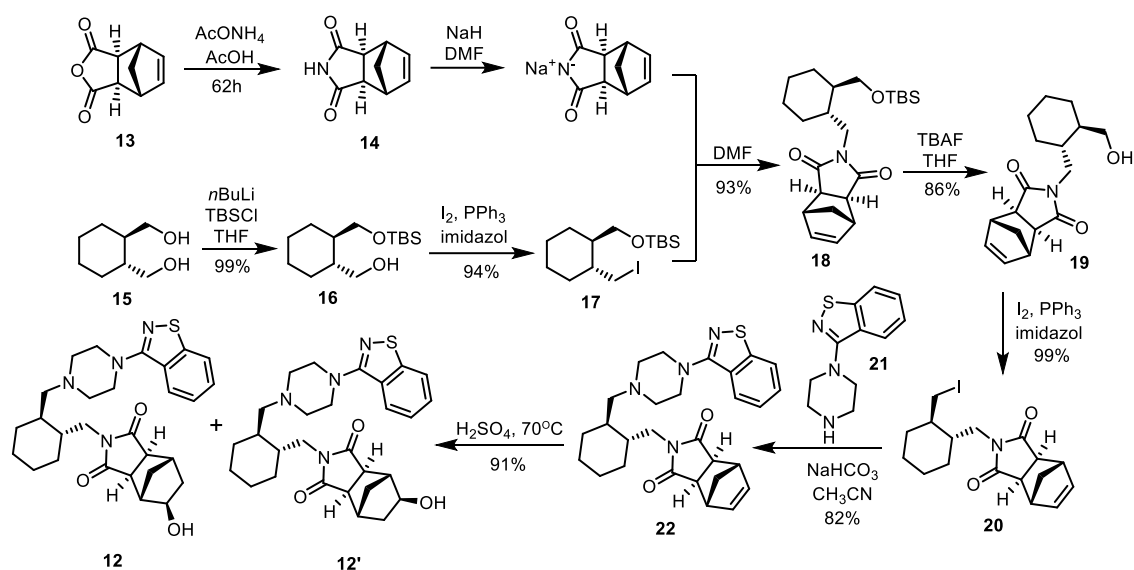
zanik wiązania podwójnego i uzyskanie pożądanego produktu alkilowania. Po zawróceniu próbki i wysuszeniu, uzyskałem produkt w postaci kruchej piany, a wydajność wyniosła 85%.

Należało jeszcze ustalić czy grupa hydroksylowa została wprowadzona w pozycję *endo* czy *egzo*. Rozstrzygnięcia dostarczyło użycie spektroskopii 2D NMR ROESY. W 6d-DMSO widoczny był sygnał pochodzący od grupy hydroksylowej alkoholu (**12,12'**) prześledzenie sprzężeń protonu grupy hydroksylowej oraz protonu znajdującego się na węglu związanym z grupą hydroksylową i ich korelacja z sąsiednimi protonami znajdującymi się w bicyklicznym rdzeniu wykazały położenie grupy hydroksylowej w pozycji *egzo*. Produktu z grupą hydroksylową w położeniu *endo* nie obserwowałem w tej reakcji, obserwowana selektywność znajduje potwierdzenie w literaturze³²⁵.

Kilkukrotne powtórzenie reakcji i wypróbowanie kilku stężeń alkeny (**22**) pozwoliło zwiększyć wydajność do 91%. Co ciekawe, prowadzenie reakcji w 30% H₂SO₄ w temperaturze pokojowej skutkuje brakiem konwersji alkeny (**22**).

3.3.2.3. Podsumowanie

Wykonanie syntezy 5β/6β-hydroksylurazydonu (**12,12'**) zgodnie z planem wygenerowanym przez program Chematica stanowi kolejny przykład walidacji poprawności jego działania, przez wykonanie syntezy. O ile w przypadku Dronedaronu porównanie z istniejącymi syntezami pokazywało, że wykonana synteza jest porównywalna z istniejącymi, to w przypadku 5β/6β-hydroksylurazydonu (**12,12'**) ocena nie pozostawia wątpliwości i jednoznacznie wskazuje wyższość komputerowo wygenerowanego planu. Pomimo długości 9 etapów całkowita wydajność syntezy wynosi 55%. Jest to 2,5 raza większa wydajność niż dotychczas raportowana³⁰⁵.

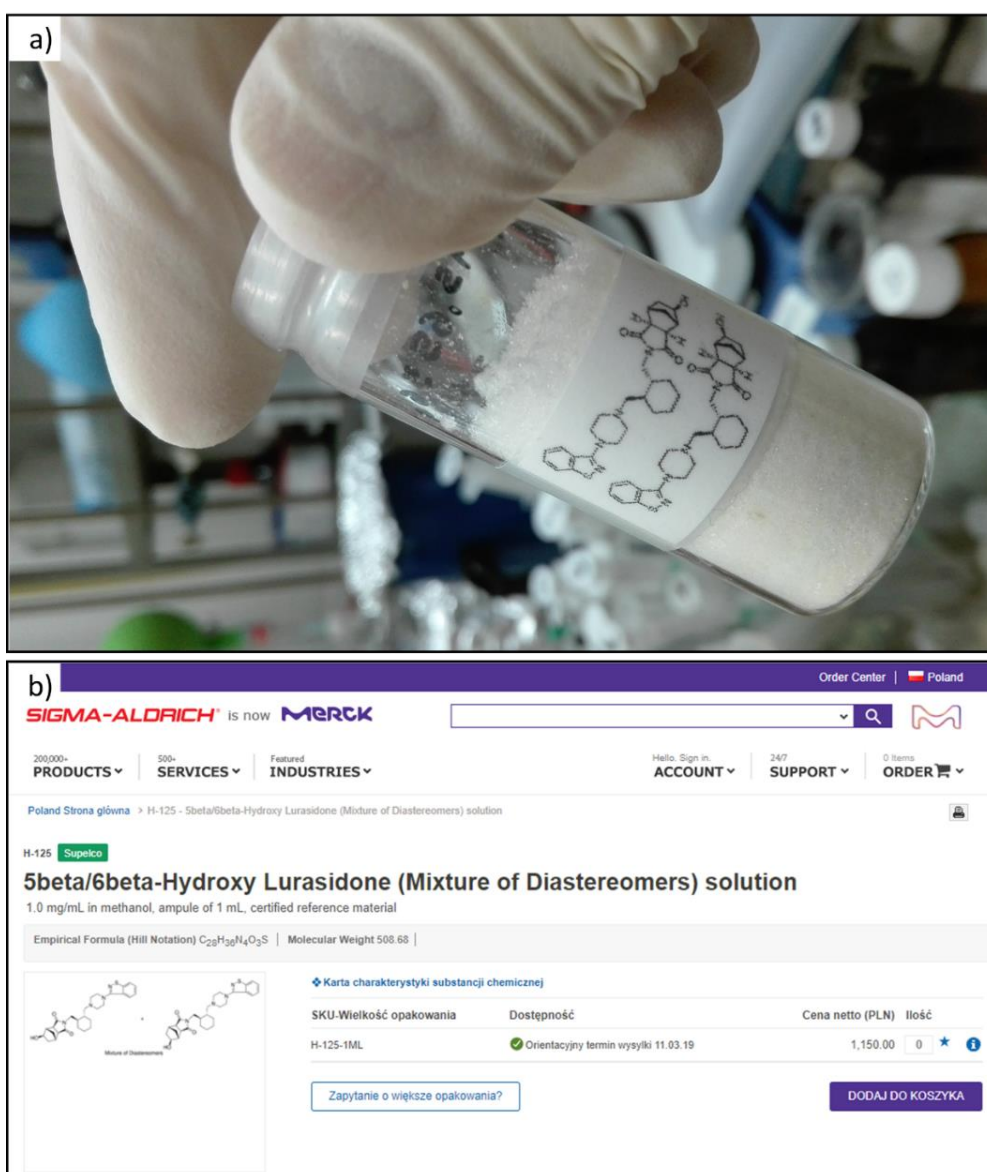


Ryc.58. Schemat syntezy 5β/6β-hydroksylurazydonu (**12,12'**) wykonanej w oparciu o plan uzyskany z programu Chematica.

Na uwagę w zaproponowanym przez Chematicę planie zasługuje umieszczenie asymetrycznego hydroksylowania na ostatnim kroku. Z perspektywy programu można to prosto wyjaśnić: obecność grupy funkcyjnej, niekompatybilnej z większością zaplanowanych etapów obniżałaby ich ocenę. Dlatego ścieżka unikająca tego problemu zyskała wyższą ocenę i znalazła się wyżej na liście zaproponowanych użytkownikowi planów syntetycznych.

Widząc pierwszy raz ten plan syntetyczny, zaproponowałem, że wolałbym najpierw wymyślić jak stereospecyficznie wprowadzić na wczesnym etapie grupę hydroksylową, a następnie utrzymując ją w postaci zabezpieczonej, przyłączyć pozostałe fragmenty cząsteczki 5β/6β-hydroksylurazydonu (**12,12'**). Nie znałem wówczas opatentowanej³⁰⁵ metody syntezy (Ryc.48c). Okazuje się, że jej twórcy kierowali się podejściem podobnym do mojej pierwszej intuicji by jak najbezpieczniej tę syntezę przeprowadzić. Tymczasem Chematica proponowała nieintuicyjne przeprowadzenie ryzykownej reakcji na samym końcu. Zdecydowałem się zaufać planowi programu i zaryzykować niepowodzenie na ostatnim etapie. Jak się okazało, ostatnią reakcję można przeprowadzić w prostych warunkach, ale już ich znalezienie nie było trywialne.

Jeżeli weźmiemy pod uwagę elastyczność i podatność planu syntetycznego na adaptacje, to właśnie plan lokujący hydroksylowanie na końcu jest na nie bardziej podatny. W razie niepowodzenia ostatniego etapu, można podjąć próbę hydroksylowania na wcześniejszym etapie. Nie ulega wówczas zmianie całość ścieżki, a jedynie kolejność ostatnich transformacji. Syntezę



Ryc.59. a) Zdjęcie fiolki z 3,10 g uzyskanego przeze mnie 5β/6β-hydroksylurazydonu (**12,12'**); b) Zrzut ekranu ze strony firmy Merck KGaA³²⁶ oferującej uzyskany przeze mnie 5β/6β-hydroksylurazydon (**12,12'**).

5 β /6 β -hydroksylurazydonu (**12,12'**) udało się zrealizować w takiej kolejności jaką zaproponowała Chematica, bez konieczności zmiany kolejności ostatnich etapów.

W wyniku prac uzyskałem 3,10 g 5 β /6 β -hydroksylurazydonu (**12,12'**). Dla moich badań, uzyskanie tego związku stanowiło potwierdzenie założonej hipotezy badawczej. Dlatego po zakończeniu prac związek został przekazany firmie Merck KGaA, który włączył go do swojej oferty.

Wykonanie syntezy 5 β /6 β -hydroksylurazydonu (**12,12'**) stanowiło jeden z kilku przykładów walidacji poprawności planowania syntezy przez Chematicę. Wszystkie syntezy związków chemicznych objętych testem prowadzonym we współpracy z firmą Merck KGaA były wykonywane w tym samym czasie, dlatego na potrzeby dowodu przez walidację eksperymentalną należy traktować je łącznie. Pomyślna realizacja tego planu jest historycznym osiągnięciem. Nigdy wcześniej program komputerowy nie wygenerował planu syntetycznego, który wykonany w laboratorium okazałby się lepszy od istniejących. Synteza 5 β /6 β -hydroksylurazydonu (**12,12'**) była najdłuższą z przeprowadzonych w ramach testu.

3.3.3. Synteza hydrazonu (**23(R,R)**)

Po zakończeniu syntetycznej walidacji planów Chematici dla małych cząsteczek, naturalne było kontynuowanie walidacji na przykładach bardziej skomplikowanych struktur. Dotychczas zrealizowane plany miały nie więcej niż dziewięć etapów. Udowodniono więc poprawność planowania programu na takim „dystansie”. Bardziej skomplikowane związki wymagają jednak zazwyczaj dłuższych syntez. Czy i wówczas plany generowane przez Chematicę nadal będą poprawne? Należało to udowodnić eksperymentalnie. Oczywiście jest to dużo większe wyzwanie, ponieważ każdy zaplanowany etap musi zostać zrealizowany. Niepowodzenie choćby na jednym etapie, niepoprawna predykcja, nieuwzględnienie czynnika warunkującego reakcję – skutkuje zarówno niewykonaniem planu jak i negatywną walidacją. Automatyczne wygenerowanie kilkunastoetapowej sekwencji o pełnej poprawności wykazanej syntetycznie dowodziłoby zaawansowania programu Chematica oraz udowadniało, że planowanie wieloetapowych syntez przestało być jedynie ludzką zdolnością. W celu wykonania takiej walidacji postanowiono wykonać syntezę totalną produktu naturalnego. Zwrócono się w stronę produktów naturalnych wyizolowanych w ostatnich latach, przez co nigdy nieuzyskanych syntetycznie. Na cel syntetyczny walidacji wybrano takamonidynę (**24**) (Ryc.65.). Jest to cytotoksyczny alkaloid wyizolowany w 2015 roku³²⁷. Należy on do rodziny alkaloidów takamanowych, a najbardziej zbliżony budową jest do takamoniny, pozbawionej jednak trzeciorzędowej grupy hydroksylowej. Ta niewielka różnica sprawia, że żadna ze strategii syntez totalnych dla takamoniny³²⁸⁻³³⁹ nie może zostać zastosowana dla takamonidyny (**24**). Postanowiono stworzyć tą nową strategię za pomocą programu komputerowego. Zgodnie z planem Chematica wygenerowała nowatorską strategię dla rdzenia alkaloidów takamanowych pozwalającą na stereokontrolowane wprowadzenie trzeciorzędowej grupy hydroksylowej. Po drobnej modyfikacji strategia ta może być również zastosowana do takamoniny.

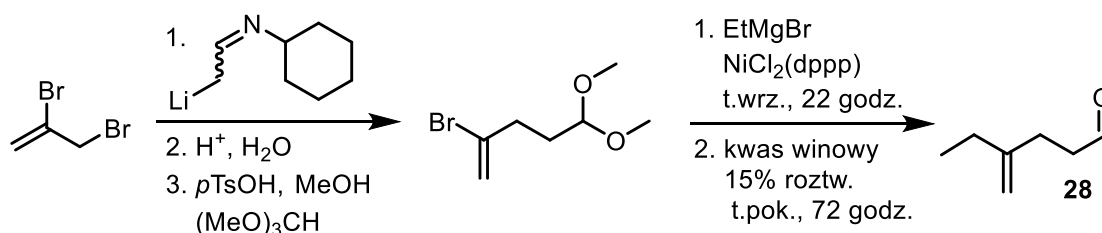
Prace nad syntezą totalną wymagają na tyle dużych nakładów pracy, że jej wykonywanie w pojedynkę rzadko jest możliwe. Syntezę totalną takamonidyny (**24**) prowadziłem wraz z Barbarą Mikulak-Klucznik. Po podziale pracy w projekcie zająłem się wykonaniem pierwszych etapów syntezy i podniesieniem ich wydajności na tyle, by móc dostarczyć jak najwięcej optycznie czynnego hydrazonu (**23(R,R)**) do następnych etapów za które dalej odpowiadała Barbara Mikulak-Klucznik.

3.3.3.1. Dyskusja planu z programu Chematica dla (**23(R,R)**)

Pierwszy wygenerowany przeze mnie plan syntezy totalnej takamonidyny, wychodził z aldehydu (**28**) i kończył się na takamonidynie (**24**). Parametry wyszukiwania pozwoliły na

zatrzymanie się analizy na znanym w literaturze aldehydzie (**28**) – zielone kółko zaznaczone żółtym rombem na (Ryc.61a). Jednak opisana w literaturze metoda³⁴⁰⁻³⁴¹ wzbudziła moje wątpliwości co do możliwości i praktycznej wygody przeprowadzenia jej w skali kilkugramowej. Dlatego po zmianie parametrów wyszukiwania wykonałem kolejną analizę retrosyntetyczną, tym razem szukając prostszej i łatwo skalowalnej metody syntezy aldehydu (**28**). Uzyskałem taką ścieżkę i po połączeniu jej z wcześniejszą analizą dysponowałem pełnym planem syntezy totalnej takamnidyny. Podział pracy w tym projekcie w przybliżeniu odpowiada dwóm składowym planom. Dlatego w niniejszej rozprawie omawiać będę wyłącznie alternatywną syntezę aldehydu (**28**) i wykonaną pierwszą reakcją z planu drugiego. Za pozostałe etapy syntezy odpowiadała Barbara Mikulak-Klucznik.

Na (Ryc.60.) przedstawiłem znaną z literatury metodę syntezy aldehydu (**28**). W pierwszej kolejności należy zaznaczyć, że aldehyd ten wykazuje niską temperaturę wrzenia, dlatego pracując z nim należy zachować szczególną ostrożność. Ta jego cecha została opisana³⁴⁰ przy okazji omawiania dlaczego ze względu na jego lotność niemożliwe było wykonanie pomiaru jego masy za pomocą spektrometrii mas. Pierwsze etapy syntezy totalnej powinny spełniać kilka wymogów: 1) wykorzystywać tanie i dostępne w dużej ilości związki wyjściowe, 2) ze względu na prowadzenie reakcji w wielogramowej skali, wykorzystanie toksycznych lub drażniących związków chemicznych nie jest wskazane, 3) reakcje powinny cechować się wysoką wydajnością i nie generować ubytków w całkowitej wydajności już na początku, 4) reakcje te powinny zachodzić szybko, kilkudniowe procedury stanowią niepotrzebne ograniczenie i spowolnienie całego procesu. Wszystkie z tych wad posiada zaprezentowana na (Ryc.60.) literaturowa metoda syntezy aldehydu (**28**).



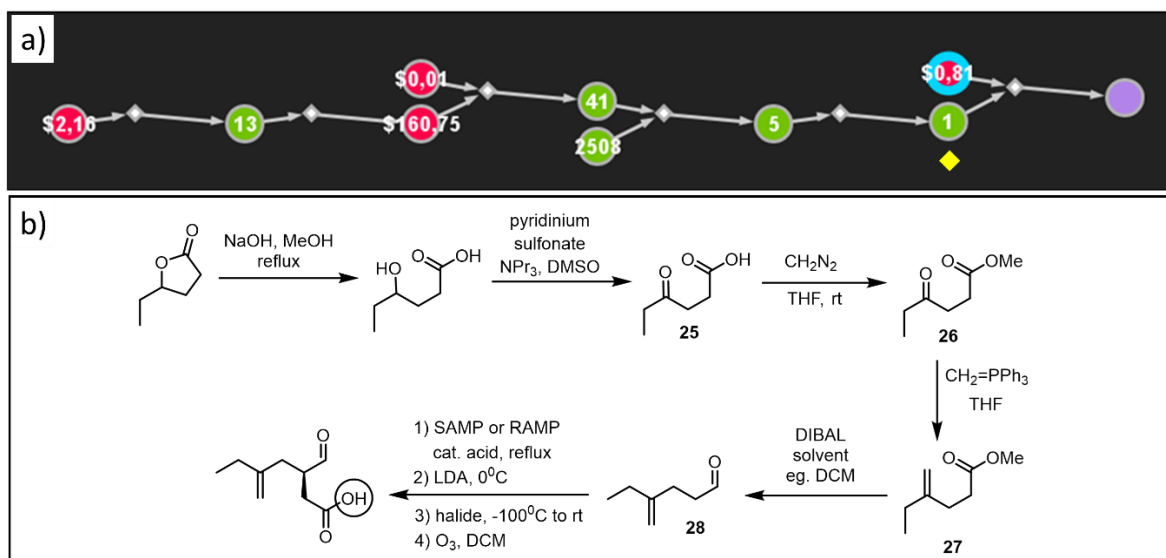
Ryc.60. Schemat syntezy aldehydu (**28**) według metody opisanej w literaturze³⁴⁰⁻³⁴¹.

Po pierwsze wyjściowy 2,3-dibromopropen jak na bromek allilowy przystało posiada drażniące właściwości. Ponadto na jego masę molową składa się głównie masa dwóch atomów bromu, które usunięte w następnych reakcjach nie wchodzą w skład produktu. Wydajność ostatnich dwóch etapów to 61%, procedura jest wieloetapowa, a hydroliza acetalu trwa trzy dni. Wszystkie te zastrzeżenia stanowiły pobudkę by za pomocą Chematici znaleźć lepszą metodę uzyskiwania aldehydu (**28**). Oczywiście nie zmniejszyłoby to jego lotności, ale być może nowa procedura ułatwiłaby pracę z nim i pozwoliła na operowanie w większej skali.

Znaleziony nowy plan syntezy wychodził z γ -heksalaktonu. Związek ten jest nietoksyczny, niedrażniący, co więcej, jest stosowany jako dodatek do żywności o certyfikowanej koszerności. Dzięki temu jest łatwo dostępny w ilościach kilogramowych za stosunkowo niską cenę. Wydaje się wyborem lepszym od 2,3-dibromopropenu również ze względu na ekonomię atomową – nie posiada masywnych grup opuszczających.

Plan Chematici (Ryc.61a-b) zakłada hydrolizę γ -heksalaktonu, utlenienie alkoholu drugorzędowego do ketonu i estryfikację grupy karboksylowej kwasu (**25**). W tak uzyskanym γ -oksoestrze (**26**) należy przeprowadzić reakcję egzometylenowania Wittiga, a uzyskanym w jej wyniku alkenem (**27**) dokonać redukcji grupy estrowej do aldehydowej przy użyciu DIBALH. Następny etap zakłada asymetryczne alkilowanie aldehydu (**28**). Zgodnie z konwencją przyjętą podczas kodowania reakcji, jeden etap w programie może wymagać przeprowadzenia więcej niż

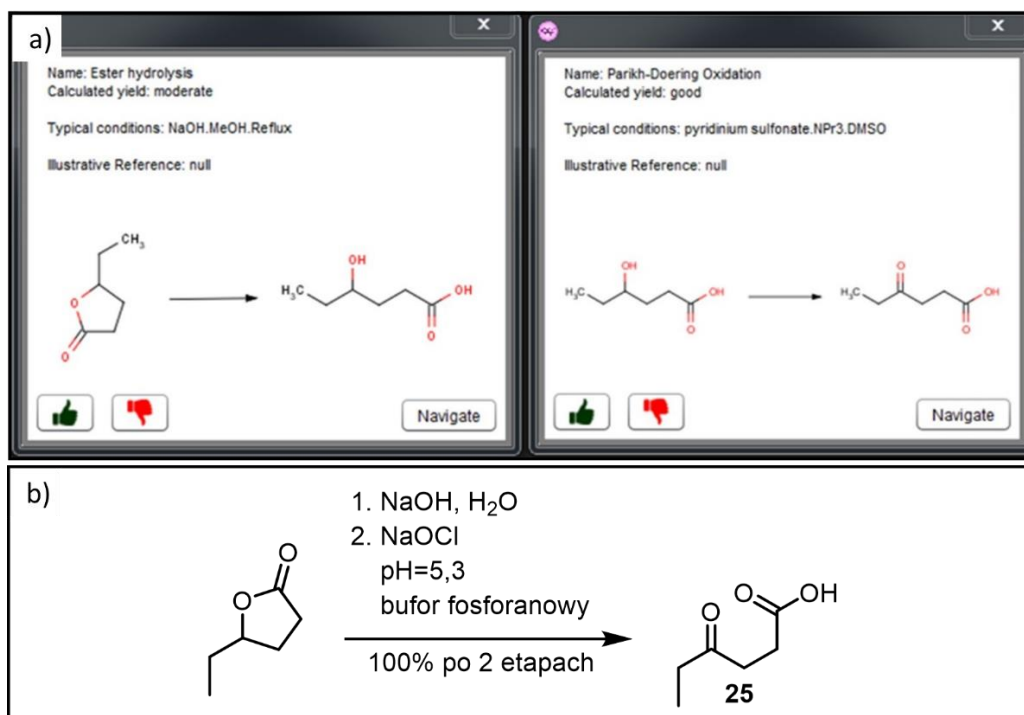
jednej reakcji. Tak jest w tym przypadku. Asymetryczne alkirowanie proponowane przez program zakłada: 1) wprowadzenie pomocnika chiralnego na grupę aldehydową, tu sugerowane jest użycie SAMP lub RAMP zgodnie z metodologią Endersa³⁴²⁻³⁴⁴, 2) deprotonowanie pozycji α hydrazonu, 3) dodanie odczynnika alkirowującego w niskiej temperaturze promującej stereoselektywne zajście reakcji, 4) usunięcie pomocnika chiralnego po reakcji i odzyskanie grupy aldehydowej. Jako, że odczynnik alkirowujący proponowany w tej reakcji zawiera niekompatybilną z warunkami alkirowania grupę karboksylową, w reakcji należy użyć go w postaci zabezpieczonej, co jest zasignalizowane za pomocą niebieskiego kółka. Faktycznie wykonałem trzy z czterech reakcji zakładanych przez ostatnią transformację. Ze względu na trwałość i niską lotność zapewnianą przez obecność w strukturze pomocnika chiralnego syntezę substratu kończyłem na uzyskaniu hydrazonu (**23(R,R)**) nie usuwając pomocnika chiralnego.



Ryc.61. a) fragment wygenerowanego przez Chematicę planu syntetycznego dla takamonidyny (**24**) omawiany w ramach niniejszej rozprawy, żółtym rombem zaznaczono miejsce połączenia dwóch planów syntetycznych. Pomiędzy żółtym rombem a takamonidyną znajduje się jeszcze pięć zaproponowanych przez Chematicę etapów syntezy – nie wchodzi one w zakres omawiany w tej pracy. Wszystkie opisy warunków reakcji podano zgodnie z propozycjami Chematici; b) ilustracja fragmentu planu syntetycznego adaptowanego do syntezy (**23(R,R)**) zawiera warunki pochodzące z reguł reakcji, grupa karboksylowa w kółku oznacza, że musi być użyta w postaci zabezpieczonej.

3.3.3.2. Adaptacja reakcji z planu, opis syntezy (**23(R,R)**)

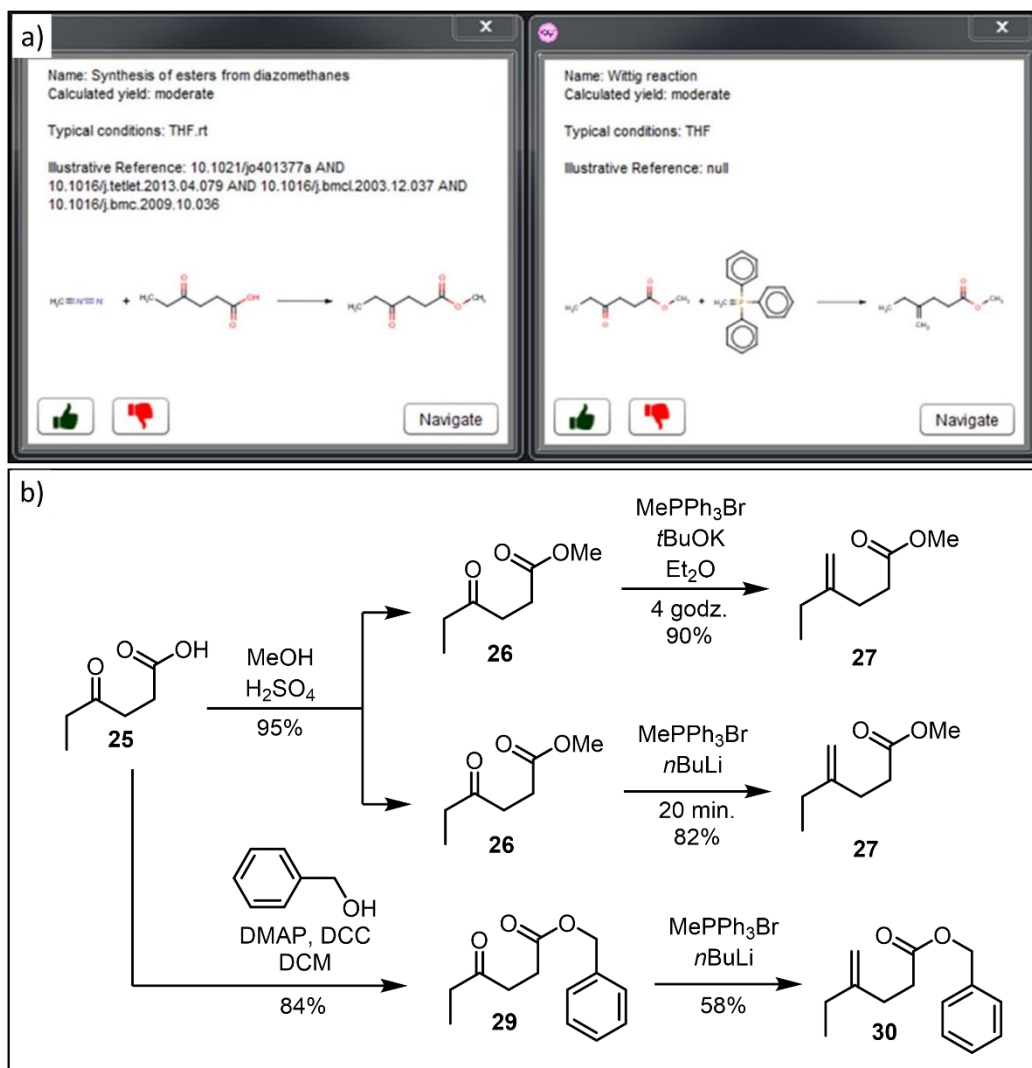
Adaptację planu przedstawionego na Ryc.61b. rozpocząłem od prób hydrolizy γ -heksalaktonu. Hydrolizę prowadziłem w środowisku zasadowym, natomiast po zakończeniu reakcji izolacja produktu w postaci kwasu polegała na zakwaszeniu mieszaniny reakcyjnej i ekstrakcji odpowiedniego γ -hydroksykwasu. Niestety po zakwaszeniu mieszaniny reakcyjnej γ -hydroksykwas spontanicznie cyklizował z powrotem do formy laktonu. Pomimo szeregu prób, nie udało mi się go uzyskać w postaci alkoholu. Dlatego postanowiłem zrealizować etapy hydrolizy i utlenienia bez izolacji kwasu. Utlenienie alkoholu do ketonu postulowane jako drugi krok, skutkuje niemożnością ponownej autocyklizacji do laktonu. Chcąc tak przeprowadzić reakcję, najpierw dokonałem hydrolizy laktonu, używając stechiometrycznej ilości NaOH. Dzięki temu po zakończeniu reakcji i odparowaniu wody, pozostałość stanowiła wyłącznie sól sodowa γ -hydroksykwasu. Taka jej postać również uniemożliwiała cyklizację do laktonu. Następnym etapem – utlenienia alkoholu – prowadziłem



Ryc.62. a) reakcje hydrolyzy estru i utlenienia alkoholu do ketonu zaproponowane przez program Chemica; b) zrealizowana syntetycznie synteza ketonu (**25**).

w środowisku wodnym. Wymuszało to uzyskanie utrwalonej postaci liniowej substratu w postaci soli sodowej. Prowadzenie reakcji w wodzie ułatwiało też precyzyjną kontrolę pH, dzięki której unikałem niepożądanego cyklizacji do laktonu. Czynnikiem utleniającym był w tym przypadku 10% wodny roztwór NaOCl. Po zakończeniu reakcji mieszaninę reakcyjną należało zakwaszyć kwasem fosforowym(V) i ekstrahować γ -oksokwas (**25**). Natychmiastowa ekstrakcja po zakwaszeniu buforu miała kluczowe znaczenie dla uzyskania produktu reakcji. W przeciwnym razie nie udało mi się wyizolować produktu. Opracowanie takiej metody syntezy kwasu (**25**) umożliwiło mi prowadzenie pierwszego etapu syntezy w skali dekagramowej.

Następnie uzyskany kwas (**25**) należało przeprowadzić w postaci estru. Chemica proponowała (Ryc.63a) wykorzystanie estru metylowego, jako metodę jego wprowadzenia sugerując wykorzystanie diazometanu. Prowadzenie reakcji z tym odczynnikiem w kilkugramowej skali nie jest przedsięwzięciem bezpiecznym. Zamiast tego postanowiłem skorzystać z estryfikacji Fischera ogrzewając kwas (**25**) w metanolu z dodatkiem H₂SO₄. Pierwsze próby kończyły się uzyskaniem śladowych ilości estru metylowego – konieczne okazało się tu dokładne wysuszenie metanolu z użyciem magnezu. Tak wysuszony rozpuszczalnik pozwolił mi uzyskać pożądaną ester (**26**) z wydajnością 95%. Tu ujawniła się pewna słabość wykorzystania krótkich fragmentów alkilowych jako bloków budulcowych. Uzyskany ester (**26**) wykazywał lotność uniemożliwiającą jego oczyszczenie na kolumnie chromatograficznej za pomocą eluentu mieszaniny heksanu z octanem etylu – w tym celu musiałem stosować mieszaninę pentanu z eterem dietylowym. Następną reakcją zaplanowaną przez Chemicę była reakcja Wittiga. Niestabilizowany ylid uzyskiwałem w dużym naczyniu Schlenka, umieściwszy w nim uprzednio suchy *t*BuOK przechowywany w komorze rękawicowej. Po zawieszeniu go w eterze dietylowym i przeniesieniu zawiesiny przy pomocy *cannuli* do drugiego naczynia Schlenka, zawierającego zawiesinę MePPh₃Br w eterze dietylowym, a następnie po dodaniu ketonu (**26**), prowadziłem reakcję kilka godzin. Przerób reakcji odbywał się



Ryc.63. a) reakcje estryfikacji i egzometylenowania Wittiga zaproponowane przez program Chematica; b) zrealizowane syntetycznie estryfikacje kwasu (**25**) i przeprowadzone dalej egzometylenowania Wittiga prowadzące do alkenów (**27**) i (**30**).

wyłącznie przy użyciu wody i eteru dietylowego. W przeciwnym razie, używając wyżej wrzących rozpuszczalników nie oddzieliłbym ich następnie od produktu, który oczyszczałem przy pomocy kolumny chromatograficznej na żelu krzemionkowym ponownie używając jako eluentów pentanu i eteru dietylowego. W przeciwieństwie do ketonu (**26**), który przy ostrożnym odparowywaniu dawało się oddzielić od eteru dietylowego i pentanu, w przypadku alkenu (**27**), okazało się to niemożliwe. Do następnej reakcji używano go w postaci roztworu w eterze i pentanie, którego stosunki molowe składników określano za pomocą ^1H NMR. Uzyskiwano w ten sposób pożądaną alken z wydajnością do 90%.

Niestety po kilku miesiącach uzyskiwania alkenu (**27**) opisanym sposobem, wydajność reakcji drastycznie spadła, do zakresu zaledwie 15-22%. Stosowanie rygorystycznie suchonych rozpuszczalników, wyprażonego szkła i odważania substratów w komorze rękawicowej nie spowodowało podniesienia wydajności. Postanowiłem wówczas uzyskać fosforowy ylid metylenowy w inny sposób. Zamiast stosować *t*BuOK, postanowiłem użyć *n*BuLi, z którym rutynowo pracowałem przy etapie asymetrycznego alkilowania. Pozwoliło to na uzyskanie ponownie

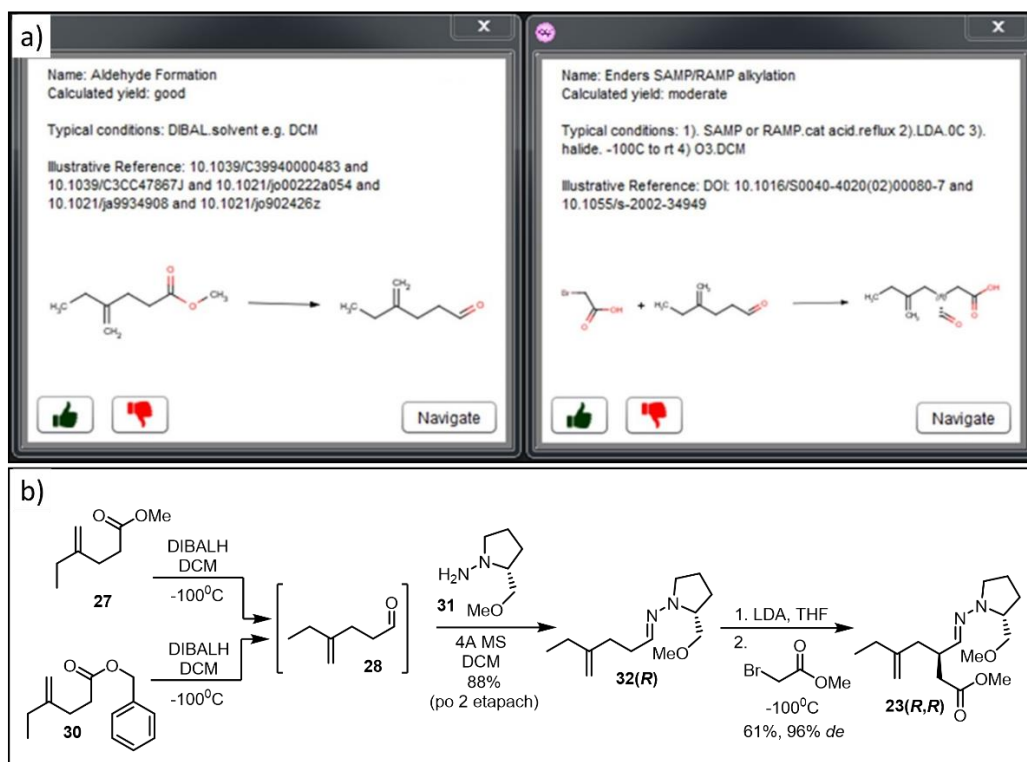
wydajności reakcji rzędu 70-80%, ostatecznie najwyższa wydajność uzyskana przez mnie tą metodą to 82%. Ponadto unikałem konieczności stosowania dwóch naczyń Schlenka, jednego na MePPh₃Br i drugiego na zawieszinę *t*BuOK. Całość reakcji prowadziłem teraz w jednym naczyniu Schlenka, nie musiałem też korzystać z komory rękawicowej – w eksperymencie kontrolnym nie zaobserwowałem różnicy wydajności reakcji przy odważaniu MePPh₃Br w komorze rękawicowej i poza nią. Zastosowana zmiana zasady spowodowała znaczące skrócenie czasu reakcji, nie eliminując jednak konieczności oczyszczania przy użyciu niskowrzących rozpuszczalników.

Stanowiło to motywację do zmiany sposobu uzyskiwania prekursora aldehydu (**28**). Pozostając wciąż w zakresie adaptacji planu z Chematici, postanowiłem użyć innego alkoholu do estryfikacji kwasu (**25**). Rodzaj podstawnika alkoksylowego w estrze nie wpływa na przebieg planu syntetycznego, wszak jest to grupa odchodząca podczas redukcji z użyciem DIBALH. Natomiast większy podstawnik powinien zmniejszyć lotność stosowanych przez mnie związków. Kierując się takim zamysłem, zamiast metanolu użyłem alkoholu benzyloвого (Ryc.63b). Zamiast estryfikacji Fischera zastosowałem estryfikację Steglich³⁴⁵, uzyskując ester (**29**) z wydajnością 84%. Zgodnie z planem ester (**29**) był możliwy do oczyszczenia za pomocą chromatografii kolumnowej z użyciem jako eluentu mieszaniny heksanu z octanem etylu. Niestety wydajność reakcji Wittiga z jego użyciem skutkowałą wydajnością 34%, którą udało się podnieść do 58%. Produkt reakcji Wittiga zawierający ester benzylovery (**30**) podobnie jak ester metylovery (**27**) okazał się lotny i w jego przypadku stosowano podobne oczyszczanie przy użyciu eteru dietylowego i pentanu.

W następnym kroku, plan Chematici zakładał redukcję estru (**27**) za pomocą DIBALH do aldehydu (**28**). Reakcję tą udało mi się zrealizować stosując zaproponowany reduktor i niską temperaturę. Uzyskiwane po reakcji Wittiga roztwory estrów (**27**) i (**30**) rozcieńczałem dodatkiem DCM i schładzałem do -100°C. Powolne wkraplanie roztworu DIBALH w DCM pozwalało mi na zatrzymanie redukcji na etapie aldehydu. Przeróbka mieszaniny reakcyjnej ponownie była utrudniona przez wysoką lotność uzyskiwanego aldehydu oraz dodatkowo przez łatwe tworzenie emulsji przez DCM i alkiloglinowe pozostałości reduktora. Opracowałem minimalizującą straty aldehydu metodę przerobu mieszaniny reakcyjnej: aldehydu (**28**) nie izolowałem w postaci czystej, lecz wysuszony roztwór po ekstrakcji bezpośrednio wykorzystywałem w następnym etapie, tj. zakładania nań pomocnika chiralnego RAMP (**31**). W przeciwieństwie do aldehydu (**28**), jego hydrazon (**32(R)**) odznacza się wysoką temperaturą wrzenia, co pozwala na oczyszczenie go po reakcji za pomocą kolumny chromatograficznej. Początkowo uzyskiwałem hydrazon (**32(R)**) z wydajnościami rzędu 15-31%. Oznaczało to zarówno obniżenie wydajności całej syntezy, ograniczenie dostępności substratu do następnych etapów jak i marnotrawstwo pomocnika chiralnego traconego w niskowydajnej reakcji. Optymalizacja tej sekwencji pozwoliła mi na podniesienie jej całkowitej wydajności do 88%.

W tym miejscu należy zaznaczyć, że ilość moli estru (**27**) lub (**30**) obliczona w oparciu o widmo ¹H NMR była kluczowa dla przeprowadzenia sekwencji. Zaniżenie ilości estru skutkowało użyciem nadmiaru DIBALH, który redukował aldehyd dalej do alkoholu. We wstępnych eksperymentach po takiej redukcji do alkoholu konieczne było jego ponowne utlenienie przy pomocy PCC. Przyjęta ilość moli aldehydu wpływała również na stosowaną ilość pomocnika chiralnego, który ze względu na jego cenę nie mógł być używany w nadmiarze.

Reakcja deprotonowania przy użyciu LDA – druga w zaproponowanej przez Chematicę sekwencji asymetrycznego alkilowania (Ryc.64.) – wymagała przygotowania LDA *in situ*. Wszelkie próby zastosowania komercyjnie dostępnego LDA skończyły się niepowodzeniem. Dopiero gdy samodzielnie generowałem LDA, a następnie dodawałem (**32(R)**) skutecznie zachodziło deprotonowanie hydrazonu. Reakcja asymetrycznego alkilowania przez długi czas zachodziła z wydajnościami rzędu 20-30%. Dopiero wielokrotne jej powtórzenie i utrzymywanie reżimu bezwodnych warunków pozwoliło mi na podniesienie jej wydajności do 61%.

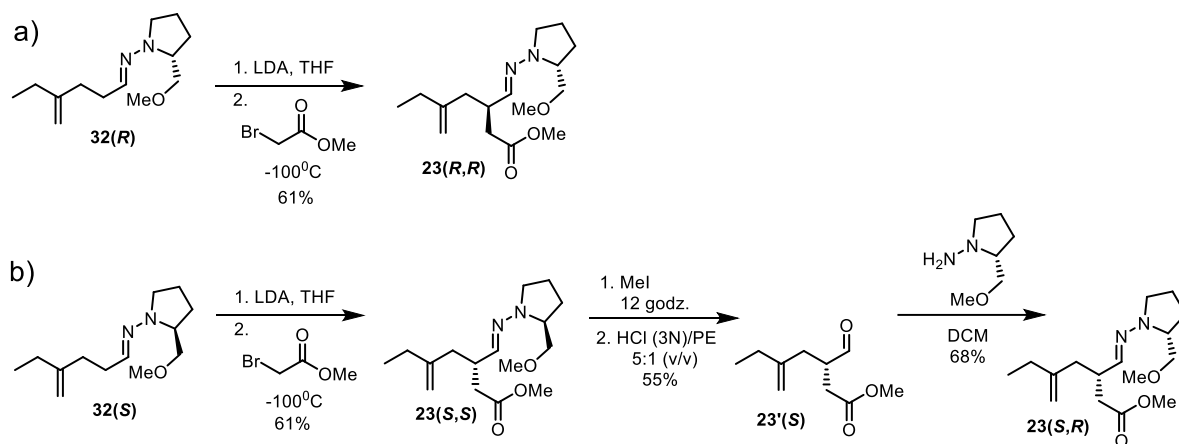


Ryc.64. a) reakcje redukcji estru do aldehydu i sekwencja asymetrycznego alkirowania Endersa zaproponowane przez program Chematica; b) wykonane syntetycznie asymetryczne alkirowanie prowadzące do hydrazone (**23(R,R)**).

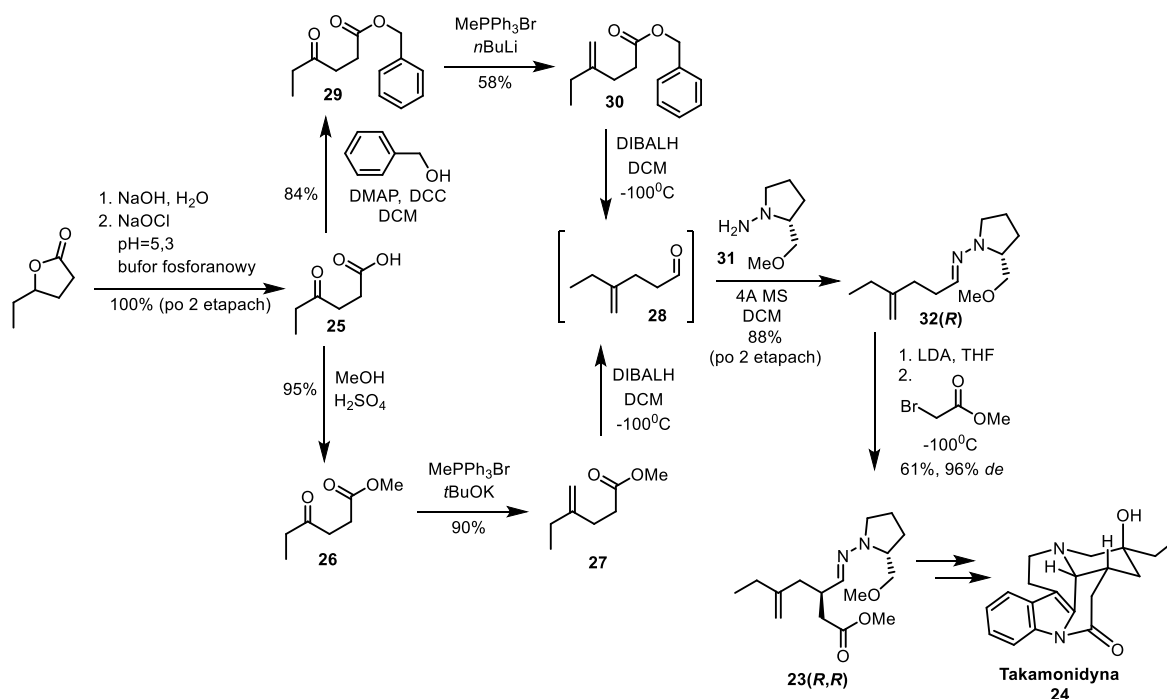
Po przeprowadzeniu reakcji asymetrycznej należało ustalić z jakim nadmiarem diastereoizomerycznym uzyskałem (**23(R,R)**). Oznaczenia dokonywałem przy wciąż obecnym pomocniku chiralnym i dwóch centrach stereogenicznych w cząsteczce.

W celu oznaczenia nadmiaru diastereoizomerycznego musiałem uzyskać jako wzorzec racemiczny produkt alkirowania, a następnie przyłączyć doń pomocnik chiralny by uzyskać wzorcową parę diastereoizomerów, możliwych do uzyskania w reakcji. Ponadto pomiar samego aldehydu po rozdziale przy użyciu chiralnego HPLC i detekcji za pomocą detektora UV był wykluczony ze względu na brak odpowiedniego chromoforu w aldehydzie. Możliwy był dopiero pomiar UV pochodnej w postaci dowolnego hydrazone. Również ze względu na wysoką lotność aldehydu (**28**) i dysponowanie nim wyłącznie w roztworze chlorowcowanych rozpuszczalników, bezpośrednie jego alkirowanie nie wchodziło w grę. Postanowiłem zatem przeprowadzić aldehyd w nieczynny optycznie hydrazone i następnie alkirować go, uzyskując mieszaninę racemiczną. Następnie usunąć hydrazone i na jego miejsce przyłączyć pomocnik chiralny. Tym sposobem uzyskałem hydrazone dimetylowy z wydajnością 44%, próba jego alkirowania w dotychczas stosowanych warunkach doprowadziła do częściowej konwersji, lecz produkt rozkładał się podczas oczyszczania za pomocą chromatografii na żelu krzemionkowym. Następnie podjąłem próbę przy użyciu hydrazone *N,N*-difenylowego, jednak w jego przypadku rozkład obserwowałem już podczas próby jego oczyszczania za pomocą kolumny na żelu krzemionkowym, zanim jeszcze przystąpiłem do alkirowania. Wobec tego postanowiłem w inny sposób uzyskać wzorzec diastereoizomeru powstającego w mniejszości. We właściwej reakcji teoretycznie powinienem uzyskać główny produkt (**23(R,R)**) o konfiguracji (*R,R*) i w mniejszości produkt (**23(R,S)**) o konfiguracji (*R,S*). Ta mieszanina diastereoizomerów nie była możliwa do rozdzielania za pomocą niechiralnej, preparatywnej

kolumny chromatograficznej. Produkt (**23(R,S)**) uzyskałem w następującej sekwencji (Ryc.65b). Prowadząc alkirowanie z pomocnikiem chiralnym SAMP otrzymałem produkt (**23(S,S)**), następnie usunąłem pomocnik chiralny SAMP, a na jego miejsce wprowadziłem pomocnik RAMP. Dzięki temu dysponowałem parą produktów alkirowania z oboma pomocnikami, szczęśliwie powstający w mniejszości izomer przy użyciu SAMP jest głównym produktem alkirowania z RAMP. Dlatego obserwowałem na HPLC wyłącznie dwa sygnały od obu spodziewanych izomerów (**23(R,R)**) i (**23(R,S)**), co pozwoliło mi na określenie nadmiaru diastereoizomerycznego w tej reakcji – wyniósł on 96%*de*.



Ryc.65. a) asymetryczne alkirowanie prowadzące do założonego produktu (**23(R,R)**); b) sposób uzyskania związku (**23(S,R)**) służącego określeniu nadmiaru diastereoizomerycznego dla reakcji z punktu a).



Ryc.66. Schemat syntezy hydrazonu (**23(R,R)**) wykonanej w oparciu o plan uzyskany z programu Chematica.

3.3.3.3. Podsumowanie

Zgodnie z założonym planem z Chematici, uzyskałem hydrazon (**23(R,R)**). Stanowi on substrat w większym planie syntezy totalnej takamonidyny (**24**). Zrealizowaną metodę jego syntezy zaprezentowałem na Ryc.66. Będąc przygotowanym na problemy związane z lotnością występującego w planie substratu (**28**), udało mi się opracować sposób prowadzenia reakcji i oczyszczania mieszaniny reakcyjnej, który nie powodował jego poważnego ubytku. Dzięki temu zaadaptowany plan Chematici mogłem wykonywać w skali wielogramowej, każdorazowo wychodząc z 10 g laktonu. Również zmiana zasady w reakcji Wittiga pozwoliła na znaczne skrócenie czasu prowadzenia reakcji oraz ułatwiła manualne jej prowadzenie. Zaproponowana przez Chematicę metoda wprowadzania pierwszego centrum stereogenicznego do cząsteczki pozwoliła na uzyskanie zadowalającego nadmiaru diastereoizomerycznego hydrazonu (**23(R,R)**).

Wszystkie reakcje założone w planie syntetycznym udało się wykonać, konieczne oczywiście było dobieranie warunków do każdego etapu, niemniej główna idea wykorzystująca zmianę grup funkcyjnych zlokalizowanych już w laktonie w pozycjach 1,4 – okazała się wykonalna, umożliwiając dalsze prace nad syntezą totalną takamonidyny (**24**).

3.4 Podsumowanie i wnioski

Podsumowując badania przedstawione w niniejszej rozprawie doktorskiej, czas odnieść się do przedstawionej we wstępie hipotezy badawczej: „Zdolność projektowania wykonalnych planów syntetycznych – cząsteczek organicznych – nie jest umiejętnością wyłącznie ludzką, lecz możliwe jest stworzenie programu komputerowego zdolność tą posiadającego.”

Począwszy od rozpoczęcia ponad pół wieku temu przez E. J. Corey’a prac nad LHASA, do dziś trwają prace stawiające sobie za cel stworzenie programu komputerowego zdolnego zrozumieć specyfikę chemii organicznej i w oparciu o to zrozumienie proponować analizy retrosyntetyczne oraz plany syntetyczne. W niniejszej pracy przedstawiłem przykłady najważniejszych prób zmierzania się z tym problemem. Wszystkie one kończyły się niepowodzeniami, a problemy odkrywane podczas tych prac zdawały się udowadniać, że jest to cel w istocie swej niemożliwy do osiągnięcia. Dopiero podejście prof. Grzybowskiego w postaci programu Chematica pozwoliło przełamać impas w dziedzinie komputerowego planowania syntez chemicznych.

Pierwszym celem moich badań było wykazanie, że *po skodyfikowaniu reguł planowania syntezy, nie ma różnicy czy reguły te stosuje człowiek czy komputer*. Realizacja tego celu wymagała stworzenia bazy reguł reakcji chemicznych. Cel ten realizowałem na początku opisanych badań wraz z innymi członkami zespołu. W wyniku tych prac wykazaliśmy²²⁰, że program komputerowy może stworzyć *de novo* już istniejące plany syntetyczne. Nie sposób było wskazać różnicy między planem zaproponowanym przez komputer, a wykonanym przez człowieka. Oznaczało to osiągnięcie pierwszego celu badań, co spotkało się ze znaczącym zainteresowaniem mediów³⁴⁶⁻³⁴⁷.

Drugim celem moich badań było: *wykazanie wykonalności syntetycznej planów wygenerowanych przez program Chematica/SynthiaTM, oraz porównanie tych ścieżek syntezy z istniejącymi i stworzonymi przez człowieka dla tych samych związków*. Realizację tego celu rozpocząłem owocną syntezą leku Dronedaron (**1**). Następnie walidację przeprowadzono na próbie 8 związków. Brałem udział w tej walidacji, wykonując z sukcesem najdłuższą zaplanowaną przez Chematicę syntezę: 5β/6β-hydroksylurazydonu (**12,12'**). W przypadku tego planu i pozostałych zaproponowanych przez program komputerowy planów syntetycznych z grupy testowej, osiągnięty rezultat był lepszy niż osiągnięty uprzednio przez człowieka. Osiągnięty rezultat wykroczał ponad założony cel badań. Nie tylko wykazano wykonalność planów syntetycznych, lecz porównanie ich

z istniejącymi jednoznacznie wskazało wyższość planów generowanych przy pomocy programu komputerowego. Należy zaznaczyć, że tak duży zbiór testowy nigdy nie został opublikowany jako dowód poprawności działania programu wspomagającego planowanie syntezy. We wstępie omówiono historyczne, nieliczne przypadki walidacji eksperymentalnej, nigdy nie była to więcej niż jedna synteza dla jednego programu, przedstawione syntezy nie były też lepsze od literaturowych. Tu na dowód przedstawiono 8 syntez²⁷² – jest to jak dotąd najmocniejszy znany w literaturze dowód poprawności działania takiego programu. Publikacja tego dowodu, rozwiązującego po ponad 50 latach jedno z najtrudniejszych wyzwań współczesnej chemii organicznej, spotkała się ze znacznym, globalnym zainteresowaniem w środowisku chemików³⁴⁸⁻³⁵³.

Nie oznacza to końca prac nad programem Chematica, lecz stanowi zaledwie wstęp do jego dalszego rozwoju i podejmowania trudniejszych wyzwań. Dalsza walidacja programu Chematica na przykładzie syntezy związku naturalnego, takamonidyny (**24**), wymagającego kilkunastoetapowej syntezy, w stopniu obejmowanym niniejszą rozprawą (pierwsze etapy syntezy) daje wynik pozytywny i nadzieje, że kierując się przyjętym założeniem możliwe będzie jeszcze lepsze planowanie syntezy chemicznej w przyszłości.

Dotychczasowa historia rozwoju metod planowania syntezy chemicznej stanowi dobry wyznacznik rozwoju całej chemii organicznej. W XVIII wieku Immanuel Kant przewidywał:

*„chemia będzie niczym więcej jak tylko systematyczną sztuką albo eksperymentalną teorią, ale nigdy nie stanie się nauką we właściwym sensie, ponieważ zasady jej są czysto empiryczne i w żaden sposób nie mogą być przedstawione a priori w oglądzie, a zatem w najmniejszym stopniu nie da się zrozumieć możliwości podstawowych zasad zjawisk chemicznych, skoro nie stosuje się do nich matematyka.”*³⁵⁴

Pisząc to Kant nie miał oczywiście pojęcia o tym jak 250 lat później będzie wyglądać chemia organiczna, a jednak mimo to wyjątkowo trafnie zidentyfikował leżące przed nią wyzwanie i kierunek rozwoju. Dotąd planowanie syntezy w ramach chemii organicznej opierało się w głównej mierze na niewymiernej intuicji chemika, później wspartego metodą retrosyntetyczną. Jednak ciągle brak było reguł i kryteriów ilościowych pozwalających porównywać i oceniać plany syntetyczne w sposób systematyczny. Wraz z oferowanymi przez komputery możliwościami subtelnego operowania na dużych drzewach retrosyntetycznych, postulowany przez Kanta przełom dokonuje się współcześnie.

4. Część eksperymentalna

4.1. Stosowane metody syntetyczne i analityczne

Odczynniki i rozpuszczalniki pochodziły ze źródeł komercyjnych (Sigma-Aldrich, ABCR, POCH, Chempur, Angene Chemicals). Wszystkie odczynniki były używane bez dodatkowego oczyszczania, chyba, że zaznaczono inaczej. Chromatografię kolumnową wykonywano z wykorzystaniem żelu krzemionkowego firmy Merck typ 60 (230-400 mesh, 40-63 μm). Postęp reakcji monitorowano za pomocą płytek do chromatografii cienkowarstwowej, typ Macherey-Nagel żel krzemionkowy 60F254 na folii aluminiowej i Merck żel krzemionkowy 60F254 na folii aluminiowej. Do wywoływania sygnałów na płytkach TLC stosowano: promieniowanie UV w długościach (254 i 365nm), pary jodu lub wywoływacz cerowo-molibdenowy (wywoływacz Hanessian).

Wszystkie rozpuszczalniki suszono za pomocą sit molekularnych (3 lub 4 \AA). Rozpuszczalniki (THF, Et_2O , DCM, MeOH, $i\text{Pr}_2\text{NH}$) stosowane w reakcjach wrażliwych na wilgoć były dodatkowo destylowane zgodnie z procedurami literaturowymi³⁵⁵ i przetrzymywane w naczyniach Schlenka w atmosferze gazu obojętnego. Procedury wykorzystujące tak przygotowane rozpuszczalniki zostały poniżej określone jako bezwodne.

Bufor fosforanowy wykonano według następującej procedury: do przygotowania 100 ml buforu: 13,8 g $\text{NaH}_2\text{PO}_4 \times 2 \text{H}_2\text{O}$ i 4,0 g $\text{Na}_2\text{HPO}_4 \times 12 \text{H}_2\text{O}$ umieszczono w kolbie płaskodennej, następnie dodano 100 ml wody dejonizowanej i mieszano przez 30-40 minut, aż do całkowitego rozpuszczenia soli. Pomiaru pH dokonywano przy użyciu pH-metru (Seven2Go, Metler Toledo) i jeżeli było to konieczne dla uzyskania pH 5,3, dodawano odpowiednio H_3PO_4 (85%) lub NaOH (stały). Tak przygotowany bufor, przetrzymywany w temperaturze pokojowej był stabilny przez 3-4 tygodnie.

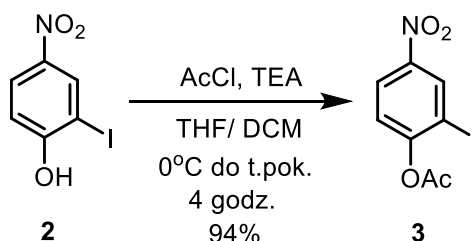
Pomiarów widm NMR dokonywano przy użyciu aparatów Bruker 400 MHz Avance III spectrometer lub Varian VNMRS 600 MHz spectrometer w temperaturach pokojowych. Przesunięcia chemiczne (δ) wyrażono w częściach na million (ppm) w odniesieniu do sygnału odpowiedniego rozpuszczalnika referencyjnego (ref: CHCl_3 [^1H : 7.26, ^{13}C : 77.2]) z dokładnością 0,01. Stałe sprzężenia (J) wyrażono w Hz z dokładnością do 0,1 Hz. Multipletowość sygnałów oznaczono jak następuje: s (singlet), d (dublet), t (tryplet), q (kwartet), qi (kwintet), sx (sekstet) i m (multiplet).

Wysokorozdzielcze widma masowe HRMS rejestrowano przy użyciu aparatu AutoSpec Premier (Waters) lub MaldiSYNAPT G2-S HDMS (Waters) a wyrażono w jednostce m/z.

Nadmiary enancjomeryczne i diastereoizomeryczne mierzono przy użyciu chiralnych kolumn HPLC model OD-H Daicel Chiralcel, za pomocą aparatu Jasco PU-2089PLUS, o detektorze: Jasco MD-2010PLUS.

4.2. Synteza Dronedaronu

4.2.1. Procedury syntetyczne zastosowane do uzyskania Dronedaronu (1)

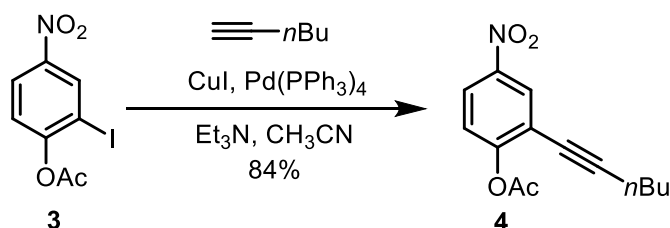


Octan 2-jodo-4-nitrofenolu (3): Do roztworu 2-jodo-4-nitrofenolu (**2**) (1,143 mmol, 0,303 g) [uzyskanego wg. procedury literaturowej²⁹⁹] w THF (0,7 ml) i DCM (0,7 ml), dodano Et₃N (0,350 ml, 1,372 mmol), a następnie schłodzono do 0°C i wkraplano AcCl (0,098 mL, 1,372 mmol). Po 30 minutach mieszania w 0°C, pozwolono mieszaninie reakcyjnej ogrzać się do temperatury pokojowej. Mieszano następnie 3,5 godziny, po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej wodę i DCM. Rozdzielono warstwy, a fazę wodną ekstrahowano trzykrotnie za pomocą DCM. Połączone warstwy organiczne suszono za pomocą bezwodnego Na₂SO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (5:1), uzyskując **3** w postaci białego ciała stałego (0,303 g, 94%).

¹H NMR (400 MHz, CDCl₃) δ 8.71 (d, *J* = 2.6 Hz, 1H), 8.27 (dd, *J* = 8.9, 2.6 Hz, 1H), 7.30 (d, *J* = 8.9 Hz, 1H), 2.43 (s, 3H).

¹³C NMR (101 MHz, CDCl₃): δ 167.57, 156.15, 145.73, 134.78, 124.74, 123.27, 90.44, 21.15.

HRMS (EI⁺): *m/z* obliczone dla: C₈H₆INO₄: 306.9342 , zmierzono: 306.9343.



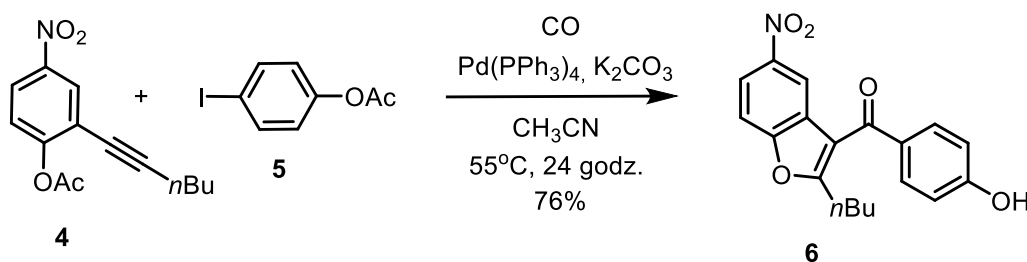
Octan 2-(heks-1-yn-1-ylo)-4-nitrofenolu (4): W kolbie okrągłodennej umieszczono **3** (3,482 mmol, 1,069 g), Pd(PPh₃)₄ (0,052 mmol, 0,060 g) i CuI (0,108 mmol, 0,021 g). Kolbę zamknięto gumowym septum, wymieniono atmosferę na argon i zamocowano balon z argonem. Następnie podawano przez strzykawkę odgazowany THF (25 mL), Et₃N (34,820 mmol, 4,850 mL) i 1-heksyn (6,963 mmol, 0,800 mL). Mieszaninę reakcyjną mieszano w temperaturze pokojowej przez 5,5 godz. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej wodę i AcOEt. Rozdzielono warstwy, a fazę wodną ekstrahowano trzykrotnie za pomocą AcOEt. Połączone warstwy organiczne suszono za pomocą bezwodnego Na₂SO₄, następnie odsączono

środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (20:1), uzyskując **4** w postaci żółtego ciała stałego (0,761 g, 84%).

^1H NMR (400 MHz, CDCl_3) δ 8.32 (d, $J = 2.7$ Hz, 1H), 8.16 (dd, $J = 8.9, 2.7$ Hz, 1H), 7.25 (d, $J = 8.9$ Hz, 1H), 2.47 (t, $J = 7.0$ Hz, 2H), 2.38 (s, 3H), 1.66 – 1.57 (m, 2H), 1.56 – 1.45 (m, 2H), 0.98 (t, $J = 7.3$ Hz, 3H).

^{13}C NMR (101 MHz, CDCl_3): δ 167.85, 156.06, 145.33, 128.43, 123.61, 123.12, 119.80, 98.59, 73.92, 30.44, 29.68, 21.90, 20.74, 19.18, 13.53.

HRMS (EI⁺): m/z obliczone dla: $\text{C}_{14}\text{H}_{15}\text{NO}_4$: 261.1001, zmierzono: 261.0997.

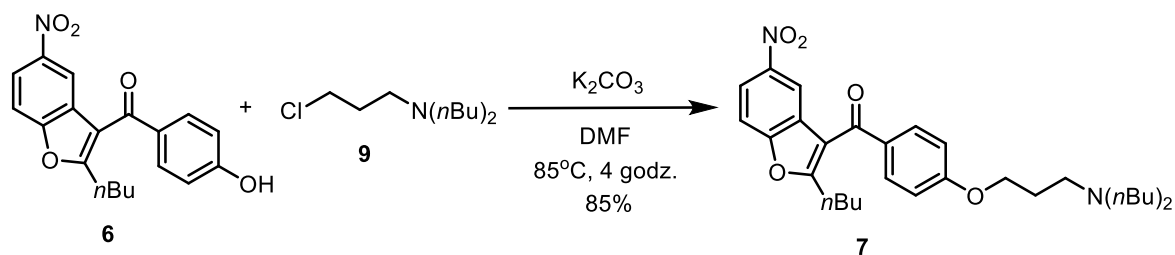


(2-butyl-5-nitrobenzofurano-3-yl)(4-hydroxyphenyl)methanone (6): W kolbie okrągłodennej umieszczono **4** (0,491 mmol, 0,128 g) i **5** (1,576 mmol, 0,913 g), $\text{Pd(PPh}_3)_4$ (0,030 mmol, 0,034 g) i bezwodny (!) K_2CO_3 (2,357 mmol, 0,326 g). Kolbę zamknięto gumowym septum, trzykrotnie wymieniano atmosferę na CO i zamocowano balon z CO. Następnie podano przez strzykawkę CH_3CN (1,2 ml), który został uprzednio odgazowany i nasycony CO. Mieszaninę reakcyjną mieszano w temperaturze 55°C przez 24 godz. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej wodę i DCM. Rozdzielono warstwy, a fazę wodną ekstrahowano trzykrotnie za pomocą DCM. Połączone warstwy organiczne suszono za pomocą bezwodnego Na_2SO_4 , następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (15:1), uzyskując **6** w postaci białego ciała stałego (0,126 g, 76%).

^1H NMR (400 MHz, CDCl_3) δ 8.36 (d, $J = 2.3$ Hz, 1H), 8.24 (dd, $J = 9.0, 2.4$ Hz, 1H), 7.81 (d, $J = 8.7$ Hz, 2H), 7.59 (d, $J = 9.0$ Hz, 1H), 6.98 (d, $J = 8.7$ Hz, 2H), 6.42 (s, 1H), 2.95 (t, 2H), 1.79 (q, $J = 7.6$ Hz, 2H), 1.37 (sx, $J = 7.4$ Hz, 2H), 0.91 (t, $J = 7.4$ Hz, 3H).

^{13}C NMR (101 MHz, CDCl_3): δ 189.35, 168.60, 168.01, 156.30, 154.65, 144.80, 135.82, 130.70, 127.59, 122.02, 120.41, 117.73, 117.01, 111.44, 29.90, 29.68, 28.09, 22.30, 21.17, 13.59.

HRMS (EI⁺): m/z obliczone dla: $\text{C}_{19}\text{H}_{17}\text{NO}_5$: 330.1107, zmierzono: 330.1104.



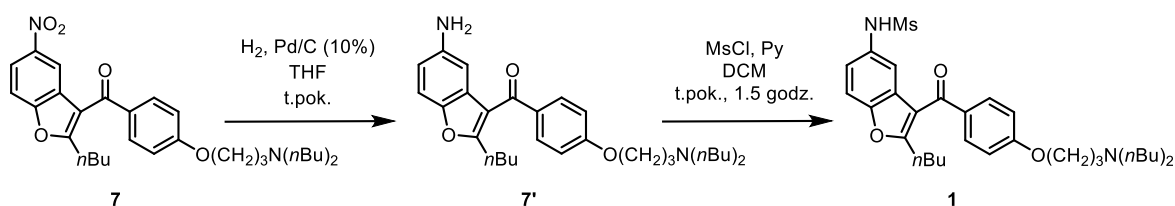
(2-butyl-5-nitrobenzofurano-3-ylo)(4-(3-(dibutylamino)propoksy)fenylo)metanon (7):

W kolbie okrągłodennej umieszczono **6** (0,730 mmol, 0,248 g), **9** (0,803 mmol, 0,165 g) i K_2CO_3 (0,730 mmol, 0,101 g). Kolbę zamknięto gumowym septum, wymieniano atmosferę na argon i zamocowano balon z argonem. Następnie podano przez strzykawkę DMF (4,5 ml). Mieszaninę reakcyjną mieszano w temperaturze 85°C przez 4 godz. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej wodę i DCM. Rozdzielono warstwy, a fazę wodną ekstrahowano trzykrotnie za pomocą DCM. Połączone warstwy organiczne suszono za pomocą bezwodnego Na_2SO_4 , następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie $CHCl_3:MeOH$ (30:1), uzyskując **7** w postaci żółtego oleju (0,316 g, 85%).

1H NMR (400 MHz, $CDCl_3$) δ 8.36 (d, 1H), 8.23 (dd, $J = 9.0, 2.3$ Hz, 1H), 7.83 (d, $J = 8.7$ Hz, 2H), 7.57 (d, $J = 9.0$ Hz, 1H), 7.00 (d, $J = 8.8$ Hz, 2H), 4.14 (t, $J = 6.3$ Hz, 2H), 2.93 (t, $J = 7.6$ Hz, 2H), 2.64 (t, $J = 6.9$ Hz, 2H), 2.51 – 2.38 (m, 4H), 2.03 – 1.91 (m, 2H), 1.83 – 1.72 (m, 2H), 1.48 – 1.40 (m, 4H), 1.38 – 1.29 (m, 6H), 0.91 (t, 9H).

^{13}C NMR (101 MHz, $CDCl_3$) δ 189.02, 167.01, 163.64, 156.30, 144.64, 131.69, 130.75, 127.94, 120.16, 117.69, 117.34, 114.48, 111.33, 66.62, 53.95, 50.33, 29.90, 29.47, 29.06, 27.93, 27.00, 22.31, 20.68, 14.05, 13.62.

HRMS (ESI+): m/z $[M+H]^+$ obliczone dla: $C_{30}H_{41}N_2O_5$: 509.3015, zmierzono: 509.3008.



***N*-(2-butyl-3-(4-(3-(dibutylamino)propoksy)benzoilo)benzofurano)metanosulfonamid (1):**

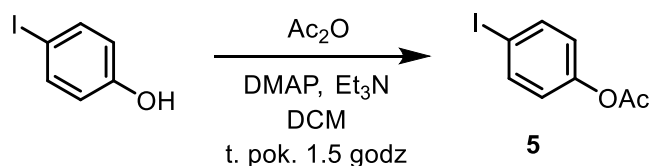
W kolbie okrągłodennej umieszczono **7** (0,230 mmol, 0,117 g) i rozpuszczono w THF (3 ml). Kolbę zamknięto gumowym septum, wymieniano atmosferę na argon, następnie ostrożnie zdjęto septum, wsypano Pd/C (10%) (0,028 g) i ponownie zamocowano septum. Wymieniono atmosferę na H_2 i zamocowano balon z H_2 . Mieszaninę reakcyjną mieszano w temperaturze pokojowej przez noc. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Mieszaninę reakcyjną przesączono przez warstwę celitu i trzykrotnie przepłukano AcOEt. Rozpuszczalniki odparowano za pomocą wyparki obrotowej, uzyskując surowy **7'** w postaci żółtego oleju (0,112 g), który użyto do następnego etapu bez oczyszczania.

Surową aminę **7'** (0.112 g) rozpuszczono w DCM (3 ml), następnie dodano pirydynę (0,351 mmol, 0,025 ml) i ostrożnie wkroplono MsCl (0,234 mmol, 0,018 ml). Mieszaninę reakcyjną mieszano w temperaturze pokojowej przez 1,5 godz. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej wodę i DCM. Rozdzielono warstwy, a fazę wodną ekstrahowano trzykrotnie za pomocą DCM. Połączone warstwy organiczne suszono za pomocą bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie CHCl₃:MeOH (40:1), uzyskując **1** w postaci żółtego oleju (0,094 g, 73% - po 2 etapach).

¹H NMR (400 MHz, CDCl₃) δ 7.83 (d, *J* = 8.8 Hz, 2H), 7.48 (d, *J* = 8.6 Hz, 1H), 7.32 – 7.26 (m, 2H), 6.98 (d, *J* = 8.8 Hz, 2H), 4.14 (t, *J* = 6.3 Hz, 2H), 2.95 (s, 3H), 2.89 (t, 2H), 2.65 (t, *J* = 6.8 Hz, 2H), 2.47 (t, 4H), 2.02 – 1.93 (m, 2H), 1.76 (qi, *J* = 7.6 Hz, 2H), 1.50 – 1.40 (m, 4H), 1.39 – 1.28 (m, 8H), 0.95 – 0.86 (m, 9H).

¹³C NMR (101 MHz, CDCl₃) δ 190.20, 165.79, 163.25, 151.84, 132.38, 131.65, 131.30, 128.26, 120.14, 116.77, 115.55, 114.29, 111.78, 66.54, 53.88, 50.35, 39.05, 30.02, 29.69, 29.13, 28.00, 26.93, 22.31, 20.68, 14.05, 13.65.

HRMS (ESI⁺): *m/z* [M+H⁺] obliczone dla: C₃₁H₄₅N₂O₅S: 557.3049 , zmierzono: 557.3048.

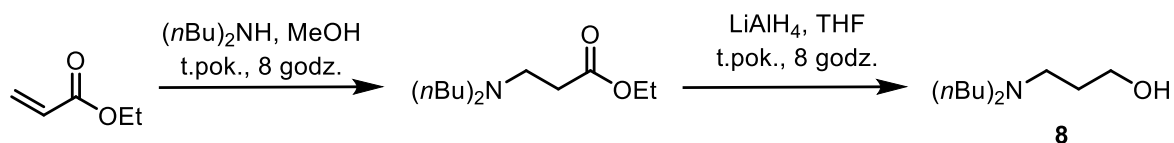


Octan 4-jodofenolu (5): W kolbie okrągłodennej umieszczono 4-jodofenol (10,06 mmol, 2,214 g) i DMAP (0,60 mmol, 0,074 g), rozpuszczono w DCM (25 ml). Kolbę zamknięto gumowym septum, wymieniało atmosferę na argon i zamocowano balon z argonem. Następnie podano przez strzykawkę Et₃N (8,05 mmol, 1,12 ml) i Ac₂O (16,10 mmol, 1,52 ml). Mieszaninę reakcyjną mieszano w temperaturze pokojowej przez 1,5 godz. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej wodę i DCM. Rozdzielono warstwy, a fazę wodną ekstrahowano trzykrotnie za pomocą DCM. Połączone warstwy organiczne suszono za pomocą bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (7:1), uzyskując **5** w postaci białego ciała stałego (2,589 g, 98%).

¹H NMR (400 MHz, CDCl₃) δ 7.71 (d, *J* = 8.8 Hz, 1H), 6.89 (d, *J* = 8.8 Hz, 1H), 2.31 (s, 2H).

¹³C NMR (101 MHz, CDCl₃) δ 168.99, 150.54, 138.47, 123.77, 89.81, 21.08.

HRMS (EI⁺): *m/z* obliczone dla: C₈H₇O₂: 261.9491 , zmierzono: 261.9500.



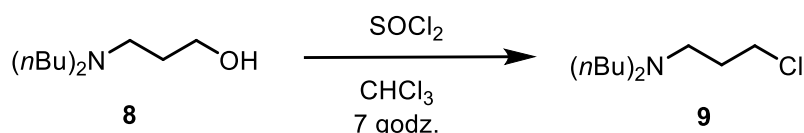
3-(dibutyloamino)propanol (8): W kolbie okrągłodennej umieszczono akrylan etylu (9,9 mmol, 0,991 g) i dibutyloaminę (9,0 mmol, 1,151 g), dodano MeOH (12 ml). Kolbę zamknięto gumowym septum. Mieszaninę reakcyjną mieszano w temperaturze pokojowej przez 8 godz. Po tym czasie rozpuszczalnik odparowano za pomocą wyparki obrotowej, uzyskując **8** w postaci klarownego, bezbarwnego oleju (2,095 g), użyto go do następnej reakcji bez oczyszczania.

Surowy ester etylowy kwasu-3-(dibutyloamino)propionowego (9,13 mmol, 2,095 g) rozpuszczono w bezwodnym THF (50 ml), a następnie porcjami wsypano LiAlH₄ (40,0 mmol, 1,518 g). Mieszaninę reakcyjną mieszano w temperaturze pokojowej przez 8 godz. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie ostrożnie dodano do mieszaniny reakcyjnej roztwór woda:dioksan (20 ml, 1:1 v/v). Rozpuszczalniki odparowano za pomocą wyparki obrotowej, uzyskując szary proszek, który następnie zawieszono w DCM (200 ml) i przesączono przez warstwę celitu. Przesącz suszono przy pomocy MgSO₄. Po odsączeniu środka suszącego rozpuszczalnik odparowano za pomocą wyparki obrotowej, a następnie suszono przy użyciu linii próżniowej uzyskując **8** w postaci bladeżółtego oleju (1,186 g, 70% - po 2 etapach).

¹H NMR (400 MHz, CDCl₃) δ 5.62 (s, *J* = 135.8 Hz, 1H), 3.77 (t, 2H), 2.62 (t, 2H), 2.47 – 2.35 (m, 4H), 1.66 (qi, *J* = 5.4 Hz, 2H), 1.44 (qi, 4H), 1.29 (sx, *J* = 7.3 Hz, 4H), 0.90 (t, *J* = 7.3 Hz, 6H).

¹³C NMR (101 MHz, CDCl₃) δ 64.70, 55.25, 53.92, 28.99, 27.82, 20.64, 13.99.

HRMS (EI⁺): *m/z* obliczone dla: C₁₁H₂₅NO: 187.1936, zmierzono: 187.1936.



***N,N*-dibutylo-(3-chloropropyl)amina (9):** W kolbie okrągłodennej umieszczono **8** (2,230 mmol, 0,418 g), rozpuszczono w CHCl₃ (5 ml) i wkroplono SOCl₂ (4,460 mmol, 0,328 ml). Kolbę zaopatrzone w chłodnicę zwrotną, ogrzano do wrzenia za pomocą łaźni olejowej i utrzymywano we wrzeniu przez 7 godzin. Po tym czasie ostudzono mieszaninę reakcyjną i dodano doń DCM i wodny roztwór K₂CO₃ (1.2 g). Rozdzielono warstwy, a fazę organiczną przemyto nasyconym wodnym roztworem NaHCO₃ (5 ml) i następnie wodą (5 ml). Warstwy organiczne suszono za pomocą bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej, uzyskując **9** w postaci żółtego oleju (0,430 g, 94%).

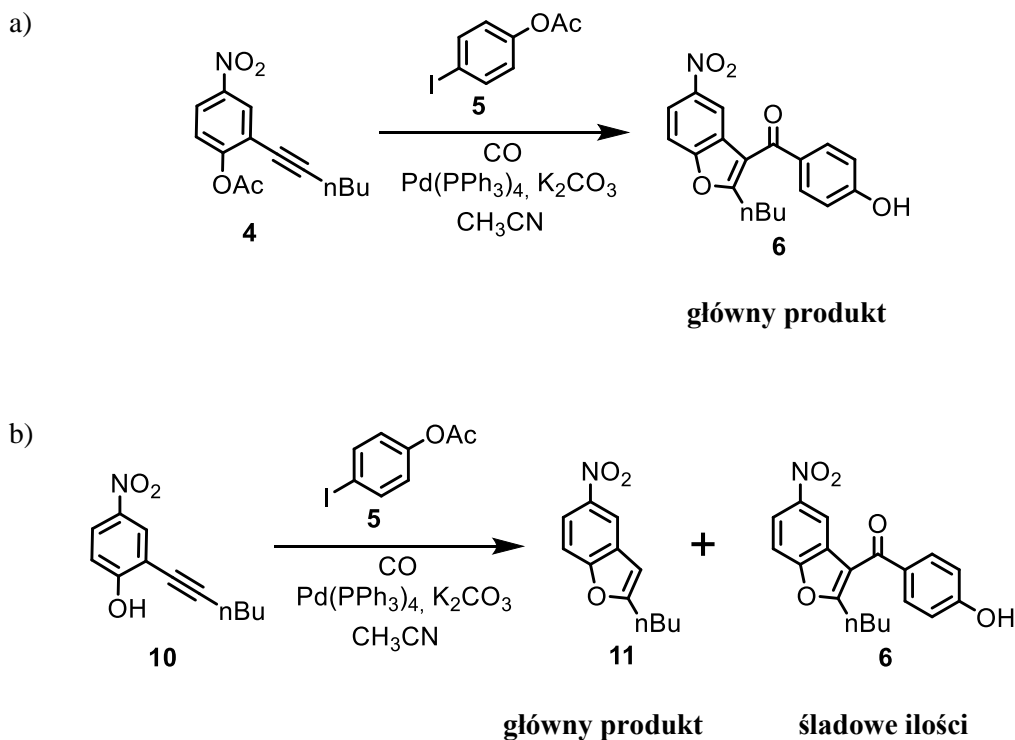
¹H NMR (400 MHz, CDCl₃) δ 3.62 (t, *J* = 6.5 Hz, 1H), 2.55 (t, *J* = 6.8 Hz, 1H), 2.40 (t, 2H), 1.89 (qi, 1H), 1.48 – 1.36 (m, 2H), 1.36 – 1.26 (m, 2H), 0.93 (t, *J* = 7.3 Hz, 3H).

¹³C NMR (101 MHz, CDCl₃) δ 54.02, 51.06, 43.45, 30.61, 29.43, 20.67, 14.05.

HRMS (EI⁺): *m/z* obliczone dla: C₁₁H₂₄NCl: 205.1597, zmierzono: 205.1596.

4.2.2. Optymalizacja warunków karbonylatywnej annulacji (4)

Kluczowym etapem syntezy Dronedaronu (1) jest karbonylatywna annulacja (4). Jak pokazano na Ryc.67b. ochrona grupy fenolowej w (4) jest niezbędna do zajścia reakcji, w przeciwnym razie, dochodzi do spontanicznej, wewnątrzcząsteczkowej cyklizacji prowadzącej do pierścienia benzofuranu Ryc.67a..



Ryc.67. a) Wykorzystanie zabezpieczonego *o*-alkinofenolu (4) prowadzi w karbonylatywnej annulacji, do pożądanego ketonu (6). b) Wykorzystanie niezabezpieczonego *o*-alkinofenolu (10) prowadzi w tych samych warunkach karbonylatywnej annulacji, do niepożądanego benzofuranu (11), pożądanego ketonu (6) obserwowane są wówczas śladowe ilości.

Tabela 1. Wybrane warunki sprawdzone podczas optymalizacji karbonylatywnej annulacji (4).

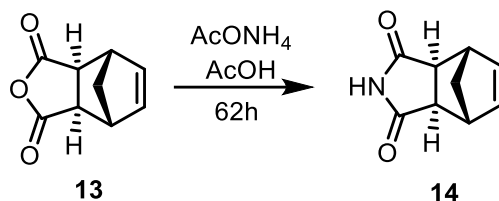
Lp.	Warunki	Wydajność
1	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (3,2 eq), K ₂ CO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), 55°C, 24 godz.	76%
2	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), K ₂ CO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), 55°C, 24 godz.	73%
3	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,0 eq), K ₂ CO ₃ (4,0 eq), Pd(PPh ₃) ₄ (5 mol %), 55°C, 24 godz.	67%
4	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (1,5 eq), K ₂ CO ₃ (4,0 eq), Pd(PPh ₃) ₄ (6 mol %), 55°C, 24 godz.	44%

5	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), K ₂ CO ₃ (3,6 eq), Pd(PPh ₃) ₄ (6 mol %), 55°C, 24 godz.	21%
6	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), K ₂ CO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), 50°C, 24 godz.	21%
7	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), K ₂ CO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), 70°C, 24 godz.	21%
8	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), K ₂ CO ₃ (4,8 eq), Pd ₂ (dba) ₃ (3 mol %), 55°C, 24 godz.	16%
9	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), Na ₂ CO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), 70°C, 48 godz.	12%
10	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), Cs ₂ CO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), 50°C, 24 godz.	9%
11	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), NaHCO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), rt, 24 godz.	9%
12	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), K ₃ PO ₄ (2,4 eq), Pd(PPh ₃) ₄ (6 mol %), rt, 24 godz.	0%
13	octan <i>o</i> -alkinofenolu (4) (1,0 eq), octan 4-jodofenolu (5) (2,4 eq), Li ₂ CO ₃ (4,8 eq), Pd(PPh ₃) ₄ (6 mol %), 70°C, 96 godz.	0%
14	<i>o</i> -alkinofenol (10) (1,0 eq), octan 4-jodofenolu (5) (2,0 eq), K ₂ CO ₃ (4,0 eq), Pd(PPh ₃) ₄ (5 mol %), 55°C, 24 godz.	4.5%*

* Związek (**11**) jest w tych warunkach głównym produktem.

4.3. Synteza 5β/6β-hydroksylurazydonu (12,12')

4.3.1. Procedury syntetyczne zastosowane do uzyskania 5β/6β-hydroksylurazydonu (12,12')

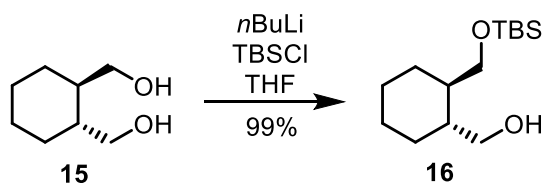


(3aR,4R,7S,7aS)-3a,4,7,7a-tetrahydro-1H-4,7-metanoizindolo-1,3(2H)-dione (14): W kolbie okrągłodennej umieszczono bezwodnik cis-5-norborneno-exo-2,3-dikarboksylowy **13** (29,00 mmol, 4,76 g), lodowaty kwas octowy (75 ml) i AcONH₄ (87,10 mmol, 6,71 g). Kolbę zaopatrzone w chłodnicę zwrotną i ogrzewano do wrzenia za pomocą czaszy grzejnej, prowadząc reakcję przez 62 godziny. Po zakończeniu reakcji rozpuszczalnik odparowano za pomocą wyparki obrotowej. Pozostały olej rozpuszczono w DCM (100 ml), odmyto nasyconym wodnym roztworem NaHCO₃, a następnie suszono za pomocą MgSO₄. Odsączono środek suszący, rozpuszczalnik odparowano za pomocą wyparki obrotowej. Następnie suszono przy użyciu linii próżniowej, uzyskując **14** w postaci białego ciała stałego (4,74 g, 100%).

¹H NMR: (400 MHz, CDCl₃): δ 8.84 (s, 1H), 6.29 (t, *J* = 1.7 Hz, 2H), 3.48 – 3.12 (m, 2H), 2.74 (d, *J* = 1.1 Hz, 2H), 1.63 – 1.54 (m, 1H), 1.47 (d, *J* = 9.9 Hz, 1H).

¹³C NMR: (101 MHz, CDCl₃) δ 178.52, 137.75, 49.19, 45.14, 42.90.

HRMS: (*m/z*): obliczone dla: C₉H₉NO₂ [M⁺•], 163.0633; zmierzono: 163.0629



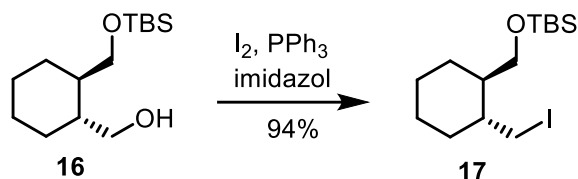
((1R,2R)-2-(((*tert*-butyldimetylosilil)oksy)metyleno)cycloheksylo)metanol (16): W kolbie okrągłodennej umieszczono diol **15** (37,17 mmol, 5,36 g) i rozpuszczono w bezwodnym THF (150 ml). Kolbę zamknięto gumowym septum, wymieniano atmosferę na argon i zamocowano balon z argonem. Kolbę schłodzono na łaźni lodowej do 0°C, po czym wkroplono do mieszaniny, przez strzykawkę roztwór *n*BuLi w cykloheksanie (3,64 mL, 2M). Przez 15 minut utrzymywano temperature 0°C, a następnie pozwolono się ogrzać mieszaninie reakcyjnej i mieszano przez 3 godziny. Po tym czasie zdjęto septum i wsypało do roztworu TBSCl (37,17 mmol, 5,60 g). Mieszano jeszcze przez 1 godzinę. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej nasycony wodny roztwór NaHCO₃ (80 ml) i AcOEt (55 ml). Przeniesiono do rozdzielacza i oddzielono warstwę wodną, ekstrahując ją jeszcze trzykrotnie za pomocą DCM (3x80 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym,

w układzie heksan:AcOEt (5:4), uzyskując **16** w postaci klarownego, bezbarwnego oleju (9,47 g, 99%).

¹H NMR: (400 MHz, CDCl₃) δ 3.61 (s, 1H), 3.57 (d, *J* = 5.0 Hz, 2H), 3.52 – 3.45 (m, 1H), 1.79 – 1.69 (m, 2H), 1.70 - 1.62 (m, 2H), 1.38 – 0.97 (m, 6H), 0.92 (s, 9H), 0.10 (t, *J* = 1.3 Hz, 6H).

¹³C NMR: (101 MHz, CDCl₃) δ 68.59, 67.43, 45.52, 44.07, 30.07, 29.86, 26.16, 26.15, 25.81, 18.17, -5.45, -5.57.

HRMS: (*m/z*): obliczone dla: C₁₄H₃₀O₂Si, [M+Na]⁺, 281.1913; zmierzono: 281.1907



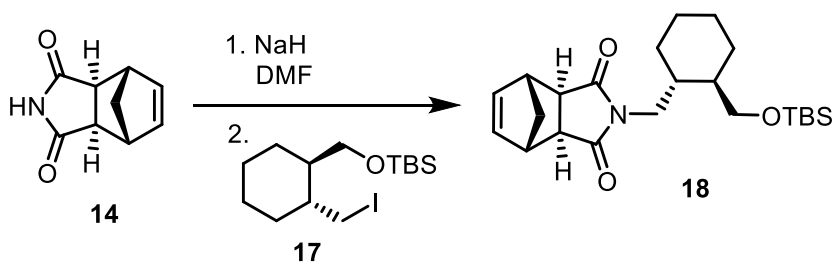
tert-butyl(((1R,2R)-2-(jodometylo)cykloheksylo)metoksy)dimetylosilan (17): W kolbie okrągłodennej umieszczono alkohol **16** (1,63 mmol, 0,42 g) i rozpuszczono w DCM (15 ml). Następnie dodano PPh₃ (3,27 mmol, 0,875 g) i imidazol (4,08 mmol, 0,28 g), mieszano w temperaturze pokojowej przez 20 minut. Po tym czasie dodano porcjami jod (3,27 mmol, 0,83 g). Mieszano przez 1 godzinę. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny nasycony wodny roztwór Na₂S₂O₃ (20 ml) i DCM (40 ml). Przeniesiono do rozdzielacza i oddzielono warstwę wodną, ekstrahując ją jeszcze trzykrotnie za pomocą DCM (3x20 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (10:1), uzyskując **17** w postaci klarownego, bezbarwnego oleju (0,57 g, 94%).

¹H NMR: (400 MHz, CDCl₃) δ 3.62 (dd, *J* = 10.2, 4.2 Hz, 1H), 3.52 (dd, *J* = 10.2, 2.7 Hz, 1H), 3.43 (dd, *J* = 9.8, 2.6 Hz, 1H), 3.32 (dd, *J* = 9.8, 5.7 Hz, 1H), 1.80 – 1.66 (m, 4H), 1.37 – 1.24 (m, 4H), 1.23 – 1.06 (m, 2H), 0.92 (s, 9H), 0.06 (s, 6H).

¹³C NMR: (101 MHz, CDCl₃) δ 65.18, 43.85, 39.20, 33.21, 29.37, 26.06, 25.92, 25.71, 18.27, 16.74, -5.45, -5.47.

HRMS: w warunkach jonizacji, obserwowano rozkład związku **17**.

Analiza elementarna: (%C; %H; %I): obliczone dla: C₁₄H₂₉IOSi: 45.65%C, 7.94%H, 34.45%I; zmierzono: 45.57%C, 7.92%H, 34.25%I



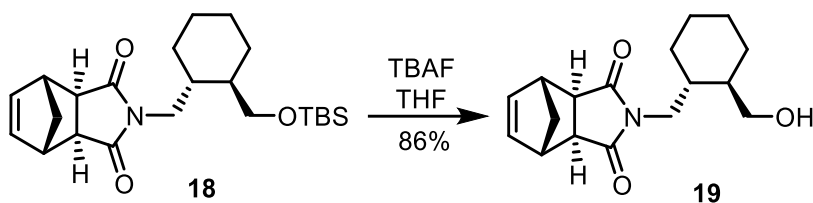
(3aR,4R,7S,7aS)-2-(((1R,2R)-2-((tert-butylodimetylosilil)oksy)metyleno)cycloheksylo)

metyleno)-3a,4,7,7a-tetrahydro-1H-4,7-metanoizindolo-1,3(2H)-dion (18): W kolbie okrągłodennej umieszczono NaH 95% (18,39 mmol, 0,44 g) i zawieszono w DMF (40 ml) następnie dodano imid **14** (16,72 mmol, 2,73 g). Kolbę zamknięto gumowym septum, wymieniono atmosferę na argon i zamocowano balon z argonem. Mieszano w temperaturze pokojowej przez 30 minut. Po czym dodano do mieszaniny reakcyjnej, przez strzykawkę, roztwór jodku **17** (16,72 mmol, 6,16 g) w DMF (40 ml). Mieszano dalej w temperaturze pokojowej przez 3 dni. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Następnie dodano do mieszaniny reakcyjnej nasycony wodny roztwór NH₄Cl (80 ml). Przeniesiono do rozdzielacza i oddzielono warstwę wodną, ekstrahując ją jeszcze trzykrotnie za pomocą AcOEt (3x100 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (5:1), uzyskując **18** w postaci białego ciała stałego (6,29 g, 93%).

¹H NMR: (400 MHz, CDCl₃) δ 6.29 (s, 2H), 3.79 – 3.64 (m, 2H), 3.57 (dd, *J* = 10.1, 5.7 Hz, 1H), 3.33 (dd, *J* = 11.9, 8.4 Hz, 1H), 3.29 (s, 2H), 2.68 (s, 2H), 1.85 – 1.73 (m, 1H), 1.72 – 1.61 (m, 3H), 1.57 – 1.48 (m, 2H), 1.33 – 0.98 (m, 6H), 0.91 (s, 9H), 0.06 (d, *J* = 2.0 Hz, 6H).

¹³C NMR: (101 MHz, CDCl₃) δ 178.34, 178.29, 137.83, 137.78, 66.00, 47.81, 47.76, 45.12, 43.04, 42.93, 42.57, 37.80, 29.64, 29.04, 25.93, 25.37, 25.16, 18.26, -5.40, -5.47.

HRMS: (*m/z*): obliczone dla: C₂₃H₃₇O₃NSi, [M+Na]⁺, 426.2440; zmierzono: 426.2428



(3aR,4R,7S,7aS)-2-(((1R,2R)-2-(hydroksymetyleno)cycloheksylo)

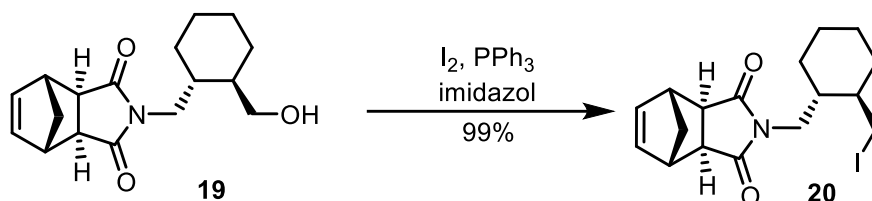
metyleno)-3a,4,7,7a-tetrahydro-1H-4,7-metanoizindolo-1,3(2H)-dion (19): W kolbie okrągłodennej umieszczono **18** (15,22 mmol, 5,87 g) rozpuszczony w THF (70 ml). Kolbę zamknięto gumowym septum, wymieniano atmosferę na argon i zamocowano balon z argonem. Do mieszącego się roztworu dodano roztwór TBAF w THF (1M, 18,9 ml). Mieszano w temperaturze pokojowej przez noc. Następnie w oparciu o TLC stwierdzono zakończenie reakcji, toteż dodano do mieszaniny reakcyjnej nasycony wodny roztwór NH₄Cl (90 ml). Przeniesiono do rozdzielacza i oddzielono warstwę wodną, ekstrahując ją jeszcze pięciokrotnie za pomocą AcOEt (5x100 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki

odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (1:1), uzyskując **19** w postaci klarownego, bezbarwnego oleju (3,65 g, 87%).

¹H NMR: (400 MHz, CDCl₃) δ 6.31 (s, 2H), 3.96 – 3.80 (m, 1H), 3.71 – 3.55 (m, 2H), 3.39 (dd, 1H), 3.29 (s, 2H), 2.71 (s, 2H), 2.45 – 2.25 (s, 1H), 1.80 – 1.62 (m, 4H), 1.56 (m, 2H), 1.37 – 1.20 (m, 4H), 1.08 (m, 2H).

¹³C NMR: (101 MHz, CDCl₃) δ 178.66, 178.56, 137.84, 137.80, 65.41, 47.85, 47.78, 45.19, 45.15, 42.79, 42.36, 41.96, 37.92, 30.84, 29.48, 25.64, 25.45.

HRMS: (*m/z*): obliczone dla: C₁₇H₂₃O₃N, [M+Na]⁺, 312.1576; zmierzono: 312.1570

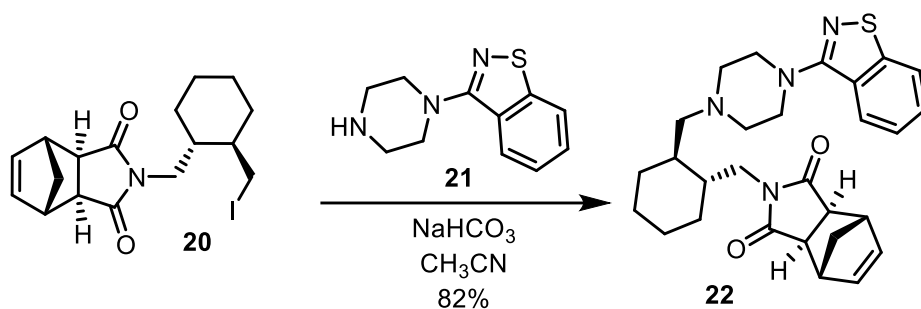


(3aR,4R,7S,7aS)-2-(((1R,2R)-2-(jodometyleno)cykloheksylo) metyleno)-3a,4,7,7a-tetrahydro-1H-4,7-metanoizindolo-1,3(2H)-dion (20): W kolbie okrągłodennej umieszczono **19** (12,06 mmol, 3,49 g) rozpuszczony w DCM (140 ml). Do roztworu dodano PPh₃ (24,12 mmol, 6,33 g) i imidazol (30,15 mmol, 2,05 g). Kolbę schłodzono do 0°C i mieszano przez 30 minut. Po tym czasie wsypano porcjami jod (26,12 mmol, 6,12 g). Reakcję prowadzono jeszcze przez 1,5 godz. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Dodano do mieszaniny reakcyjnej nasycony wodny roztwór Na₂S₂O₃ (100 ml). Przeniesiono do rozdzielacza i oddzielono warstwę wodną, ekstrahując ją jeszcze czterokrotnie za pomocą DCM (4x100 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (1:1), uzyskując **20** w postaci klarownego, bezbarwnego oleju (4,77 g, 99%).

¹H NMR: (400 MHz, CDCl₃) δ 6.31 (s, 2H), 3.61 (dd, 1H), 3.48 – 3.39 (m, 2H), 3.38 – 3.31 (m, 1H), 3.30 (s, 2H), 2.72 (s, 2H), 1.80 (d, 1H), 1.75 – 1.59 (m, 4H), 1.55 (d, 1H), 1.37 – 1.17 (m, 4H), 1.13 – 0.95 (m, 2H).

¹³C NMR: (101 MHz, CDCl₃) δ 178.38, 178.25, 137.80, 137.78, 47.84, 47.80, 45.15, 45.13, 43.03, 42.05, 41.24, 40.25, 32.78, 29.63, 25.28, 25.06, 15.03.

HRMS: (*m/z*): obliczone dla: C₁₇H₂₂O₂NI, [M+Na]⁺, 422.0593; zmierzono: 422.0582



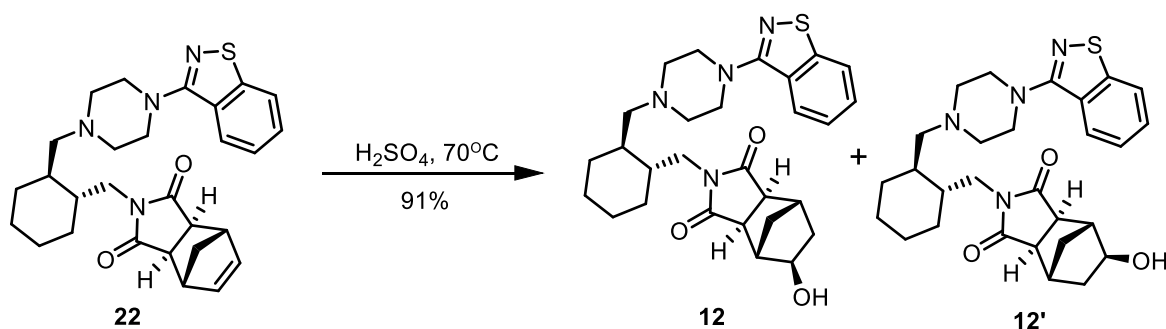
(3aR,4R,7S,7aS)-2-(((1R,2R)-2-((4-(benzo[d]izotiazol-3-yl)piperazyn-1-ylo)metyleno)cykloheksylo)metyleno)-3a,4,7,7a-tetrahydro-1H-4,7-metanoizoindolo-1,3(2H)-dion (22):

W kolbie okrągłodennej umieszczono **20** (1,27 mmol, 0,509 g) rozpuszczony w CH₃CN (7 ml). Do roztworu dodano **21** (2,16 mmol, 0,475 g) i NaHCO₃ (1,66 mmol, 0,139 g). Kolbę zamknięto gumowym septum, wymieniano atmosferę na argon i zamocowano balon z argonem. Mieszano w temperaturze pokojowej przez 22 godziny. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Dodano do mieszaniny reakcyjnej wodę (20 ml). Przeniesiono do rozdzielacza i oddzielono warstwę wodną, ekstrahując ją jeszcze czterokrotnie za pomocą AcOEt (4x20 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (2:1), po wysuszeniu uzyskując **22** w postaci kruchej białej piany (0,514 g, 82%).

¹H NMR: (400 MHz, CDCl₃) δ 7.92 (d, 1H), 7.81 (d, 1H), 7.47 (t, 1H), 7.36 (t, 1H), 6.30 (s, 2H), 3.97 (dd, 1H), 3.54 (t, 4H), 3.34 (dd, 1H), 3.29 (s, 2H), 2.69 (s, 2H), 2.68 – 2.58 (m, 5H), 2.25 (dd, *J* = 12.5, 1H), 1.90 (d, 1H), 1.68 (d, 2H), 1.56 (dd, 3H), 1.40 (dd, 1H), 1.27 (t, 2H), 1.15 (d, 1H), 1.08 – 0.97 (m, 2H).

¹³C NMR: (101 MHz, CDCl₃) δ 178.40, 164.07, 152.74, 137.84, 137.78, 128.11, 127.44, 123.96, 123.80, 120.51, 63.66, 53.50, 50.18, 47.83, 47.79, 45.13, 42.92, 42.66, 40.74, 37.56, 30.77, 29.91, 25.42, 25.02.

HRMS: (*m/z*): obliczone dla: C₂₈H₃₄N₄O₂S, [M+H]⁺, 491.2481; zmierzono: 491.2472



(3aR,4S,5R,7S,7aS)-2-(((1R,2R)-2-((4-(benzo[d]izotiazol-3-yl)piperazyn-1-ylo)metyleno)cykloheksylo)metyleno)-5-hydroksyheksahydro-1H-4,7-metanoizoindolo-1,3(2H)-dion (12,12'):

W kolbie okrągłodennej umieszczono **22** (1,02 mmol, 0,502 g) rozpuszczono w 33% roztworze H₂SO₄ (6 ml). Zamocowano chłodnicę zwrotną i ogrzano za pomocą łaźni olejowej do 70 °C. Kontynuowano ogrzewanie przez noc. Nazajutrz zdjęto łaźnię grzejącą i pozwolono

mieszaniu reakcyjnej ostygnąć do temperatury pokojowej. Mieszano jeszcze przez 1,5 godziny. Po tym czasie w oparciu o TLC stwierdzono zakończenie reakcji. Do mieszaniny reakcyjnej dodano ostrożnie wodny roztwór K_2CO_3 (200 ml), AcOEt (100 ml) i wodę (50 ml). Przeniesiono do rozdzielacza i oddzielono warstwę wodną, ekstrahując ją jeszcze trzykrotnie za pomocą AcOEt (3x100 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego $MgSO_4$, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, jako eluentu używając wyłącznie AcOEt, po wysuszeniu uzyskując **12,12'** w postaci kruchej białej piany (0,475 g, 91%).

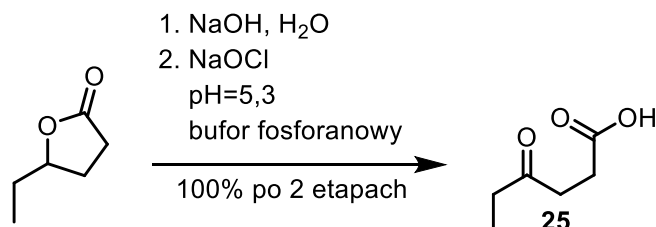
1H NMR: (400 MHz, $CDCl_3$) δ 7.92 (d, $J = 8.2$ Hz, 1H), 7.82 (d, $J = 8.1$ Hz, 1H), 7.47 (t, $J = 7.5$ Hz, 1H), 7.36 (t, $J = 7.6$ Hz, 1H), 3.97 (m, 2H), 3.54 (m, 4H), 3.34 (dd, $J = 13.1, 10.3$ Hz, 1H), 2.75 (d, $J = 4.0$ Hz, 1H), 2.64 (m, 6H), 2.52 (s, 2H), 2.24 (dd, $J = 12.5, 6.5$ Hz, 1H), 1.99 – 1.77 (m, 3H), 1.75 - 1.63 (m, 3H), 1.60 - 1.45 (m, 3H), 1.45 - 1.35 (m, 1H), 1.32 - 0.95 (m, 6H).

^{13}C NMR: (101 MHz, $CDCl_3$) δ 178.68, 178.51, 164.09, 152.72, 128.10, 127.48, 123.97, 120.53, 77.23, 72.97 (d), 63.67, 53.49, 50.18, 47.91 (d, $J = 6.1$ Hz), 47.45 (d, $J = 2.9$ Hz), 44.66 (d, $J = 4.8$ Hz), 42.89 (d, $J = 2.8$ Hz), 40.67 (d), 40.54 (d, $J = 5.3$ Hz), 38.68 (d, $J = 2.8$ Hz), 37.57 (s), 30.76 (s), 29.84 (s), 29.54 (s), 25.40 (s), 25.00 (s).

HRMS: (m/z): obliczone dla: $C_{28}H_{36}N_4O_3S$, $[M+H]^+$, 509.2586; zmierzono: 509.2583

4.4. Synteza hydrazonu (**23(R,R)**) w syntezie totalnej Takamonidyny (**24**)

4.4.1. Procedury syntetyczne zastosowane do uzyskania hydrazonu (**23(R,R)**)

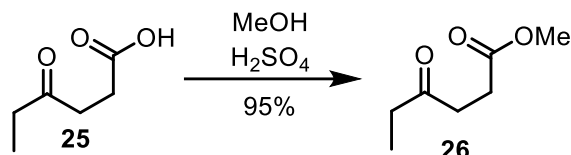


Kwas 4-oksokapronowy (25): W kolbie okrągłodennej umieszczono NaOH (87,61 mmol, 3,80 g) i rozpuszczono w wodzie destylowanej (80 ml). Do roztworu dodano przez strzykawkę γ -heksalaktone (87,61 mmol, 10,00 g), a następnie zamocowano chłodnicę zwrotną, za pomocą łaźni olejowej ogrzano kolbę do $70^\circ C$ i kontynuowano ogrzewanie przez noc (ok. 16 godzin). Następnie odparowano wodę za pomocą wyparki obrotowej, uzyskując białe ciało stałe (13,5 g), które użyto bezpośrednio w następnym etapie. Rozpuszczono je w buforze fosforanowym (730 ml, pH = 5,3; przygotowanie opisane w ogólnych metodach powyżej). Do mieszającego się roztworu dodano 10% wodny roztwór NaClO (134 ml). Reakcję prowadzono w temperaturze pokojowej przez 24 godziny. Po tym czasie mieszaninę reakcyjną podzielono na 2 równe części, każdą z nich potraktowano identycznie, zakwaszając przy użyciu 85% wodnego roztworu H_3PO_4 do osiągnięcia pH = 3,6, następnie natychmiast ekstrahowano przy użyciu AcOEt (4x250 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego $MgSO_4$, następnie odsączono środek suszący, a rozpuszczalniki odparowano za pomocą wyparki obrotowej, uzyskując **25** w postaci bezbarwnych kryształów (11,34 g, 100% - po 2 etapach).

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 9.64 (s, 1H), 2.75 (dd, $J = 9.6, 3.9$ Hz, 2H), 2.66 (dd, $J = 9.8, 4.1$ Hz, 2H), 2.50 (q, $J = 7.3$ Hz, 2H), 1.10 (t, $J = 7.4$ Hz, 3H).

$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 209.34, 178.48, 36.34, 35.84, 27.77, 7.74.

HRMS: (m/z): obliczone dla: $\text{C}_6\text{H}_{10}\text{O}_3$, $[\text{M}+\text{H}]^+$, 130.0635; zmierzono: 130.0630.

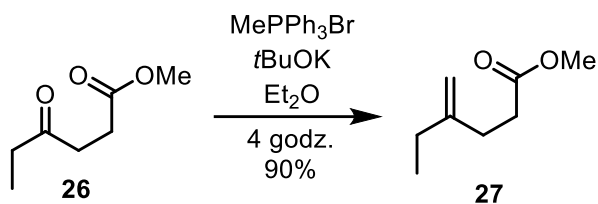


Ester metylowy kwasu 4-oksokapronowego (26): W kolbie okrągłodennej umieszczono **25** (91,97 mmol, 11,97 g), utarte sita molekularne 3\AA i suchy (!) MeOH (62 ml). Do tak przygotowanego roztworu, w trakcie intensywnego mieszania, dodano 98% H_2SO_4 (18,39 mmol, 0,98 ml). Następnie zamocowano chłodnicę zwrotną, zabezpieczono ją gumowym septum, a następnie wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Kolbę umieszczono na łaźni olejowej i ogrzano do 65°C , reakcję prowadzono przez noc (ok. 16 godzin). Po tym czasie w oparciu o TLC (wywołane jodem) stwierdzono zakończenie reakcji. Kolbę wyjęto z łaźni olejowej, ochłodzono do temperatury pokojowej, odsączono sita molekularne, a przesącz przeniesiono do rozdzielacza. Dodano nasycony roztwór NaHCO_3 (100 ml) i ekstrahowano czterokrotnie przy pomocy Et_2O (4×100 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO_4 , następnie odsączono środek suszący, a rozpuszczalnik odparowano ostrożnie za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie niskowrzących rozpuszczalników pentan: Et_2O (3:1), które po zebraniu frakcji zawierających **26** ostrożnie odparowano przy pomocy wyparki obrotowej (500 mbar, 22°C) uzyskując **26** w postaci jasnożółtego, klarownego oleju (12,62 g, 95%).

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 3.70 (s, 3H), 2.75 (t, $J = 6.6$ Hz, 2H), 2.62 (t, $J = 6.5$ Hz, 2H), 2.50 (q, $J = 7.3$ Hz, 2H), 1.10 (t, $J = 7.4$ Hz, 3H).

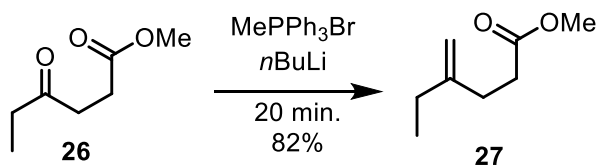
$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 209.33, 173.24, 51.67, 36.56, 35.82, 27.72, 7.69.

HRMS: (m/z): obliczone dla: $\text{C}_7\text{H}_{12}\text{O}_3\text{Na}$, $[\text{M}+\text{Na}]^+$, 167.0684; zmierzono: 167.0678.



Procedura z *t*BuOK

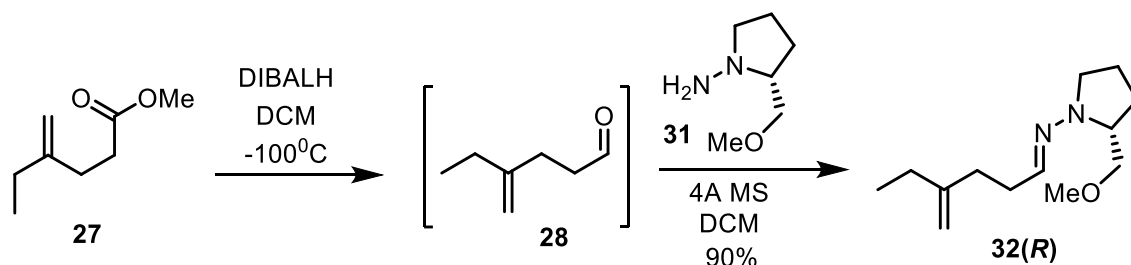
Ester metylowy kwasu 4-metylenokapronowego (27): W naczyniu Schlenka, uprzednio wyprażonym w piecu (240°C, >16 godz.), umieszczono *t*BuOK (12,95 mmol, 1,45 g). Następnie zamocowano gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Podano przez septę suchy Et₂O (10 ml) i mieszano 15-20 minut w temperaturze pokojowej. W międzyczasie w drugim naczyniu Schlenka, uprzednio wyprażonym w piecu (240°C, >16 godz.), umieszczono Ph₃PMeBr (12,95 mmol, 4,62 g). Następnie zamocowano gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Podano przez septę suchy Et₂O (20 ml). Naczynia Schlenka połączono cannulą i używając nadciśnienia argonu przeniesiono zawieszynę *t*BuOK do zawiesziny Ph₃PMeBr. Roztwór Ph₃PMeBr natychmiast przybrał jasnożółtą barwę. Tak przygotowany roztwór mieszano w temperaturze pokojowej przez 2 godziny. Po tym czasie podano przez septę roztwór ketonu **26** (6,33 mmol, 0,912 g) w suchym Et₂O (2 ml). Reakcję prowadzono jeszcze przez 4 godziny, ogrzewając naczynie reakcyjne w łaźni olejowej do 35°C. Po tym czasie w oparciu o TLC (wywołane jodem) stwierdzono zakończenie reakcji. Roztwór przeniesiono do rozdzielacza zawierającego nasycony wodny roztwór NH₄Cl (80 ml). Oddzielono warstwy i fazę wodną ekstrahowano czterokrotnie przy pomocy Et₂O (4x70 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalnik odparowano powoli za pomocą wyparki obrotowej (500 mbar, 22 °C). Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie niskowrzących rozpuszczalników pentan:Et₂O (10:1→4:1), które po zebraniu frakcji zawierających **27** ostrożnie zatężono przy pomocy wyparki obrotowej (500 mbar, 22 °C) uzyskując 1,047 g roztworu **27** w Et₂O/*n*-pentanie. W oparciu o ¹H NMR obliczono zawartość **27** w roztworze (0,810 g, 90%). Taki roztwór był używany w następnym etapie bez dalszego oczyszczania.



Procedura z *n*BuLi

Ester metylowy kwasu 4-metylenokapronowego (27): W naczyniu Schlenka, uprzednio wyprażonym w piecu (240°C, >16 godz.), umieszczono Ph₃PMeBr (19,32 mmol, 6,90 g). Następnie zamocowano gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Podano przez septę suchy Et₂O (45 ml) i mieszano w temperaturze pokojowej przez 20 minut. Po tym czasie powoli podano przez septę *n*BuLi (1,6 M roztwór w heksanie, 19,32 mmol, 12,07 ml). Roztwór pierwotnie przybrał barwę jasnożółtą, a następnie pomarańczową. Tak

przygotowany roztwór mieszano w temperaturze pokojowej przez 1 godzinę. Po tym czasie podano przez septę roztwór ketonu **26** (12,70 mmol, 1,83 g) w suchym Et₂O (2 ml). Tuż po podaniu ketonu obserwowano formowanie się osadu, który w przypadku zbyt stężonego roztworu może prowadzić do całkowitego zestalenia się mieszaniny reakcyjnej. W trakcie tego procesu obserwowano zmianę barwy roztworu z pomarańczowego do ciemnobrązowego i ostatecznie do jasnobieżowego. Reakcję prowadzono przez 20 minut w temperaturze pokojowej. Po tym czasie w oparciu o TLC (wywołane jodem) stwierdzono całkowity zanik **26**. Roztwór przeniesiono do rozdzielacza zawierającego wodę (25 ml) i nasycony wodny roztwór NH₄Cl (25 ml). Oddzielono warstwy i fazę wodną ekstrahowano trzykrotnie przy pomocy Et₂O (3x50 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, a rozpuszczalnik odparowano powoli za pomocą wyparki obrotowej (500 mbar, 22 °C). Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie niskowrzących rozpuszczalników pentan:Et₂O (10:1→4:1), które po zebraniu frakcji zawierających **27** ostrożnie zatężono przy pomocy wyparki obrotowej (500 mbar, 22 °C) uzyskując 3,19 g roztworu **27** w Et₂O/*n*-pentanie. W oparciu o ¹H NMR obliczono zawartość **27** w roztworze (1,48 g, 82%). Taki roztwór był używany w następnym etapie bez dalszego oczyszczania.



(R)-N-(2-(metoksymetylo)pirrolidyn-1-ylo)-4-metylenoheksano-1-hydrazon (32(R)):

Redukcja estru do aldehydu (28)

W kolbie kulistej zawierającej roztwór **27** w Et₂O/*n*-pentanie. (9,928 mmol, 1,41 g) i DCM (22 ml), zamocowano gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Następnie schłodzono do -100°C za pomocą łaźni aceton/ciekły azot. Podawano ze strzykawki przez septę roztwór DIBALH (1 M w DCM, 9,93 mmol, 9,90 ml) przez 6 minut. Tak sporządzony roztwór mieszano w -100°C przez 30 minut. Po tym czasie zazwyczaj następowało zestalenie roztworu. Wyjmowano go z łaźni chłodzącej i ogrzewano ostrożnie suszarką do ponownego rozpuszczenia. Po odzyskaniu ciekłej konsystencji umieszczano kolbę w łaźni -100°C na kolejne 30 minut. Po tym czasie w oparciu o TLC (wywołane jodem) stwierdzono niemal całkowity zanik **27**. Reakcję skończono dodając przez septę nasycony wodny roztwór NH₄Cl (13 ml) i wyjmując kolbę z łaźni chłodzącej. Mieszano przez 40 minut pozwalając ogrzać się do temperatury pokojowej (wcześniejsze przystąpienie do ekstrakcji skutkuje powstaniem nierozdzielalnej, gęstej emulsji). Następnie dodano kolejną porcję nasyconego wodnego roztworu NH₄Cl (13 ml), przeniesiono roztwór do rozdzielacza i ekstrahowano CHCl₃ (4x25 ml) bardzo delikatnie wytrząsając – w przeciwnym wypadku powstaje gęsta, nierozdzielalna emulsja. Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO₄, następnie odsączono środek suszący, prowadząc sączenie roztworu aldehydu **28**, bezpośrednio do kolby okrągłodennej, w której prowadzony będzie następny etap.

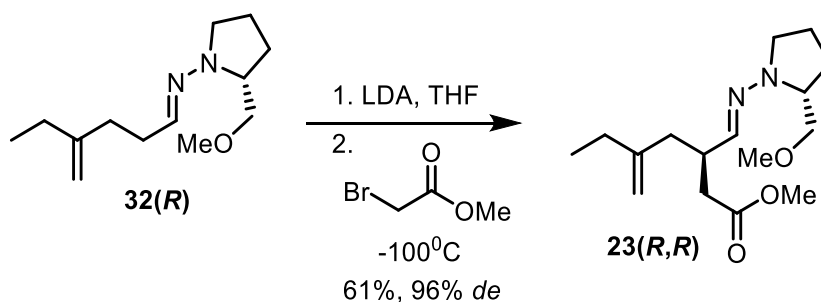
Tworzenie hydrazonu (32(R)) z aldehydu (28)

Do kolby kulistej zawierającej roztwór **28** w DCM/CHCl₃, uzyskany w poprzednim etapie, wsypano utarte sita molekularne 3Å, zamocowano gumową septę i balon z argonem i do mieszającego się roztworu ostrożnie wkroplono przez septę RAMP **31** (9,93 mmol, 1,33 ml). Mieszano w temperaturze pokojowej przez 5,5 godziny. Po tym czasie w oparciu o TLC (UV, λ = 254 nm + pary jodu), stwierdzono zakończenie reakcji. Odsączono sita molekularne, a rozpuszczalniki odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (4:1), po zebraniu i połączeniu frakcji zawierających **32(R)**, odparowano przy pomocy wyparki obrotowej uzyskując **32(R)** w postaci bezbarwnego, klarownego oleju (1,954 g, 88% - po 2 etapach).

¹H NMR (400 MHz, CDCl₃) δ 6.66 (t, *J* = 5.4 Hz, 1H), 4.76 (s, 2H), 3.59 (dd, *J* = 8.3, 3.0 Hz, 1H), 3.52 – 3.26 (m, 6H), 2.74 (q, *J* = 8.4 Hz, 1H), 2.39 (m, 2H), 2.27 – 2.19 (m, 2H), 2.07 (q, *J* = 7.4 Hz, 2H), 2.01 – 1.85 (m, 3H), 1.80 (d, 1H), 1.05 (t, *J* = 7.4 Hz, 3H).

¹³C NMR (101 MHz, CDCl₃) δ 150.61, 138.54, 108.01, 104.30, 74.88, 63.47, 59.19, 50.46, 34.20, 31.41, 28.84, 26.63, 22.15, 12.35.

HRMS: (*m/z*): obliczone dla: C₁₃H₂₅N₂O [M+H]⁺, 225.1967; zmierzono: 225.1960.



Ester metylowy kwasu (R)-3-(((R)-2-(metoksymetylo)pirrolidyn-1-ylo)imino)metylo-5-metylenoheptanowego (23(R,R)): Naczynie Schlenka, uprzednio wyprażone w piecu (240°C, >16 godz.), zaopatrzone w gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Podano przez septę suchy (!) THF (14 ml) i suchą *i*Pr₂NH (5,75 mmol, 0,806 ml). Następnie za pomocą łaźni aceton/suchy lód schłodzono naczynie reakcyjne do -70 °C. Po czym podano przez septę *n*BuLi (1,6 M roztwór w heksanie, 5,31 mmol, 3,32 ml). Mieszano dalej w temperaturze -70°C przez 5 minut i zmieniono na łaźnię lód/wodny roztwór NH₄Cl o temperaturze -10°C, w której chłodzono jeszcze przez 55 minut. Po tym czasie powoli dodano przez septę roztwór **32(R)** (4,426 mmol, 0,993 g) w suchym THF (7 ml). Reakcję prowadzono przez 6 godzin w temperaturze -10°C. Po tym czasie zmieniono łaźnię na aceton/ciekły azot i schłodzono do -100°C. Po czym wkroplono powoli przez septę bromooctan metylu (13,28 mmol, 1,26 ml). Mieszanka reakcyjna przybrała barwę pomarańczową. Reakcję prowadzono dalej przez noc (>16 godz.), pozwalając się ogrzać do temperatury pokojowej. Nazajutrz TLC (UV, λ = 254 nm) wykazało pełną konwersję **32(R)**. Rozpuszczalnik odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (8:1→4:1), uzyskując **23(R,R)** w postaci bladopomarańczowego, klarownego oleju (0,806 g, 61%, 96,43% *de*).

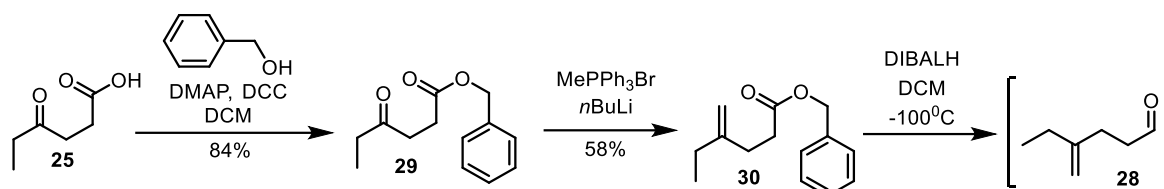
¹H NMR (400 MHz, CDCl₃) δ 6.57 (d, *J* = 4.7 Hz, 1H), 4.82 (d, *J* = 1.5 Hz, 1H), 4.77 (s, 1H), 3.67 (s, 3H), 3.57 (dd, *J* = 9.0, 3.4 Hz, 1H), 3.45 – 3.26 (m, 6H), 3.03 – 2.91 (m, 1H), 2.71 (dd, *J* = 16.5, 8.2 Hz, 1H), 2.53 (dd, *J* = 15.6, 8.3 Hz, 1H), 2.42 (dd, *J* = 15.6, 5.7 Hz, 1H), 2.27 (dd, *J* = 14.0, 7.3 Hz, 1H), 2.15 (dd, *J* = 14.1, 7.9 Hz, 1H), 2.07 (q, *J* = 7.5 Hz, 2H), 1.99 – 1.75 (m, 4H), 1.05 (t, *J* = 7.4 Hz, 3H).

¹³C NMR (101 MHz, CDCl₃) δ 173.29, 148.50, 138.85, 110.43, 74.56, 63.33, 59.18, 51.30, 49.83, 40.54, 37.19, 36.61, 28.38, 26.56, 21.96, 12.21.

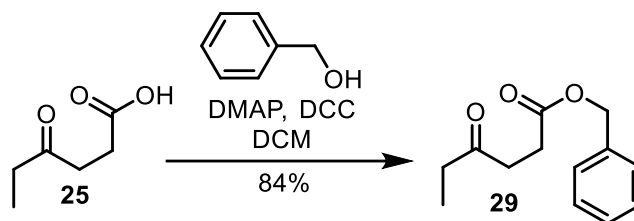
HRMS: (*m/z*): obliczone dla: C₁₆H₂₉N₂O₃ [M+H]⁺, 297.2178; zmierzono: 297.2171.

Alternatywna metoda uzyskania aldehydu (28)

Zważywszy na wysoką lotność związków (26), (27) i (28), a co za tym idzie trudność w ich oczyszczaniu, opracowano metodę uzyskiwania aldehydu (28) z prekursorów mniej lotnych.



Ryc.68. Alternatywna metoda uzyskiwania aldehydu (28).

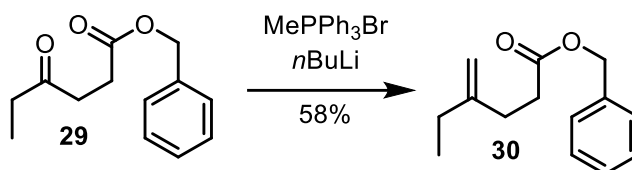


Ester benzylový kwasu 4-oksokapronowego (29): W kolbie okrągłodennej umieszczono **25** (24,65 mmol, 3,21 g), DMAP (4,93 mmol, 0,602 g) i suchy DCM (50 ml). Naczynie zaopatrzone w gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Następnie za pomocą łaźni lodowej schłodzono naczynie reakcyjne do 0°C. Zdjęto septum i dodano do roztworu DCC (29,58 mmol, 6,103 g). Po czym ponownie zaopatrzone naczynie w gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Następnie podano przez septę alkohol benzylový (49,30 mmol, 5,10 ml). Pozwolono reakcji osiągnąć temperaturę pokojową i prowadzono reakcję przez następne 18 godzin. Po tym czasie w roztworze był obecny brązowy osad. Odsączono go przy pomocy lejka Büchnera z warstwą celitu na bibule. Osad przemyto DCM. Przesącz odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (7:1), uzyskując **29** w postaci bladeżółtego, klarownego oleju (4,57 g, 84%).

¹H NMR (400 MHz, CDCl₃) δ 7.45 – 7.30 (m, 5H), 5.14 (s, 2H), 2.76 (dd, *J* = 9.8, 3.9 Hz, 2H), 2.67 (dd, *J* = 9.6, 3.8 Hz, 2H), 2.49 (q, *J* = 7.3 Hz, 2H), 1.09 (t, *J* = 7.3 Hz, 3H).

^{13}C NMR (101 MHz, CDCl_3) δ 209.29, 172.65, 135.88, 128.54, 128.17, 66.46, 36.59, 35.91, 28.05, 7.75.

HRMS: (m/z): obliczone dla: $\text{C}_{13}\text{H}_{16}\text{O}_3$, $[\text{M}+\text{H}]^+$, 220.1099; zmierzono: 220.1092.

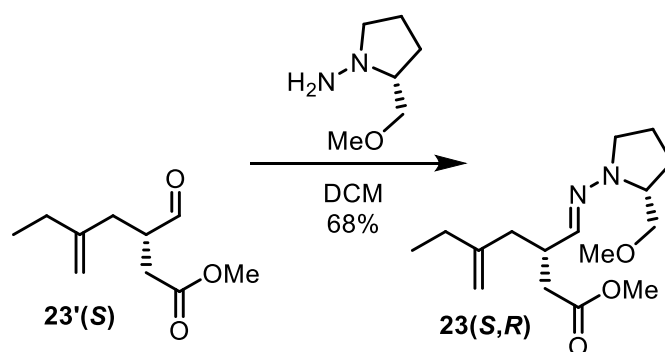


Ester benzyłowy kwasu 4-metylenokapronowego (30): W naczyniu Schlenka, uprzednio wyprazonym w piecu (240°C , >16 godz.), umieszczono Ph_3PMeBr (25,88 mmol, 9,29 g), po czym zaopatrzone je w gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Podano przez septum suchy Et_2O (70 ml), mieszano przez 20 minut w temperaturze pokojowej, po czym dodano $n\text{BuLi}$ (1,6 M w heksanie, 24,89 mmol, 15,55 ml). Skutkiem czego roztwór z bezbarwnego przybrał barwę jasnożółtą, a następnie pomarańczową. Mieszano przez 1 godzinę w temperaturze pokojowej. Po tym czasie wkroplono roztwór **29** (19,91 mmol, 4,38 g) w Et_2O (2 ml). Tuż po podaniu ketonu obserwowano formowanie się osadu, który w przypadku zbyt stężonego roztworu może prowadzić do całkowitego zestalenia się mieszaniny reakcyjnej. W trakcie tego procesu obserwowano zmianę barwy roztworu z pomarańczowego do ciemnobrązowego i ostatecznie do jasnobłękowego. Reakcję prowadzono przez 20 minut w temperaturze pokojowej. Po tym czasie w oparciu o TLC (wywołane jodem) stwierdzono niemal całkowity zanik **29**. Roztwór przeniesiono do rozdzielacza zawierającego wodę (25 ml) i nasycony wodny roztwór NH_4Cl (80 ml). Oddzielono warstwy i fazę wodną ekstrahowano trzykrotnie przy pomocy Et_2O (3×70 ml). Połączone fazy organiczne suszono przy pomocy bezwodnego MgSO_4 , następnie odsączono środek suszący, a rozpuszczalnik odparowano powoli za pomocą wyparki obrotowej (500 mbar, 22°C). Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie niskowrzących rozpuszczalników pentan: Et_2O (10:1 \rightarrow 4:1), które po zebraniu frakcji zawierających **30** ostrożnie zatężono przy pomocy wyparki obrotowej (500 mbar, 22°C) uzyskując roztwór **30** w $\text{Et}_2\text{O}/n$ -pentanie. W oparciu o ^1H NMR obliczono zawartość **30** w roztworze (2,53 g, 58%). Taki roztwór był używany w następnym etapie bez dalszego oczyszczenia, identycznie jak ester metylowy **27** opisany wcześniej, nie obserwowano różnic wydajności pomiędzy estrem metylowym a benzyłowy w sekwencji redukcji przy pomocy DIBALH i tworzeniu hydrazonu.

4.4.2. Określenie nadmiaru diastereoizomerycznego dla **23(R,R)**

Nadmiar enancjomeryczny uzyskiwany w reakcji asymetrycznego alkilowania przy pomocy RAMP, określono za pomocą wysokosprawnej chromatografii cieczowej, z wykorzystaniem chiralnej kolumny.

W tym celu porównano produkt alkilowania **23(R,R)** uzyskany w opisanej powyżej procedurze, z mieszaniną epimerów zawierającą **23(R,R)** i **23(S,R)**. Sposób uzyskania związku **23(S,R)** – zilustrowano poniżej. Próby wykonania alkilowania za pomocą bromooctanu metylu, na optycznie nieczynnych pochodnych aldehydu **28** – nie powiodły się. Bezpośrednie alkilowanie aldehydu **28** nie mogło zostać zrealizowane ze względu na jego wysoką lotność i uzyskiwanie go wyłącznie w roztworze chlorowcowanych rozpuszczalników. Przekształcenie go w odpowiednie hydrazony: dimetylohydrazon i difenylhydrazon, i alkilowanie w takiej postaci, również nie mogło zostać zastosowane ze względu na nietrwałość tych hydrazonów. Aldehyd **28** i jego alkilowana pochodna były stabilne w postaci hydrazonów RAMP i SAMP. Dlatego to nimi się posłużono w celu określenia nadmiaru diastereoizomerycznego.



Ester metylowy kwasu (S)-3-(((R)-2-(metoksymetylo)pirrolidyn-1-ylo)imino)metylo)-5-metylenoheptanowego (23(S,R)**):** Do roztworu aldehydu **23'(S)** (0,409 mmol, 0,0753 g) (uzyskanego za pomocą standardowych metod opisanych w punkcie 3.3.3.2), w suchym DCM (1,5 ml), wsypano utarte sita molekularne 3Å. Po czym zaopatrzone kolbę w gumowe septum i wymieniono atmosferę na argon, mocując na koniec w septum balon z argonem. Przez septum podano RAMP (0,409 mmol, 0,055 ml). Reakcję prowadzono w temperaturze pokojowej przez 16 godzin. Po tym czasie TLC (UV, $\lambda = 254$ nm) wykazało pełną konwersję **23'(S)**. Odsączono sita molekularne, a rozpuszczalnik odparowano za pomocą wyparki obrotowej. Pozostałość oczyszczano za pomocą chromatografii kolumnowej na żelu krzemionkowym, w układzie heksan:AcOEt (10:1→7:1), uzyskując **23(S,R)** w postaci bezbarwnego, klarownego oleju (0,0819 g, 68%).

¹H NMR (400 MHz, CDCl₃) δ 9.77 (s, 1H), 4.91 (s, 1H), 3.70 (s, 3H), 3.07-2.98 (m, 1H), 2.69 (dd, 1H), 2.55 (dd, 1H), 2.46 (dd, 1H), 2.16 (dd, 1H), 2.07 (bq, 2H), 1.08 (t, 3H).

¹³C NMR (101 MHz, CDCl₃) δ 202.55, 172.30, 146.82, 111.35, 51.75, 45.60, 35.31, 32.58, 28.23, 12.08.

5. Bibliografia

1. E. J. Corey, *Pure Appl. Chem.*, **1967**, *14.1*, 19-38.
2. E. J. Corey, W. T. Wipke, R. D. Cramer III, W. J. Howe, *J. Am. Chem. Soc.*, **1972**, *94(2)*, 421-430.
3. E. J. Corey, W. T. Wipke, *Science*, **1969**, *166(3902)*, 178-192.
4. E. J. Corey, W. J. Howe, H. W. Orf, D. A. Pensak, G. Petersson, *J. Am. Chem. Soc.*, **1975**, *97(21)*, 6116-6124.
5. A. K. Long, J. C. Kappos, *J. Chem. Inf. Comput. Sci.*, **1994**, *34(4)*, 915-921.
6. E. J. Corey, R. D. Cramer III, W. J. Howe, *J. Am. Chem. Soc.*, **1972**, *94(2)*, 440-459.
7. G. Büchi, W. Hofheinz, J. V. Paukstelis, *J. Am. Chem. Soc.*, **1966**, *88(17)*, 4113-4114.
8. G. Büchi, W. Hofheinz, J. V. Paukstelis, *J. Am. Chem. Soc.*, **1969**, *91(23)*, 6473-6478.
9. R. C. Kelly, I. Schletter, *J. Am. Chem. Soc.*, **1973**, *95(21)*, 7156-7157.
10. E. J. Corey, W. J. Howe, D. A. Pensak, *J. Am. Chem. Soc.*, **1974**, *96(25)*, 7724-7737.
11. M. H. Todd, *Chem. Soc. Rev.*, **2005**, *34(3)*, 247-266.
12. E. J. Corey, W. L. Jorgensen, *J. Am. Chem. Soc.*, **1976**, *98(1)*, 203-209.
13. E. J. Corey, A. P. Johnson, A. K. Long, *J. Org. Chem.*, **1980**, *45(11)*, 2051-2057.
14. E. J. Corey, A. K. Long, J. Mulzer, H. W. Orf, A. P. Johnson, A. P. Hewett, *J. Chem. Inf. Comput. Sci.*, **1980**, *20(4)*, 221-230.
15. M. A. Ott, J. H. Noordik, *J. Chem. Inf. Comput. Sci.*, **1997**, *37(1)*, 98-108.
16. E. J. Corey, A. K. Long, S. D. Rubenstein, *Science*, **1985**, *228(4698)*, 408-418.
17. P. Mata, A. M. Lobo, C. Marshall, A. P. Johnson, *J. Chem. Inf. Comput. Sci.*, **1994**, *34(3)*, 491-504.
18. A. P. Johnson, C. Marshall, P. N. Judson, *Recl. Trav. Chim. Pays-Bas*, **1992**, *111(6)*, 310-316.
19. E. J. Corey, W. L. Jorgensen, *J. Am. Chem. Soc.*, **1976**, *98(1)*, 189-203.
20. E. J. Corey, A. K. Long, G. I. Lotto, S. D. Rubenstein, *Recl. Trav. Chim. Pays-Bas*, **1992**, *111(6)*, 304-309.
21. [dostęp 11.07.2019]: cheminf.cmbi.ru.nl/cheminf/lhasa/
22. E. L. van Rozendaal, M. A. Ott, H. W. Scheeren, *Recl. Trav. Chim. Pays-Bas*, **1994**, *113(5)*, 297-303.
23. E. J. Corey, H. W. Orf, D. A. Pensak, *J. Am. Chem. Soc.*, **1976**, *98(1)*, 210-221.
24. E. J. Corey, A. K. Long, T. W. Greene, J. W. Miller, *J. Org. Chem.*, **1985**, *50(11)*, 1920-1927.
25. E. J. Corey, R. D. Balanson, *Tetrahedron Lett.*, **1973**, *14(34)*, 3153-3156.
26. P. Judson, *Knowledge-based Expert Systems in Chemistry: Not Counting on Computers*, wyd. 1, RCS Publishing, Cambridge 2009.
27. E. J. Corey, R. D. Cramer III, W. J. Howe, *J. Am. Chem. Soc.*, **1972**, *94(2)*, 440-459.
28. E. J. Corey, W. T. Wipke, R. D. Cramer III, W. J. Howe, *J. Am. Chem. Soc.*, **1972**, *94(2)*, 431-439.
29. E. J. Corey, N. F. Feiner, *J. Org. Chem.*, **1980**, *45(5)*, 757-764.
30. E. J. Corey, N. F. Feiner, *J. Org. Chem.*, **1980**, *45(5)*, 765-780.
31. G. Büchi, W. D. Macleod, *J. Am. Chem. Soc.*, **1962**, *84(16)*, 3205-3206.
32. G. Büchi, W. D. MacLeod, J. Padilla, *J. Am. Chem. Soc.*, **1964**, *86(20)*, 4438-4444.
33. S. Danishefsky, D. Dumas, *Chem. Comm. (London)*, **1968**, *21*, 1287-1288.
34. F. Näf, G. Ohloff, *Helv. Chim. Acta*, **1974**, *57(6)*, 1868-1870.
35. M. Bertrand, P. Teisseire, G. Pelerin, *Tetrahedron Lett.*, **1980**, *21(21)*, 2055-2056.
36. T. V. Magee, G. Stork, P. Fludzinski, *Tetrahedron Lett.*, **1995**, *36(42)*, 7607-7610.
37. R. N. Mirrington, K. J. Schmalzl, *J. Org. Chem.*, **1972**, *37(18)*, 2871-2877.
38. K. Kaliappan, G. S. R. S. Rao, *J. Chem. Soc., Perkin Trans. 1*, **1997**, *(9)*, 1385-1390.
39. G. Q. Xu, G. Q. Lin, B. F. Sun, *Org. Chem. Front.*, **2017**, *4(10)*, 2031-2033.
40. K. Yamada, Y. Kyotani, S. Manabe, M. Suzuki, *Tetrahedron*, **1979**, *35(3)*, 293-298.
41. R. B. Zhao, Y. L. Wu, *Chinese J. Chem.*, **1991**, *9(4)*, 377-380.
42. A. Srikrishna, G. Satyanarayana, *Tetrahedron Asymmetr.*, **2005**, *16(24)*, 3992-3997.
43. E. J. Corey, R. D. Balanson, *J. Am. Chem. Soc.*, **1974**, *96(20)*, 6516-6517.
44. A. Weise, H. G. Scharnow, *Z. Chem.*, **1979**, *19(2)*, 49-56.
45. W. G. Dauben, A. C. Ashcraft, *J. Am. Chem. Soc.*, **1963**, *85(22)*, 3673-3676.
46. E. J. Corey, Z. Arnold, J. Hutton, *Tetrahedron Lett.*, **1970**, *11(4)*, 307-310.

47. E. J. Corey, R. Noyori, *Tetrahedron Lett.*, **1970**, 11(4), 311-313.
48. R. B. Woodward, J. Gosteli, I. Ernest, R. J. Friary, G. Nestler, H. Raman, *J. Am. Chem. Soc.*, **1973**, 95(20), 6853-6855.
49. E. J. Corey, N. M. Weinshenker, T. K. Schaaf, W. Huber, *J. Am. Chem. Soc.*, **1969**, 91(20), 5675-5677.
50. E. J. Corey, T. K. Schaaf, W. Huber, U. Koelliker, N. M. Weinshenker, *J. Am. Chem. Soc.*, **1970**, 92(2), 397-398.
51. E. J. Corey, R. B. Mitra, H. Uda, *J. Am. Chem. Soc.*, **1964**, 86(3), 485-492.
52. J. E. McMurry, *J. Am. Chem. Soc.*, **1968**, 90(24), 6821-6825.
53. J. A. Marshall, G. L. Bundy, W. I. Fanta, *J. Org. Chem.*, **1968**, 33(10), 3913-3922.
54. P. A. Grieco, M. Nishizawa, T. Oguri, S. D. Burke, N. Marinovic, *J. Am. Chem. Soc.*, **1977**, 99(17), 5773-5780.
55. A. F. Barrero, J. E. Oltra, A. Barragán, M. Álvarez, *J. Chem. Soc., Perkin Trans. 1*, **1998**, (24), 4107-4114.
56. A. F. Barrero, J. E. Oltra, A. Barragán, *Tetrahedron Lett.*, **1995**, 36(2), 311-314.
57. A. F. Barrero, J. E. Oltra, M. Álvarez, *Tetrahedron Lett.*, **2000**, 41(40), 7639-7643.
58. H. Iio, M. Isobe, T. Kawai, T. Goto, *Tetrahedron*, **1979**, 35(8), 941-948.
59. P. A. Grieco, M. Nishizawa, S. D. Burke, N. Marinovic, *J. Am. Chem. Soc.*, **1976**, 98(6), 1612-1613.
60. S. Danishefsky, T. Kitahara, P. F. Schuda, S. J. Etheredge, *J. Am. Chem. Soc.*, **1976**, 98(10), 3028-3030.
61. S. Danishefsky, P. F. Schuda, T. Kitahara, S. J. Etheredge, *J. Am. Chem. Soc.*, **1977**, 99(18), 6066-6075.
62. G. R. Kieczkowski, R. H. Schlessinger, *J. Am. Chem. Soc.*, **1978**, 100(6), 1938-1940.
63. M. Isobe, H. Iio, T. Kawai, T. Goto, *J. Am. Chem. Soc.*, **1978**, 100(6), 1940-1942.
64. H. Iio, M. Isobe, T. Kawai, T. Goto, *J. Am. Chem. Soc.*, **1979**, 101(20), 6076-6081.
65. T. Wakamatsu, H. Hara, Y. Ban, *J. Org. Chem.*, **1985**, 50(1), 108-112.
66. S. Danishefsky, P. Schuda, K. Kato, *J. Org. Chem.*, **1976**, 41(7), 1081-1088.
67. R. D. Clark, C. H. Heathcock, *J. Org. Chem.*, **1976**, 41(8), 1396-1403.
68. P. A. Grieco, K. Hiroi, J. J. Reap, J. A. Noguez, *J. Org. Chem.*, **1975**, 40(10), 1450-1453.
69. K. Kondo, M. Sodeoka, M. Mori, M. Shibasaki, *Synthesis*, **1993**, 9, 920-930.
70. F. Zutterman, H. De Wilde, R. Mijngheer, P. De Clercq, M. Vandewalle, *Tetrahedron*, **1979**, 35(20), 2389-2396.
71. B. M. Trost, Y. Nishimura, K. Yamamoto, *J. Am. Chem. Soc.*, **1979**, 101(5), 1328-1330.
72. C. J. Rizzo, A. B. Smith III, *Tetrahedron Lett.*, **1988**, 29(23), 2793-2796.
73. R. M. Bettolo, P. Tagliatesta, A. Lupi, D. Bravetti, *Helv. Chim. Acta*, **1983**, 66(7), 1922-1928.
74. A. Lupi, M. Patamia, R. M. Bettolo, *Helv. Chim. Acta*, **1988**, 71(4), 872-875.
75. E. J. Corey, M. A. Tius, J. Das, *J. Am. Chem. Soc.*, **1980**, 102(5), 1742-1744.
76. J. Justicia, J. E. Oltra, J. M. Cuerva, *J. Org. Chem.*, **2005**, 70(21), 8265-8272.
77. F. Bilodeau, L. Dubé, P. Deslongchamps, *Tetrahedron*, **2003**, 59(16), 2781-2791.
78. G. Bélanger, P. Deslongchamps, *J. Org. Chem.*, **2000**, 65(21), 7070-7074.
79. M. Toyota, Y. Nishikawa, K. Fukumoto, *Tetrahedron*, **1994**, 50(38), 11153-11166.
80. M. Toyota, Y. Nishikawa, K. Fukumoto, *Tetrahedron Lett.*, **1995**, 36(30), 5379-5382.
81. M. Toyota, Y. Nishikawa, K. Fukumoto, *Tetrahedron*, **1996**, 52(31), 10347-10362.
82. M. Toyota, Y. Nishikawa, K. Fukumoto, *Tetrahedron Lett.*, **1994**, 35(35), 6495-6498.
83. R. A. Holton, R. M. Kennedy, H. B. Kim, M. E. Krafft, *J. Am. Chem. Soc.*, **1987**, 109(5), 1597-1600.
84. M. Toyota, M. Sasaki, M. Ihara, *Org. Lett.*, **2003**, 5(8), 1193-1195.
85. R. E. Ireland, J. D. Godfrey, S. Thaisrivongs, *J. Am. Chem. Soc.*, **1981**, 103(9), 2446-2448.
86. R. E. Ireland, W. C. Dow, J. D. Godfrey, S. Thaisrivongs, *J. Org. Chem.*, **1984**, 49(6), 1001-1013.
87. T. Tanaka, K. Murakami, O. Okuda, T. Inoue, T. Kuroda, K. Kamei, T. Murata, H. Yoshino, T. Imanishi, S.-W. Kim, C. Iwata, *Chem. Pharm. Bull.*, **1995**, 43(2), 193-197.
88. T. Tanaka, O. Okuda, K. Murakami, H. Yoshino, H. Mikamiyama, A. Kanda, S.-W. Kim, C. Iwata, *Chem. Pharm. Bull.*, **1995**, 43(8), 1407-1411.

89. J. E. McMurry, A. Andrus, G. M. Ksander, J. H. Musser, M. A. Johnson, *Tetrahedron*, **1981**, 37, 319-327.
90. S. P. Tanis, Y. H. Chuang, D. B. Head, *J. Org. Chem.*, **1988**, 53(21), 4929-4938.
91. E. E. Van Tamelen, S. R. Zawacky, R. K. Russell, J. G. Carlson, *J. Am. Chem. Soc.*, **1983**, 105(1), 142-143.
92. J. Wolinsky, R. Novak, R. Vasileff, *J. Org. Chem.*, **1964**, 29(12), 3596-3598.
93. L. Parra Rapado, V. Bulugahapitiya, P. Renaud, *Helv. Chim. Acta*, **2000**, 83(7), 1625-1632.
94. D. Seebach, R. Naef, *Helv. Chim. Acta*, **1981**, 64(8), 2704-2708.
95. S. V. Pansare, V. A. Adsool, *Org. Lett.*, **2006**, 8(10), 2035-2037.
96. K. Hiroya, K. Ogasawara, *Chem. Commun.*, **1998**, (18), 2033-2034.
97. S. F. Martin, G. J. Pacofsky, R. P. Gist, W. C. Lee, *J. Am. Chem. Soc.*, **1989**, 111(19), 7634-7636.
98. E. J. Corey, S. Kim, S. E. Yoo, K. C. Nicolaou, L. S. Melvin Jr, D. J. Brunelle, J. R. Falck, E. J. Trybulski, R. Lett, P. W. Sheldrake, *J. Am. Chem. Soc.*, **1978**, 100(14), 4620-4622.
99. S. F. Martin, G. J. Pacofsky, R. P. Gist, W. C. Lee, *J. Am. Chem. Soc.*, **1989**, 111(19), 7634-7636.
100. B. Chandra, D. Fu, S. G. Nelson, *Angew. Chem. Int. Edit.*, **2010**, 49(14), 2591-2594.
101. J. Mulzer, H. M. Kirstein, J. Buschmann, C. Lehmann, P. Luger, *J. Am. Chem. Soc.*, **1991**, 113(3), 910-923.
102. R. B. Woodward, E. Logusch, K. P. Nambiar, K. Sakan, D. E. Ward, B. W. Au-Yeung, C. H. Chen, R. B. Chenevert, A. Fliri, K. Frobel, H.-J. Gais, D. G. Garratt, K. Hayakawa, W. Heggie, D. P. Hesson, D. Hoppe, I. Hoppe, J. A. Hyatt, D. Ikeda, P. A. Jacobi, K. S. Kim, Y. Kobuke, K. Kojima, K. Krowicki, V. J. Lee, T. Leutert, S. Malchenko, J. Martens, R. S. Matthews, B. S. Ong, J. B. Press, T. V. Rajan Babu, G. Rousseau, H. M. Sauter, M. Suzuki, K. Tatsuta, L. M. Tolbert, E. A. Truesdale, I. Uchida, Y. Ueda, T. Uyehara, A. T. Vasella, W. C. Vladuchick, P. A. Wade, R. M. Williams, H. N.-C. Wong, *J. Am. Chem. Soc.*, **1981**, 103(11), 3210-3213.
103. E. J. Corey, E. J. Trybulski, L. S. Melvin Jr, K. C. Nicolaou, J. A. Secrist, R. Lett, P. W. Gheldrake, J. R. Falck, D. J. Brunelle, M. F. Haslanger, S. Kim, S. Yoo, *J. Am. Chem. Soc.*, **1978**, 100(14), 4618-4620.
104. T. Nakata, M. Fukui, T. Oishi, *Tetrahedron Lett.*, **1988**, 29(18), 2219-2222.
105. A. F. Sviridov, M. S. Ermolenko, D. V. Yashunsky, V. S. Borodkin, N. K. Kochetkov, *Tetrahedron Lett.*, **1987**, 28(33), 3835-3838.
106. A. F. Sviridov, M. S. Ermolenko, D. V. Yashunsky, V. S. Borodkin, N. K. Kochetkov, *Tetrahedron Lett.*, **1987**, 28(33), 3839-3842.
107. K. C. Nicolaou, Z. Yang, J. J. Liu, H. Ueno, P. G. Nantermet, R. K. Guy, C. F. Claiborne, J. Renaud, E. A. Couladouros, K. Paulvannan, E. J. Sorensen, *Nature*, **1994**, 367(6464), 630.
108. R. A. Holton, C. Somoza, H. B. Kim, F. Liang, R. J. Biediger, P. D. Boatman, M. Shindo, C. C. Smith, S. Kim, H. Nadizadeh, Y. Suzuki, C. Tao, P. Vu, S. Tang, P. Zhang, K. K. Murthi, L. N. Gentile, J. H. Liu, *J. Am. Chem. Soc.*, **1994**, 116(4), 1597-1598.
109. R. A. Holton, H.-B. Kim, C. Somoza, F. Liang, R. J. Biediger, P. D. Boatman, M. Shindo, C. C. Smith, S. Kim, H. Nadizadeh, Y. Suzuki, C. Tao, P. Vu, S. Tang, P. Zhang, K. K. Murthi, L. N. Gentile, J. H. Liu, *J. Am. Chem. Soc.*, **1994**, 116(4), 1599-1600.
110. T. V. Magee, W. G. Bornmann, R. C. Isaacs, S. J. Danishefsky, *J. Org. Chem.*, **1992**, 57(12), 3274-3276.
111. J. J. Masters, J. T. Link, L. B. Snyder, W. B. Young, S. J. Danishefsky, *Angew. Chem. Int. Edit.*, **1995**, 34(16), 1723-1726.
112. S. J. Danishefsky, J. J. Masters, W. B. Young, J. T. Link, L. B. Snyder, T. V. Magee, D. K. Jung, R. C. A. Isaacs, W. G. Bornmann, C. A. Alaimo, C. A. Coburn, M. J. Di Grandi, *J. Am. Chem. Soc.*, **1996**, 118(12), 2843-2859.
113. T. V. Magee, W. G. Bornmann, R. C. Isaacs, S. J. Danishefsky, *J. Org. Chem.*, **1992**, 57(12), 3274-3276.
114. M. J. Di Grandi, C. A. Coburn, R. C. Isaacs, S. J. Danishefsky, *J. Org. Chem.*, **1993**, 58(27), 7728-7731.
115. N. L. Allinger, G. L. Wang, B. B. Dewhurst, *J. Org. Chem.*, **1974**, 39(12), 1730-1735.

116. S. Ahmad, M. Yousaf, A. Mansha, N. Rasool, A. F. Zahoor, F. Hafeez, S. M. A. Rizvi, *Synthetic Commun.*, **2016**, 46(17), 1397-1416.
117. E. L. van Rozendaal, **1994**, *Some approaches to the synthesis of taxol and its derivatives: total-synthesis based on a LHASA analysis and semi-synthesis starting from taxine B*. (Niepublikowana rozprawa doktorska). Nijmegen: Uniwersytet im. Radbouda w Nijmegen, [dostęp: 17.07.2019]: https://repository.ubn.ru.nl/bitstream/handle/2066/30052/mmubn000001_181282496.pdf
118. E. L. van Rozendaal, S. J. Kuster, E. T. Rump, H. W. Scheeren, *Synthetic Commun.*, **1994**, 24(3), 367-374.
119. E. J. Corey, *Angew. Chem. Int. Edit.*, **1991**, 30(5), 455-465.
120. A. P. Johnson, C. Marshall, *J. Chem. Inf. Comput. Sci.*, **1992**, 32(5), 418-425. & A. P. Johnson, C. Marshall, *J. Chem. Inf. Comput. Sci.*, **1992**, 32(5), 426-429.
121. A. K. Long, J. C. Kappos, S. D. Rubenstein, G. E. Walker, *J. Chem. Inf. Comput. Sci.*, **1994**, 34(4), 922-933.
122. T. W. Greene, P. G. Wuts, *Protective groups in organic synthesis*, wyd. 3., Wiley&Sons, inc. Nowy Jork, Chichester, Weinheim, Brisbane, Toronto, Singapore, **1998**
123. [dostęp: 26.04.2019]: <https://cen.acs.org/articles/92/i33/Reflections-ChemDraw.html>
124. H. Gelernter, N. S. Sridharan, A. J. Hart, S. C. Yen, F. W. Fowler, H. J. Shue, *The discovery of organic synthetic routes by computer*. [w:] *New Concepts I* (s. 113-150), Springer, Berlin, Heidelberg 1973.
125. H. L. Gelernter, A. F. Sanders, D. L. Larsen, K. K. Agarwal, R. H. Boivie, G. A. Spritzer, J. E. Searleman, *Science*, **1977**, 197(4308), 1041-1049.
126. H. W. Whitlock, *J. Am. Chem. Soc.*, **1962**, 84(17), 3412-3413
127. F. S. Fawcett, *Chem. Rev.*, **1950**, 47(2), 219-274.
128. P. Roach, R. Warmuth, *Angew. Chem. Int. Edit.*, **2003**, 42(26), 3039-3042.
129. J. Gauthier, P. Deslongchamps, *Can. J. Chem.*, **1967**, 45(3), 297-300.
130. V. J. Lee, K. L. Rinehart Jr, *J. Antibiot.*, **1980**, 33(4), 408-415.
131. V. J. Lee, A. R. Branfman, T. R. Herrin, K. L. Rinehart Jr, *J. Am. Chem. Soc.*, **1978**, 100(13), 4225-4236.
132. J. D. Benstock, D. J. Berndt, K. K. Agarwal, *Discrete Appl. Math.*, **1988**, 19(1-3), 45-63.
133. H. Gelernter, J. R. Rose, C. Chen, *J. Chem. Inf. Comput. Sci.*, **1990**, 30(4), 492-504.
134. D. Krebsbach, H. Gelernter, S. M. Sieburth, *J. Chem. Inf. Comput. Sci.*, **1998**, 38(4), 595-604.
135. W. Katouda, **2004**, *Development of Hybrid Web System for Supporting Process Research.*, [dostęp: 18.07.2019]: https://www.sumitomo-chem.co.jp/english/rd/report/files/docs/20040108_nb2.pdf
136. M. Bersohn, *Bull. Chem. Soc. Jpn.*, **1972**, 45(6), 1897-1903.
137. M. Takahashi, I. Dogane, M. Yoshida, H. Yamachika, T. Takabatake, M. Bersohn, *J. Chem. Inf. Comput. Sci.*, **1990**, 30(4), 436-441.
138. T. Takabatake, **2009**, *SYNSUP, Synthetic Route Design System*, [dostęp: 18.07.2019]: https://www.sumitomo-chem.co.jp/english/rd/report/files/docs/20090204_6uf.pdf
139. A. Tanaka, T. Kawai, T. Takabatake, N. Oka, H. Okamoto, M. Bersohn, *Tetrahedron*, **2007**, 63(41), 10226-10236.
140. A. Tanaka, T. Kawai, T. Takabatake, N. Oka, H. Okamoto, M. Bersohn, *Tetrahedron Lett.*, **2006**, 47(38), 6733-6737.
141. A. Tanaka, H. Okamoto, M. Bersohn, *J. Chem. Inf. Model.*, **2010**, 50(3), 327-338.
142. I. Dogane, T. Takabatake, M. Bersohn, *Recl. Trav. Chim. Pays-Bas*, **1992**, 111(6), 291-296.
143. A. Tanaka, T. Kawai, T. Takabatake, H. Okamoto, M. Bersohn, *J. Comput. Aided Chem.*, **2009**, 10, 104-117.
144. Y. Gao, J. M. Klunder, R. M. Hanson, H. Masamune, S. Y. Ko, K. B. Sharpless, *J. Am. Chem. Soc.*, **1987**, 109(19), 5765-5780.
145. Controulis, J., Rebstock, M. C., & Crooks, H. M. (1949). *Journal of the American Chemical Society*, 71(7), 2463-2468.
146. D. Szulczewski, F. Eng, *Chloramphenicol* [w:] *Analytical profiles of drug substances*, (vol. 4, s. 47-90). Elsevier Academic Press, **1975**.
147. L. M. Rice, B. S. Sheth, J. W. Wheeler, *J. Heterocyclic Chem.*, **1973**, 10, 731-735.
148. K. C. Nicolaou, E. J. Sorensen, *Classics in Total Synthesis*, Wiley-VCH, Weinheim **1996**

149. [dostęp: 18.07.2019]: <https://www.foreignaffairs.com/articles/united-states/2009-03-01/japan-fallacy>
150. [dostęp: 18.07.2019]: <https://www.theguardian.com/commentisfree/2010/nov/30/britains-economy-what-happened-japan>
151. K. Satoh, K. Funatsu, *J. Chem. Inf. Comput. Sci.*, **1999**, 39(2), 316-325.
152. K. Funatsu, S. I. Sasaki, *Tetrahedron Comput. Meth.*, **1988**, 1(1), 27-37.
153. K. Funatsu, C. A. Del Carpio, S. I. Sasaki, *Tetrahedron Comput. Meth.*, **1988**, 1(1), 39-51.
154. K. Hori, H. Sadatomi, A. Miyamoto, T. Kuroda, M. Sumimoto, H. Yamamoto, *Molecules*, **2010**, 15(11), 8289-8304.155.
155. K. Satoh, S. Azuma, H. Satoh, K. Funatsu, *J. Chem. Softw.*, **1998**, 4(3), 101-111.
156. F. Feng, L. Lai, J. Pei, *Front. Chem.*, **2018**, 6(a. 199), 1-10.
157. K. Satoh, A. Imura, A. Miyadera, K. Kanai, Y. Yukimoto, *Chem. Pharm. Bull.*, **1998**, 46(4), 587-590.
158. T. Toshiaki, S. Sadahiro, S. Tatsuru, O. Naoki, H. Keiichi, E. Tutomu, A. Toshifumi, Patent: PCT/JP96/01373, WO96/37470, Opublikowano: 28.11.1996.
159. K. Hori, H. Sadatomi, K. Okano, M. Sumimoto, A. Miyamoto, S. Hayashi, H. Yamamoto, *J. Comput. Aided Chem.*, **2007**, 8, 12-18.
160. A. Tanaka, T. Kawai, T. Matsumoto, T. Takabatake, H. Okamoto, K. Funatsu, *J. Comput. Aided Chem.*, **2008**, 9, 81-91.
161. [dostęp: 18.07.2019]: <https://trademark.trademarkia.com/aiphoskosp-76622676.html>
162. S. Hanessian, J. Franco, B. Larouche, *Pure Appl. Chem.*, **1990**, 62(10), 1887-1910.
163. S. Hanessian, M. Botta, B. Larouche, A. Boyaroglu, *J. Chem. Inf. Comput. Sci.*, **1992**, 32(6), 718-722.
164. S. Hanessian, *Curr. Opin. Drug Discovery Dev.*, **2005**, 8, 798-819.
165. E. V. Gordeeva, D. E. Lushnikov, N. S. Zefirov, *Tetrahedron Comput. Meth.*, **1990**, 3(6), 445-459.
166. E. V. Gordeeva, D. E. Lushnikov, N. S. Zefirov, *Tetrahedron*, **1992**, 48(18), 3789-3804.
167. D. E. Lushnikov, N. S. Zefirov, *J. Chem. Inf. Comput. Sci.*, **1992**, 32(4), 317-322.
168. L. A. Kas'jan, L. Y. Gnedenko, M. F. Galafeeva, M. Y. Kornilov, P. A. Krasutsky, N. V. Averina, N. S. Zefirov, *Tetrahedron Lett.*, **1986**, 27(25), 2921-2922.
169. N. S. Zefirov, E. V. Gordeeva, S. S. Tratch, *J. Chem. Inf. Comput. Sci.*, **1988**, 28(4), 188-193.
170. N. S. Zefirov, E. V. Gordeeva, *J. Org. Chem.*, **1988**, 53(3), 527-532.
171. L. Baumer, G. Sala, G. Sello, *J. Am. Chem. Soc.*, **1991**, 113(7), 2494-2500.
172. L. Baumer, I. Compagnari, G. Sala, G. Sello, *Recl. Trav. Chim. Pays-Bas*, **1992**, 111(6), 297-303.
173. G. Sello, *J. Chem. Inf. Comput. Sci.*, **1994**, 34(1), 120-129.
174. E. S. Blurock, *J. Chem. Inf. Comput. Sci.*, **1990**, 30(4), 505-510.
175. W. T. Wipke, T. M. Dyott, *J. Am. Chem. Soc.*, **1974**, 96(15), 4834-4842.
176. W. T. Wipke, G. I. Ouchi, S. Krishnan, *Artifi. Intell.*, **1978**, 11(1-2), 173-193.
177. W. T. Wipke, W. J. Howe, *Computer-Assisted Organic Synthesis*, American Chemical Society, Waszyngton, **1976**.
178. G. Mehta, R. Barone, M. Chanon, *Eur. J. Org. Chem.*, **1998**, 7, 1409-1412.
179. H. Satoh, K. Funatsu, *J. Chem. Inf. Comput. Sci.*, **1995**, 35(1), 34-44.
180. H. Satoh, K. Funatsu, *J. Chem. Inf. Comput. Sci.*, **1996**, 36(2), 173-184.
181. P. Azario, R. Barone, M. Chanon, *J. Org. Chem.*, **1988**, 53(4), 720-724.
182. G. Mehta, R. Barone, P. Azario, F. Barberis, M. Arbelot, M. Chanon, *Tetrahedron*, **1992**, 48(41), 8953-8962.
183. W. T. Wipke, D. Rogers, *J. Chem. Inf. Comput. Sci.*, **1984**, 24(2), 71-81
184. J. B. Hendrickson, *J. Am. Chem. Soc.*, **1977**, 99(16), 5439-5450.
185. J. B. Hendrickson, *Angew. Chem. Int. Edit.*, **1990**, 29(11), 1286-1295.
186. J. B. Hendrickson, *Acc. Chem. Res.*, **1986**, 19(9), 274-281.
187. J. B. Hendrickson, A. G. Toczko, *J. Chem. Inf. Comput. Sci.*, **1989**, 29(3), 137-145.
188. J. B. Hendrickson, P. Huang, *J. Chem. Inf. Comput. Sci.*, **1989**, 29(3), 145-151.
189. D. F. Chodosh, J. Hill, L. Shpilsky, W. L. Mendelson, *Recl. Trav. Chim. Pays-Bas*, **1992**, 111(6), 247-254.
190. P. Y. Johnson, I. Burnstein, J. Crary, M. Evans, T. Wang, *Designing an expert system for organic synthesis: the need for strategic planning*. American Chemical Society, Waszyngton, **1989**.
191. R. Doenges, B. T. Groebel, H. Nickelsen, J. Sander, *J. Chem. Inf. Comput. Sci.*, **1985**, 25(4), 425-430.

192. R. Moll, *J. Chem. Inf. Comput. Sci.*, **1994**, 34(1), 117-119.
193. J. Gasteiger, W. D. Ihlenfeldt, P. Röse, *Recl. Trav. Chim. Pays-Bas*, **1992**, 111(6), 270-290.
194. W. D. Ihlenfeldt, J. Gasteiger, *Angew. Chem. Int. Edit.*, **1996**, 34(23-24), 2613-2633.
195. R. Barone, M. Attolini, M. Arbelot, M. Chanon, *Eur. J. Org. Chem.*, **2006**, 22, 5184-5190.
196. W. L. Jorgensen, E. R. Laird, A. J. Gushurst, J. M. Fleischer, S. A. Gothe, H. E. Helson, G. D. Paderes, S. Sinclair, *Pure Appl. Chemistry*, **1990**, 62(10), 1921-1932.
197. J. Gasteiger, C. Jochum, *EROS a computer program for generating sequences of reactions*. [w:] *Organic Compounds* (s. 93-126). Springer, Berlin, Heidelberg **1978**.
198. J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Löw, M. Marsili, H. Saller, K. Yuki, A new treatment of chemical reactivity: *Development of EROS, an expert system for reaction prediction and synthesis design*. [w:] *Organic Synthesis, Reactions and Mechanisms* (s. 19-73). Springer, Berlin, Heidelberg **1987**.
199. P. Röse, J. Gasteiger, *Anal. Chim. Acta*, **1990**, 235, 163-168.
200. R. Höllering, J. Gasteiger, L. Steinhauer, K. P. Schulz, A. Herwig, *J. Chem. Inf. Comput. Sci.*, **2000**, 40(2), 482-494.
201. I. M. Socorro, J. M. Goodman, *J. Chem. Inf. Model.*, **2006**, 46(2), 606-614.
202. I. M. Socorro, K. Taylor, J. M. Goodman, *Org. Lett.* **2005**, 7(16), 3541-3544.
203. M. S. Molchanova, S. S. Tratch, N. S. Zefirov, *J. Phys. Org. Chem.*, **2003**, 16(8), 463-474.
204. J. Bauer, R. Herges, E. Fontain, I. Ugi, *Chimia*, **1985**, 39(2), 43-53.
205. I. Ugi, J. Bauer, K. Bley, A. Dengler, A. Dietz, E. Fontain, B. Gruber, R. Herges, M. Knauer, K. Reitsam, N. Stein, *Angew. Chem. Int. Edit.*, **1993**, 32(2), 201-227.
206. J. Bauer, *Tetrahedron Comput. Meth.*, **1989**, 2(5), 269-280.
207. N. S. Zefirov, I. I. Baskin, V. A. Palyulin, *J. Chem. Inf. Comput. Sci.*, **1994**, 34(4), 994-999.
208. E. V. Gordeeva, V. V. Shcherbukhin, N. S. Zefirov, *Tetrahedron Comput. Meth.*, **1990**, 3(6), 429-443.
209. V. V. Shcherbukhin, N. S. Zefirov, *J. Chem. Inf. Comput. Sci.*, **1995**, 35(1), 159-164.
210. E. Fontain, K. Reitsam, *J. Chem. Inf. Comput. Sci.*, **1991**, 31(1), 96-101.
211. P. Gund, *Computer-assisted Organic Synthetic Analysis*. [w:] *Annual Reports in Medicinal Chemistry* (vol. 12, s. 288-297). Elsevier Academic Press **1977**.
212. M. A. Ott, J. H. Noordik, *Recl. Trav. Chim. Pays-Bas*, **1992**, 111(6), 239-246.
213. A. Dengler, E. Fontain, M. Knauer, N. Stein, I. Ugi, *Recl. Trav. Chim. Pays-Bas*, **1992**, 111(6), 262-269.
214. G. Sello, M. Termini, *Tetrahedron*, **1997**, 53(10), 3729-3756.
215. J. H. Noordik, *The CAOS/CAMM Center: The Dutch national academic facility for computer-assisted organic synthesis and modeling* [w:] *Cheminformatics*. (s. 7-36), IOS Press, Amsterdam **2004**.
216. A. Cook, A. P. Johnson, J. Law, M. Mirzazadeh, O. Ravitz, A. Simon, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **2012**, 2(1), 79-107.
217. W. A. Warr, *Mol. Inform.*, **2014**, 33(6-7), 469-476.
218. I. I. Baskin, T. I. Madzhidov, I. S. Antipin, A. A. Varnek, *Russ. Chem. Rev.*, **2017**, 86(11), 1127-1156.
219. O. Engkvist, P. O. Norrby, N. Selmi, Y. H. Lam, Z. Peng, E. C. Sherer, W. Amberg, T. Erhard, L. A. Smyth, *Drug Discov. Today*, **2018**, 23(6), 1203-1218.
220. S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, B. A. Grzybowski, *Angew. Chem. Int. Edit.*, **2016**, 55(20), 5904-5937.
221. K. N. Houk, F. Liu, *Acc. Chem. Res.*, **2017**, 50(3), 539-543.
222. P. P. Plehiers, G. B. Marin, C. V. Stevens, K. M. Van Geem, *J. Cheminformatics*, **2018**, 10(1), 11.
223. C. W. Coley, W. H. Green, K. F. Jensen, *J. Chem. Inf. Model.*, **2019**, 59, 2529-2537
224. [dostęp: 18.07.2019]: <https://www.elsevier.com/solutions/reaxys/how-reaxys-works/synthesis-planner>
225. [dostęp: 18.07.2019]: <https://www.cas.org/products/scifinder/sciplanner>
226. B. A. Grzybowski, K. J. Bishop, B. Kowalczyk, C. E. Wilmer, *Nature Chem.*, **2009**, 1(1), 31.
227. K. J. Bishop, R. Klajn, B. A. Grzybowski, *Angew. Chem. Int. Edit.*, **2006**, 45(32), 5348-5354.
228. M. Fialkowski, K. J. Bishop, V. A. Chubukov, C. J. Campbell, B. A. Grzybowski, *Angew. Chem. Int. Edit.*, **2005**, 44(44), 7263-7269.

229. M. Kowalik, C. M. Gothard, A. M. Drews, N. A. Gothard, A. Weckiewicz, P. E. Fuller, B. A. Grzybowski, K. J. Bishop *Angew. Chem. Int. Edit.*, **2012**, 51(32), 7928-7932.
230. Informacja od Witolda Zielińskiego, przedstawiciela Elsevier, właściciela Reaxys, podczas szkolenia z obsługi bazy Reaxys, 27.03.2019 IChO PAN Warszawa.
231. O. Ravitz, *Drug Discov. Today: Technologies*, **2013**, 10(3), 443-449.
232. J. Law, Z. Zsoldos, A. Simon, D. Reid, Y. Liu, S. Y. Khew, A. P. Johnson, S. Major, R. A. Wade, H. Y. Ando, *J. Chem. Inf. Model.*, **2009**, 49(3), 593-602.
233. [dostęp: 18.07.2019]: <http://www.chemconnector.com/tag/simbiosys-archem/>
234. [dostęp: 18.07.2019]: <https://www.chemanager-online.com/en/news-opinions/headlines/wiley-acquires-simbiosys>
235. A. Bøgevig, H. J. Federsel, F. Huerta, M. G. Hutchings, H. Kraut, T. Langer, P. Löw, C. Oppawsky, T. Rein, H. Saller, *Org. Process. Res. Dev.*, **2015**, 19(2), 357-368.
236. M. H. Segler, M. Preuss, M. P. Waller, *Nature*, **2018**, 555(7698), 604.
237. M. H. Segler, M. P. Waller, *Chem.: Eur. J. Chemistry*, **2017**, 23(25), 5966-5971.
238. M. A. Kayala, C. A. Azencott, J. H. Chen, P. Baldi, *J. Chem. Inf. Model.*, **2011**, 51(9), 2209-2222.
239. C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, *ACS Cent. Sci.*, **2017**, 3(5), 434-443.
240. B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. L. Nguyen, S. Ho, J. Sloane, P. Wender, V. Pande, *ACS Cent. Sci.*, **2017**, 3(10), 1103-1113.
241. J. N. Wei, D. Duvenaud, A. Aspuru-Guzik, *ACS Cent. Sci.*, **2016**, 2(10), 725-732.
242. C. A. Grambow, A. Jamal, Y. P. Li, W. H. Green, J. Zador, Y. V. Suleimanov, *J. Am. Chem. Soc.*, **2018**, 140(3), 1035-1048.
243. P. Schwaller, T. Gaudin, D. Lanyi, C. Bekas, T. Laino, *Chem. Sci.*, **2018**, 9(28), 6091-6098.
244. J. H. Chen, P. Baldi, *J. Chem. Inf. Model.*, **2009**, 49(9), 2034-2043.
245. C. W. Coley, W. Jin, L. Rogers, T. F. Jamison, T. S. Jaakkola, W. H. Green, R. Barzilay, K. F. Jensen, *Chem. Sci.*, **2019**, 10(2), 370-377.
246. G. Sello, *J. Chem. Inf. Comput. Sci.*, **1992**, 32(6), 713-717.
247. A. Cadeddu, E. K. Wylie, J. Jurczak, M. Wampler-Doty, B. A. Grzybowski, *Angew. Chem. Int. Edit.*, **2014**, 53(31), 8108-8112.
248. C. W. Coley, L. Rogers, W. H. Green, K. F. Jensen, *ACS Cent. Sci.*, **2017**, 3(12), 1237-1245.
249. F. Pennerath, G. Niel, P. Vismara, P. Jauffret, C. Laurencço, A. Napoli, *J. Chem. Inf. Model.*, **2010**, 50(2), 221-239
250. D. Rappoport, A. Aspuru-Guzik, *J. Chem. Theory Comput.*, **2019**, 15(7), 4099-4112.
251. I. A. Watson, J. Wang, C. A. Nicolaou, *J. Cheminformatics*, **2019**, 11(1), 1:1-12.
252. J. L. Baylon, N. A. Cilfone, J. R. Gulcher, T. W. Chittenden, *J. Chem. Inf. Model.*, **2019**, 59(2), 673-688.
253. K. Lin, Y. Xu, J. Pei, L. Lai, **2019**, *Automatic Retrosynthetic Pathway Planning Using Template-free Models*, *arXiv preprint* [dostęp: 19.07.2019]: <https://arxiv.org/abs/1906.02308>
254. S. Avramova, N. Kochev, P. Angelov, *Data*, **2018**, 3(2), 14: 1-6.
255. X. Liu, P. Li, S. Song, **2019**, *Decomposing Retrosynthesis into Reactive Center Prediction and Molecule Generation*, *bioRxiv* [dostęp: 19.07.2019]: <https://www.biorxiv.org/content/biorxiv/early/2019/06/21/677849.full.pdf>
256. N. Arai, S. Yoshikawa, N. Yasuo, R. Yoshino, M. Sekijima, *J. Bioinform. Comput. Biol.*, **2018**, 16(03), 1840016: 1-13.
257. J. H. Winter, *J. Chem. Inf. Comput. Sci.*, **1985**, 25(4), 389-391.
258. G. Pirok, N. Máté, J. Varga, J. Szegezdi, M. Vargyas, S. Dóránt, F. Csizmadia, *J. Chem. Inf. Model.*, **2006**, 46(2), 563-568.
259. A. Masoumi, M. Antoniazzi, M. Soutchanski, **2015**, *Modeling Organic Chemistry and Planning Organic Synthesis*. [w:] *GCAI*, (s. 176-195).
260. B. B. Masek, D. S. Baker, R. J. Dorfman, K. DuBrucq, V. C. Francis, S. Nagy, B. L. Richey, F. Soltanshahi, *J. Chem. Inf. Model.*, **2016**, 56(4), 605-620.
261. C. W. Coley, W. H. Green, K. F. Jensen, *Acc. Chem. Res.*, **2018**, 51(5), 1281-1289.
262. J. S. Schreck, C. W. Coley, K. J. Bishop, *ACS Cent. Sci.*, **2019**, 5(6), 970-981.

263. Q. Huang, L. L. Li, S. Y. Yang, *J. Chem. Inf. Model.*, **2011**, 51(10), 2768-2777.
264. T. Nakayama, *J. Chem. Inf. Comput. Sci.*, **1995**, 35(5), 885-893.
265. J. B. Hendrickson, *J. Am. Chem. Soc.*, **1971**, 93(25), 6847-6854.
266. N. S. Zefirov, *Acc. Chem. Res.*, **1987**, 20(7), 237-243.
267. J. Schwan, M. Christmann, *Chem. Soc. Rev.*, **2018**, 47(21), 7985-7995.
268. N. Z. Burns, P. S. Baran, R. W. Hoffmann, *Angew. Chem. Int. Edit.*, **2009**, 48(16), 2854-2867.
269. J. Li, M. D. Eastgate, *Org. Biomol. Chem.*, **2015**, 13(26), 7164-7176.
270. N. Schneider, D. M. Lowe, R. A. Sayle, G. A. Landrum, *J. Chem. Inf. Model.*, **2015**, 55(1), 39-53.
271. L. Wittgenstein, *Dociekania filozoficzne*, PWN, Warszawa **2004**.
272. T. Klucznik, B. Mikulak-Klucznik, M. P. McCormack, H. Lima, S. Szymkuć, M. Bhowmick, K. Molga, Y. Zhou, L. Rickershauser, E. P. Gajewska, A. Toutchkine, P. Dittwald, M. P. Startek, G. J. Kirkovits, R. Roszak, A. Adamski, B. Sieredzińska, M. Mrsksich, S. L. J. Trice, B. A. Grzybowski, *Chem*, **2018**, 4(3), 522-532.
273. B. A. Grzybowski, S. Szymkuć, E. P. Gajewska, K. Molga, P. Dittwald, A. Wołos, T. Klucznik, *Chem*, **2018**, 4(3), 390-398.
274. K. Molga, P. Dittwald, B. A. Grzybowski, *Chem*, **2019**, 5(2), 460-473.
275. T. Badowski, K. Molga, B. A. Grzybowski, *Chem. Sci.*, **2019**, 10(17), 4640-4651.
276. K. Molga, E. Gajewska, S. Szymkuć, B. A. Grzybowski, *React. Chem. Eng.*, **2019**, Advance Art. 1-48.
277. D. Weininger, *J. Chem. Inf. Comput. Sci.*, **1988**, 28(1), 31-36.
278. [dostęp: 18.07.2019]: <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>
279. F. S. Emami, A. Vahid, E. K. Wylie, S. Szymkuć, P. Dittwald, K. Molga, B. Grzybowski, *Angew. Chem. Int. Edit.*, **2015**, 54(37), 10797-10801.
280. S. Sana, A. M. M. Tasneem, K. C. Rajanna, P. K. Saiprakash, *Synth. Commun.*, **2009**, 39, 2949-2953.
281. R. P. A. Cruz, O. Ottoni, C. A. M. Abella, L. B. Aquino, *Tetrahedron Lett.*, **2001**, 42, 1467-1469.
282. D. Liang, H. Luo, Y.-F. Liu, Z.-Y. Hao, Y. Wang, C.-L. Zhang, Q.-J. Zhang, R.-Y. Chen, D.-Q. Yu, *Tetrahedron*, **2013**, 69, 2093-2097.
283. L. Li, D. Qiu, J. Shi, Y. Li, *Org. Lett.*, **2016**, 18, 3726-3729.
284. W. Beker, E. P. Gajewska, T. Badowski, B. A. Grzybowski, *Angew. Chem. Int. Edit.*, **2019**, 58(14), 4515-4519.
285. P. G. K. Clark, L. C. C. Vieira, C. Tallant, O. Fedorov, D. C. Singleton, C. M. Rogers, O. P. Monteiro, J. M. Bennett, R. Baronio, S. Mueller, S. Müller, D. L. Daniels, J. Mendez, S. Knapp, P. E. Brennan, D. J. Dixon, *Angew. Chem. Int. Ed.*, **2015**, 54, 6217-6221.
286. A. Miki, M. Tatsuno, M. Katagi, M. Nishikawa, H. Tsuchihashi, *J. Anal. Toxicol.*, **2002**, 26, 87-93.
287. G. A. Archer, L. H. Sternbach, *Chem. Rev.*, **1968**, 68(6), 747-784.
288. L. H. Sternbach, *Angew. Chem. Int. Edit.*, **1971**, 10(1), 34-43.
289. M. Getlik, D. Smil, C. Zepeda-Velazquez, Y. Bolshan, G. Poda, H. Wu, A. P. Dong, E. Kuznetsova, R. Marcellus, G. Senisterra, L. Dombrowski, T. Hajian, T. Kiyota, M. Schapira, C. H. Arrowsmith, P. J. Brown, M. Vedadi, R. Al-awar, *J. Med. Chem.*, **2016**, 59, 2478-2496.
290. H. C. Wu, M. J. Cheng, C. F. Peng, S. C. Yang, H. S. Chang, C. H. Lin, C. J. Wang, I. S. Chen, *Phytochemistry*, **2012**, 82, 118-127.
291. I. S. Kim, M. Y. Ngai, M. J. Krische, *J. Am. Chem. Soc.*, **2008**, 130(44), 14891-14899.
292. I. S. Kim, M. Y. Ngai, M. J. Krische, *J. Am. Chem. Soc.*, **2008**, 130(20), 6340-6341.
293. T. Sugawara, M. Adachi, K. Sasakura, A. Kitagawa, *The J. Org. Chem.*, **1979**, 44(4), 578-586.
294. R. A. Glennon, C. Chaurasia, M. Titeler, *J. Med. Chem.*, **1990**, 33(10), 2777-2784.
295. R. A. Glennon, E. Schubert, J. M. Jacyno, J. A. Rosecrans, *J. Med. Chem.*, **1980**, 23(11), 1222-1226.
296. M. J. Hewlins, A. H. Jackson, A. M. Oliveira-Campos, P. V. Shannon, *J. Chem. Soc., Perkin Trans. 1*, **1981**, 2906-2911.
297. S. H. Hohnloser, H. J. Crijns, M. Van Eickels, C. Gaudin, R. L. Page, C. Torp-Pedersen, S. J. Connolly, *N. Engl. J. Med.*, **2009**, 360(7), 668-678.
298. Y. Hu, Y. Zhang, Z. Yang, R. Fathi, *J. Org. Chem.*, **2002**, 67(7), 2365-2368.
299. D. R. Chancellor, K. E. Davies, O. De Moor, C. R. Dorgan, P. D. Johnson, A. G. Lambert, D. Lawrence, C. Lecci, C. Maillol, P. J. Middleton, G. Nugent, S. D. Poignant, A. C. Potter, P. D. Price,

- R. J. Pye, R. Storer, J. M. Tinsley, R. van Well, R. Vickers, J. Vile, F. J. Wilkes, F. X. Wilson, S. P. Wren, G. M. Wynn, *J. Med. Chem.*, **2011**, 54(9), 3241-3250.
300. M. Banwell, D. Hockless, *Chem. Commun.*, **1997**, (23), 2259-2260.
301. S. Thorand, N. Krause, *J. Org. Chem.*, **1998**, 63(23), 8551-8553.
302. S. Korneev, H. Rosemeyer, *Helv. Chim. Acta*, **2013**, 96(2), 201-216.
303. D. O. Kiesewetter, W. C. Eckelman, *J. Label. Compd. Radiopharm.*, **2004**, 47(13), 953-969.
304. G. H. Chu, M. Gu, J. A. Cassel, S. Belanger, G. J. Stabley, R. N. DeHaven, N. Conway-James, M. Koblish, P. J. Little, D. L. DeHaven-Hudkins, R. E. Dolle, *Bioorg. Med. Chem. Lett.*, **2006**, 16(3), 645-648.
305. J. Gubin, P. Chatelain, J. Lucchetti, G. Rosseells, H. Inion, Patent: US5223510 (1993), Opublikowano: 29.06.1993.
306. J. M. Meyer, A. D. Loebel, E. Schweizer, *Expert Opin. Investig. Drugs*, **2009**, 18, 1715-1726.
307. M. Y. Katteboina, N. R. Pilli, R. Mullangi, R. R. Seelam, S. R. Satla, *Biomed. Chromatogr.*, **2016**, 30(7), 1065-1074.
308. P. M. Burden, T. H. Ai, H. Q. Lin, M. Akinci, M. Costandi, T. M. Hambley, G. A. Johnston, *J. Med. Chem.*, **2000**, 43(24), 4629-4635.
309. Y. Uozumi, L. Sang-Yong, T. Hayashi, *Tetrahedron Lett.*, **1992**, 33(47), 7185-7188.
310. M. Breuning, T. Häuser, C. Mehler, C. Däschlein, C. Strohmam, A. Oechsner, H. Braunschweig, *Beilstein J. Org. Chem.*, **2009**, 5(1), 81.
311. C. S. Korapala, J. Qin, G. K. Friestad, *Org. Lett.*, **2007**, 9(21), 4243-4246.
312. G. K. Friestad, A. Ji, J. Baltrusaitis, C. S. Korapala, J. Qin, *J. Org. Chem.*, **2012**, 77(7), 3159-3180.
313. P. J. Garegg, B. Samuelsson, *J. Chem. Soc., Perkin Trans. 1*, **1980**, 2866-2869.
314. J. Y. Wu, C. D. Kuo, C. Y. Chu, M. S. Chen, J. H. Lin, Y. J. Chen, H. F. Liao, *Molecules*, **2014**, 19(6), 6911-6928.
315. A. Khan, C. M. Marson, R. A. Porter, *Chem. Commun.*, **2001**, 31(11), 1753-1764.
316. M. Więcek, T. Kottke, X. Ligneau, W. Schunack, R. Seifert, H. Stark, J. Handzlik, K. Kieć-Kononowicz, *Bioorg. Med. Chem.*, **2011**, 19(9), 2850-2858.
317. U. P. Dhokte, H. C. Brown, *J. Org. Chem.*, **1997**, 62(4), 865-869.
318. F. Osterkamp, H. Wehlan, U. Koert, M. Wiesner, P. Raddatz, S. L. Goodman, *Tetrahedron*, **1999**, 55(35), 10713-10734.
319. G. W. Kabalka, T. M. Shoup, N. M. Goudgaon, *J. Org. Chem.*, **1989**, 54(25), 5930-5933.
320. G. W. Kabalka, T. M. Shoup, N. M. Goudgaon, *Tetrahedron Lett.*, **1989**, 30(12), 1483-1486.
321. H. C. Brown, G. Zweifel, *J. Am. Chem. Soc.*, **1961**, 83(2), 486-487.
322. H. C. Brown, G. Zweifel, *J. Am. Chem. Soc.*, **1960**, 82(17), 4708-4712.
323. M. Nussim, Y. Mazur, F. Sondheimer, *J. Org. Chem.*, **1964**, 29(5), 1120-1131.
324. T. Gibson, *Tetrahedron Lett.*, **1982**, 23(2), 157-158.
325. H. Zhang, S. Fu, M. Tao, Z. Liao, C. Qian, X. Chen, *Res. Chem. Intermediat.*, **2016**, 42(4), 3325-3332.
326. [dostęp: 23.02.2019]: <https://www.sigmaaldrich.com/catalog/search?term=5beta%2F6beta-Hydroxy+Lurasidone+%28Mixture+of+Diastereomers%29+solution&interface=Product%20Name&N=0+&mode=mode%20matchpartialmax&lang=pl®ion=PL&focus=productN=0%20220003048%20219853286%20219853218>
327. K. H. Lim, V. J. Raja, T. D. Bradshaw, S. H. Lim, Y. Y. Low, T. S. Kam, *J. Nat. Prod.*, **2015**, 78(5), 1129-1138.
328. M. Ihara, F. Setsu, M. Shohda, N. Taniguchi, K. Fukumoto, *Heterocycles*, **1994**, 37(1), 289-292.
329. M. Ihara, F. Setsu, M. Shohda, N. Taniguchi, Y. Tokunaga, K. Fukumoto, *J. Org. Chem.*, **1994**, 59(18), 5317-5323.
330. M. Lounasmaa, D. D. Belle, A. Tolvanen, *Tetrahedron Lett.*, **1995**, 36(39), 7141-7144.
331. M. Lounasmaa, K. Karinen, D. D. Belle, A. Tolvanen, *Tetrahedron*, **1998**, 54(1-2), 157-164.
332. B. Danieli, G. Lesma, S. Maccellini, D. Passarella, A. Silvani, *Tetrahedron: Asymmetry*, **1999**, 10(20), 4057-4064.
333. B. Danieli, G. Lesma, D. Passarella, A. Sacchetti, A. Silvani, *Tetrahedron Lett.*, **2001**, 42(41), 7237-7240.

334. M. Suzuki, M. Ihara, *Heterocycles*, **2000**, 52(3), 1083-1085.
335. T. L. Ho, E. Gorobets, *Tetrahedron*, **2002**, 58(24), 4969-4973.
336. C. Y. Chen, B. R. Chang, M. R. Tsai, M. Y. Chang, N. C. Chang, *Tetrahedron*, **2003**, 59(47), 9383-9387.
337. C. Y. Chen, M. Y. Chang, R. T. Hsu, S. T. Chen, N. C. Chang, *Tetrahedron Lett.*, **2003**, 44(47), 8627-8630.
338. T. L. Ho, Q. X. Lin, *Tetrahedron*, **2008**, 64(45), 10401-10405.
339. D. B. England, A. Padwa, *J. Org. Chem.*, **2008**, 73(7), 2792-2802.
340. A. Gansäuer, D. Worgull, K. Knebel, I. Huth, G. Schnakenburg, *Angew. Chem. Int. Edit.*, **2009**, 48(47), 8882-8885.
341. R. W. Barnhart, X. Wang, P. Noheda, S. H. Bergens, J. Whelan, B. Bosnich, *J. Am. Chem. Soc.*, **1994**, 116(5), 1821-1830.
342. A. Job, C. F. Janeck, W. Bettray, R. Peters, D. Enders, *Tetrahedron*, **2002**, 12(58), 2253-2329.
343. H. Ahlbrecht, E. O. Düber, D. Enders, H. Eichenauer, P. Weuster, *Tetrahedron Lett.*, **1978**, 19(39), 3691-3694.
344. D. Enders, H. Kipphardt, P. Fey, *Org. Synth.*, **1987**, 65, 183.
345. B. Neises, W. Steglich, *Angew. Chem. Int. Edit.*, **1978**, 17(7), 522-524.
346. [dostęp: 23.07.2019]: <https://www.chemistryworld.com/news/software-could-revolutionisechemistry/1017236.article>
347. [dostęp: 23.07.2019]: <http://blogs.sciencemag.org/pipeline/archives/2016/04/12/the-algorithms-are-coming>
348. [dostęp: 23.07.2019]: https://cen.acs.org/analytical-chemistry/CENs-Year-Chemistry-2018/96/i49?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+cen_latestnews+%28Chemical+%26+Engineering+News%3A+Latest+News%29
349. [dostęp: 23.07.2019]: <https://www.heise.de/tp/features/Auf-dem-Weg-zum-Chemieroboter-4003178.html>
350. [dostęp: 23.07.2019]: <https://blogs.sciencemag.org/pipeline/archives/2018/03/06/retrosynthesis-here-it-comes>
351. [dostęp: 23.07.2019]: <https://phys.org/news/2018-03-chemists-chematica-ability-synthetic-routes.html>
352. [dostęp: 23.07.2019]: https://cen.acs.org/articles/96/i10/Chemists-test-computer-planned-syntheses.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+cen_latestnews+%28Chemical+%26+Engineering+News%3A+Latest+News%29
353. [dostęp: 23.07.2019]: <http://crosstalk.cell.com/blog/five-chem-papers-that-made-an-impact>
354. I. Kant, *Metafizyczne podstawy przyrodoznawstwa* [w:] „*Dziela zebrane t. III*”, Wyd. Naukowe UMK, Toruń **2013**.
355. J. Leonard, B. Lygo, G. Procter, *Advanced Practical Organic Chemistry*, wyd. 3., CRC Press, Boca Raton **2013**.
356. K. Sakai, T. Kobori, T. Fujisawa, *Tetrahedron Lett.*, **1981**, 22(2), 115-118.
357. M. J. Dimsdale, R. F. Newton, D. K. Rainey, C. F. Webb, T. V. Lee, S. M. Roberts, *Chem. Comm.*, **1977**, (20), 716-716.
358. A. D. Baxter, S. M. Roberts, F. Scheinmann, B. J. Wakefield, R. F. Newton, *Chem. Comm.*, **1983**, (17), 932-933.
359. S. M. Ali, C. B. Chapleo, M. A. Finch, S. M. Roberts, G. T. Woolley, R. J. Cave, R. F. Newton, *J. Chem. Soc., Perkin Trans. 1*, **1980**, 2093-2097.
360. M. A. Finch, S. M. Roberts, G. T. Woolley, R. F. Newton, *J. Chem. Soc., Perkin Trans. 1*, **1981**, 1725-1728.
361. A. D. Baxter, F. Binns, T. Javed, S. M. Roberts, P. Sadler, F. Scheinmann, B. J. Wakefield, M. Lynch, R. F. Newton, *J. Chem. Soc., Perkin Trans. 1*, **1986**, 889-900.
362. J. F. Bickley, V. Jadhav, S. M. Roberts, M. G. Santoro, A. Steiner, P. W. Sutton, *Synlett*, **2003**, (08), 1170-1174.
363. S. B. Usmani, E. Takahisa, Y. Kobayashi, *Tetrahedron Lett.*, **1998**, 39(7), 601-604.

364. Y. Kobayashi, M. G. Muruges, M. Nakano, E. Takahisa, S. B. Usmani, T. Ainai, *J. Org. Chem.*, **2002**, 67(20), 7110-7123.
365. J. E. McMurry, *J. Am. Chem. Soc.*, **1968**, 90(24), 6821-6825.
366. E. J. Corey, M. Ohno, R. B. Mitra, P. A. Vatakencherry, *J. Am. Chem. Soc.*, **1964**, 86(3), 478-485.
367. J. A. Marshall, G. L. Bundy, W. I. Fanta, *J. Org. Chem.*, **1968**, 33(10), 3913-3922.
368. P. G. Sammes, L. J. Street, R. J. Whitby, *J. Chem. Soc., Perkin Trans. 1*, **1986**, 281-289.
369. J. A. Marshall, W. I. Fanta, H. Roebke, *J. Org. Chem.*, **1966**, 31(4), 1016-1020.
370. Y. Yasuhara, *J. Org. Chem.*, **1968**, 33, 4512.
371. J. A. Hyatt, P. W. Reynolds, *J. Org. Chem.*, **1984**, 49(2), 384-385.
372. E. J. Corey, R. K. Bakshi, S. Shibata, *J. Am. Chem. Soc.*, **1987**, 109(18), 5551-5553.
373. E. J. Corey, R. K. Bakshi, S. Shibata, C. P. Chen, V. K. Singh, *J. Am. Chem. Soc.*, **1987**, 109(25), 7925-7926.
374. E. J. Corey, S. Shibata, R. K. Bakshi, *J. Org. Chem.*, **1988**, 53(12), 2861-2863.
375. G. A. Poulton, C. A. Boulet, *Synth. Commun.*, **1990**, 20(21), 3347-3353.
376. I. Shimizu, Y. Lee, Y. Fujiwara, *Synlett*, **2000**, 2000(09), 1285-1286.
377. A. Covarrubias-Zúñiga, A. Gonzalez-Lucas, M. M. Domínguez, *Tetrahedron*, 2003, 59(11), 1989-1994.
378. H. Nakamura, K. Yasui, Y. Kanda, P. S. Baran, *J. Am. Chem. Soc.*, **2019**, 141(4), 1494-1497.

Streszczenie rozprawy doktorskiej

Komputerowo wspomagana synteza organiczna związków biologicznie aktywnych

mgr inż. Tomasz Klucznik

Próby stworzenia programów komputerowych zdolnych planować syntezy chemiczne dowolnych cząsteczek organicznych – sięgają lat 60-tych XX wieku. Jednak pomimo znacznego wysiłku próby te były w dużej mierze nieskuteczne, rodząc pytanie, czy planowanie syntezy organicznej można sprowadzić do podstawy algorytmicznej. Nadrzędnym celem mojej pracy doktorskiej było *wykazanie, że przy pomocy odpowiednio zaprojektowanej bazy danych o reakcjach cząsteczek organicznych wraz z algorytmami pomocniczymi – taka redukcja jest rzeczywiście możliwa, a planowanie syntezy może być sformułowane raczej jako problem algorytmiczny niż forma „sztuki” realizowanej i dostępnej tylko dla ludzkich ekspertów.*

Weryfikacja tej hipotezy wymagała wielu lat wspólnego wysiłku ode mnie i moich kolegów, co ostatecznie doprowadziło do powstania platformy oprogramowania o nazwie Chematica. Po jej zakończeniu Chematica musiała zostać zweryfikowana poprzez eksperymentalne wykonanie komputerowo zaprojektowanych planów syntetycznych. Ten wysiłek eksperymentalnej walidacji na przykładach nietrywialnych, biologicznie aktywnych cząsteczek, był moim głównym wkładem w projekt opisanym w niniejszej rozprawie.

Rozprawa rozpoczyna się historycznym przeglądem dziedziny, kilkoma wczesnymi próbami rozwiązania problemu, a następnie opisem podstaw koncepcyjnych programu Chematica. Następnie opisuję udaną walidację komputerowo zaprojektowanych syntez: leku Dronedaron, metabolitu 5 β /6 β -hydroksylurazydonu i hydrazonu będącego półproduktem wykorzystanym w syntezie totalnej takamonidyny (złożonego, bioaktywnego produktu naturalnego).

Co ciekawe, plany syntetyczne wygenerowane przez maszynę okazały się bardziej efektywne niż plany syntetyczne wcześniej zaprojektowane przez ludzkich ekspertów. Wyniki te stanowią pierwszą w historii demonstrację udanego projektu syntetycznego *in silico* i zapowiadają nową erę chemii organicznej, w której chemicy będą mogli przeprowadzać planowanie złożonych syntez za pomocą komputerów.

Summary of doctoral thesis

Computer-assisted organic synthesis of biologically active compounds

MSc. Eng. Tomasz Klucznik

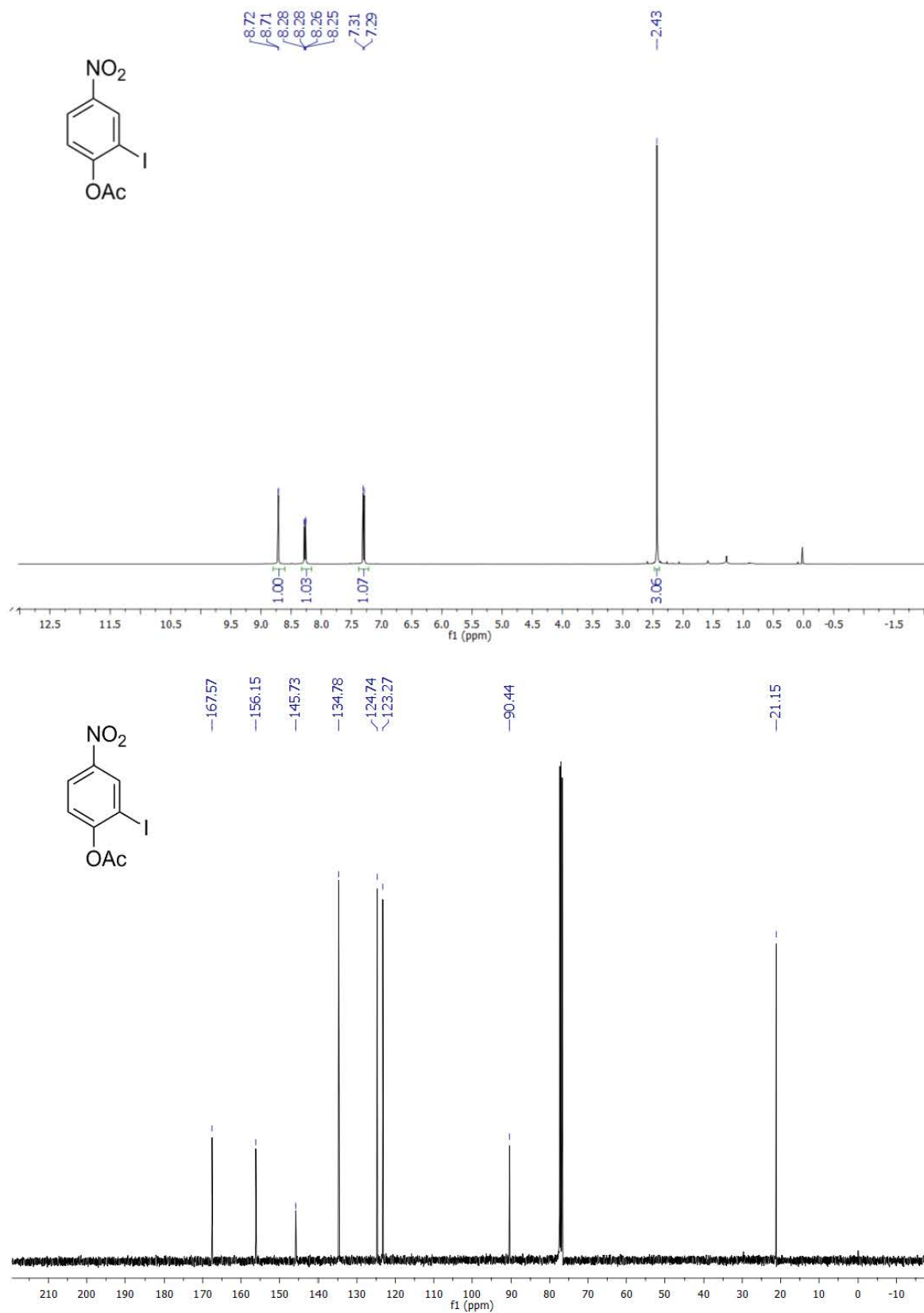
Attempts to create computer programs capable of planning syntheses of arbitrary target molecules date back to the 1960s – despite significant effort, however, these attempts have been largely unsuccessful raising the question whether organic-synthetic planning can be at all reduced to an algorithmic basis. The overriding aim of my doctoral work has been to *demonstrate that with appropriately designed reaction knowledge base and supporting algorithms such reduction is, indeed, possible and synthetic planning can be framed as an algorithmic problem rather than a form of “art” pursued by and available only to human experts.*

Verification of this hypothesis required many years of concerted effort by me and my colleagues, ultimately giving rise to a software platform called Chematica. Upon its completion, Chematica had to be validated by experimental execution of computer-designed routes. This experimental effort – yielding non-trivial, biologically active targets – has been my major contribution described in this thesis.

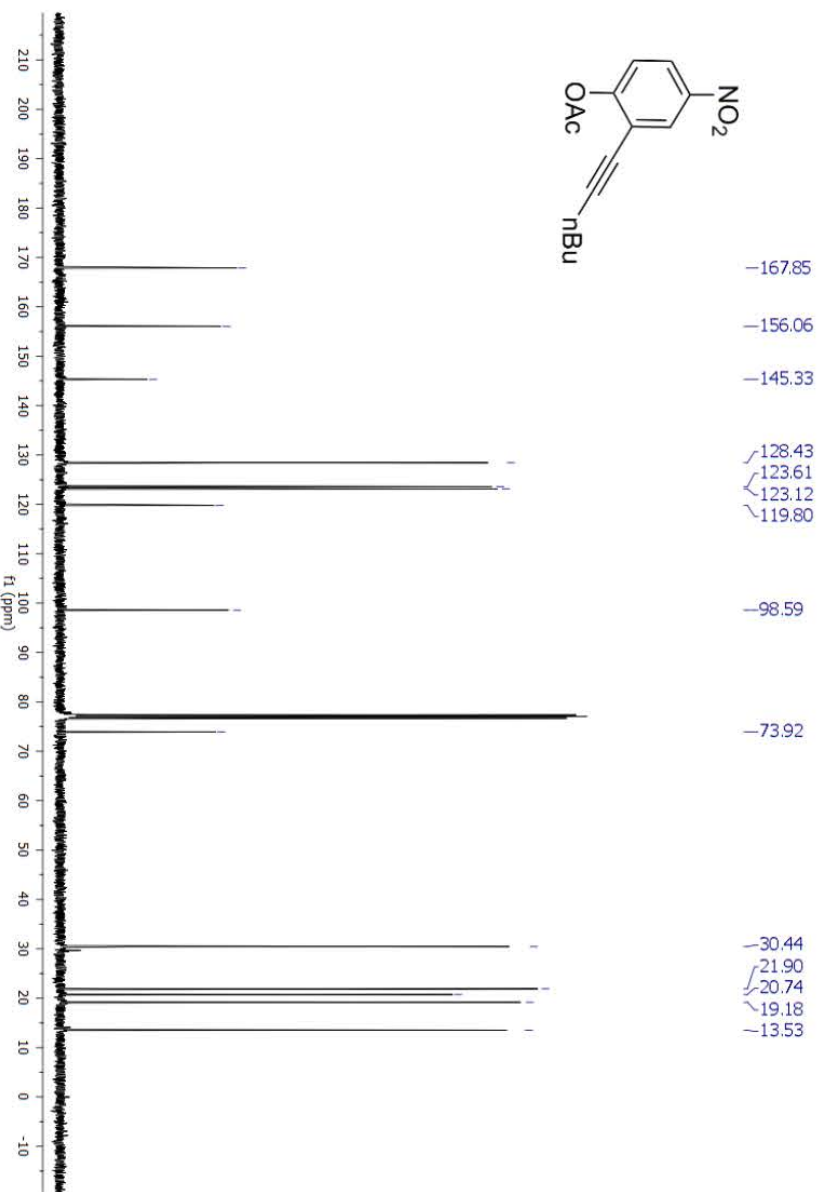
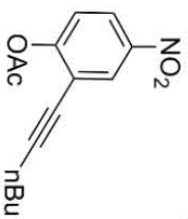
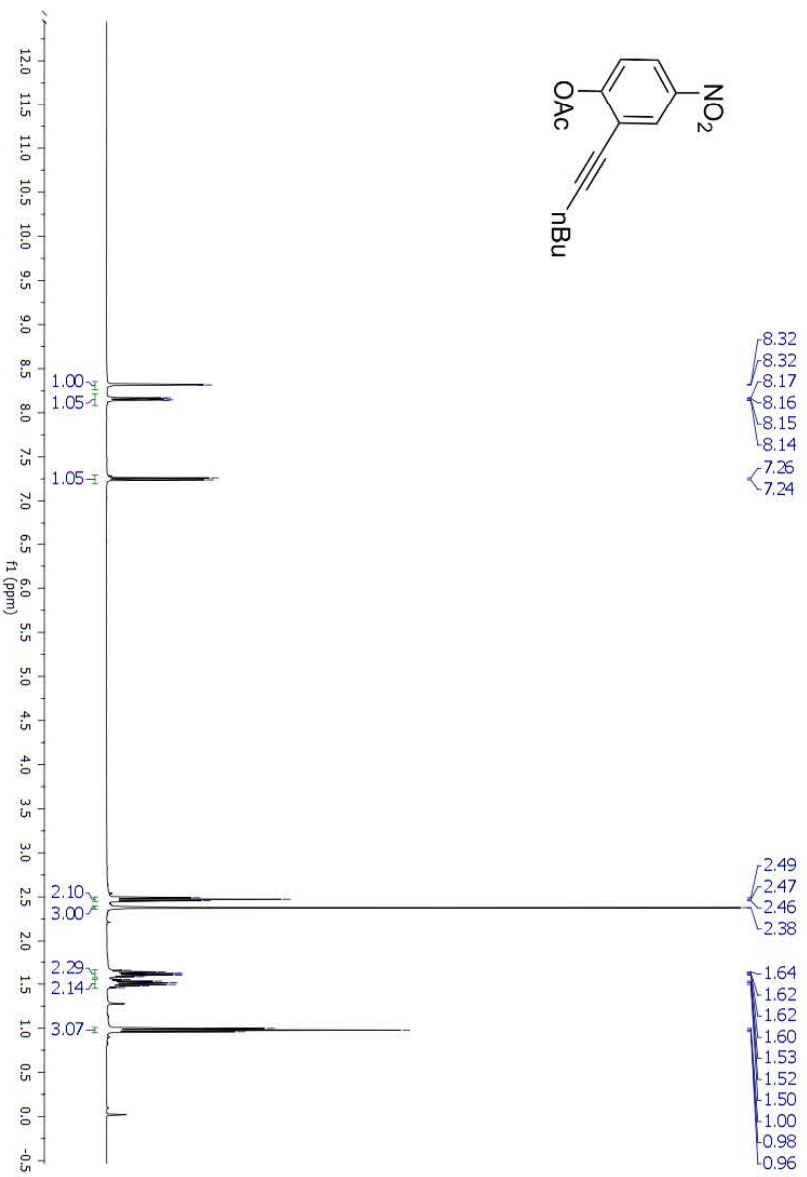
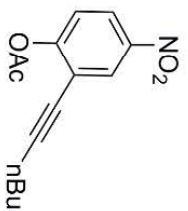
The thesis begins with a historical overview of the field, some early attempts and solutions, and then the conceptual basis of Chematica. It then describes successful validation of computer-designed syntheses of the Dronedarone drug, of 5 β /6 β -hydroxylurasidone metabolite of the Lurasidone drug, and of a hydrazone intermediate used for the total synthesis of tacamonidine (a complex bioactive natural product).

Remarkably, the synthetic plans generated by the machine proved more efficacious than previously available routes designed by human experts. These results constitute the first-ever demonstration of successful in silico synthetic design and herald a new era of organic chemistry in which chemists will be able to perform high-level chemical planning with the help of computers.

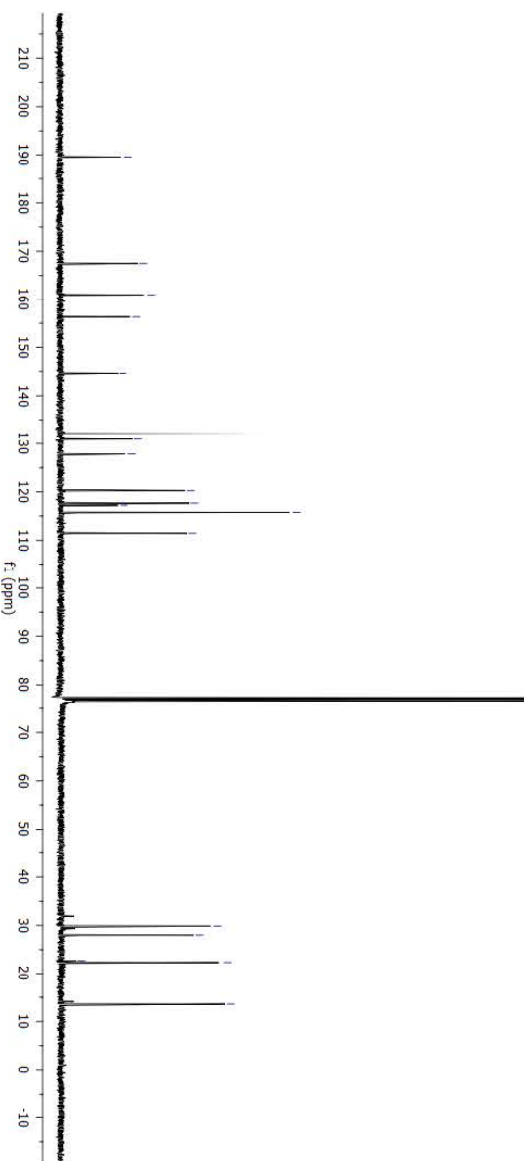
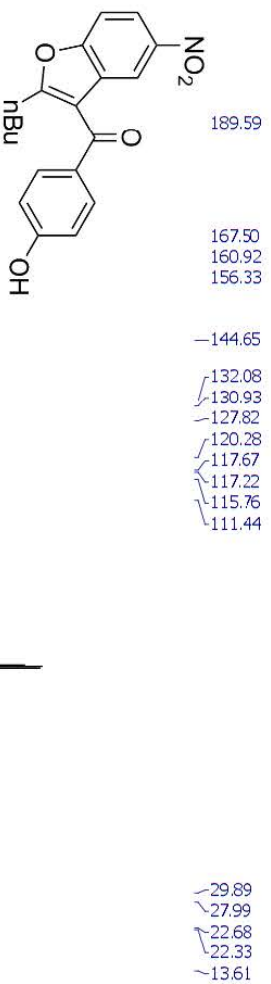
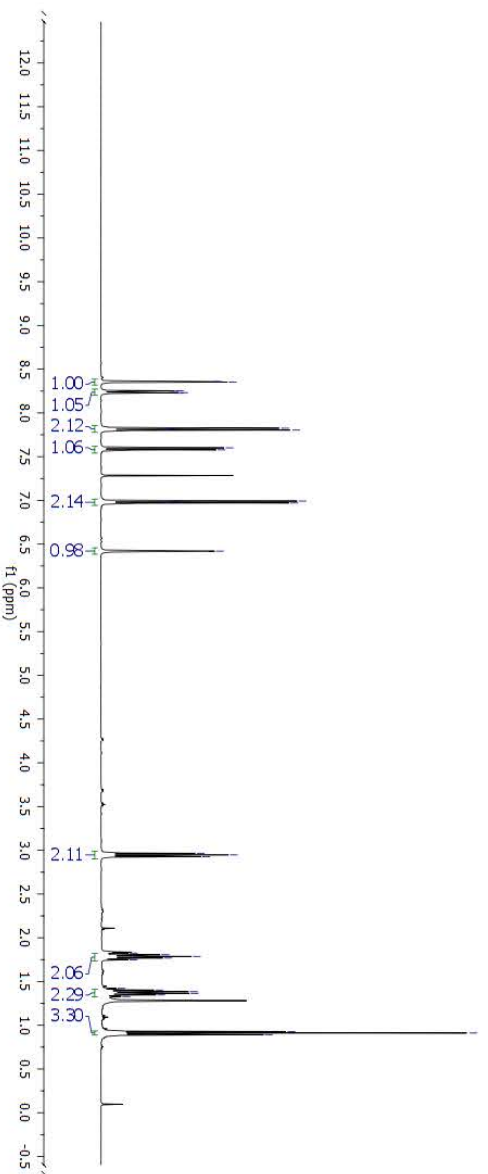
Appendiks I. Widma NMR i chromatogramy



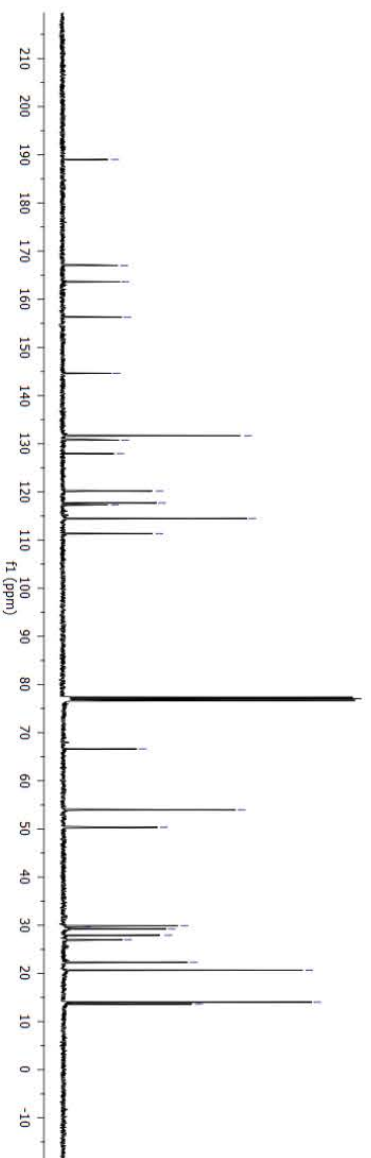
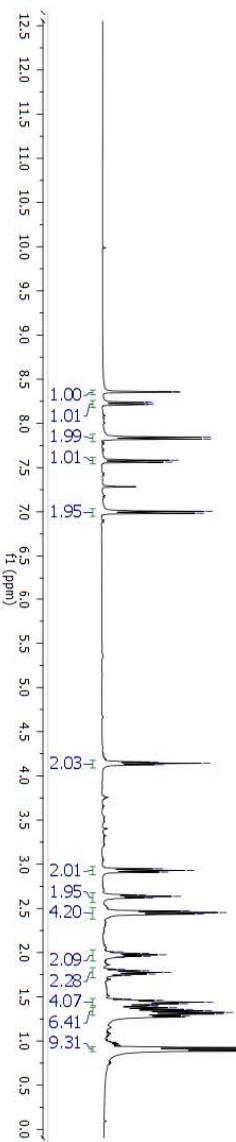
Ryc.69. Widma ¹H NMR (góra), ¹³C NMR (dół), dla związku 3.



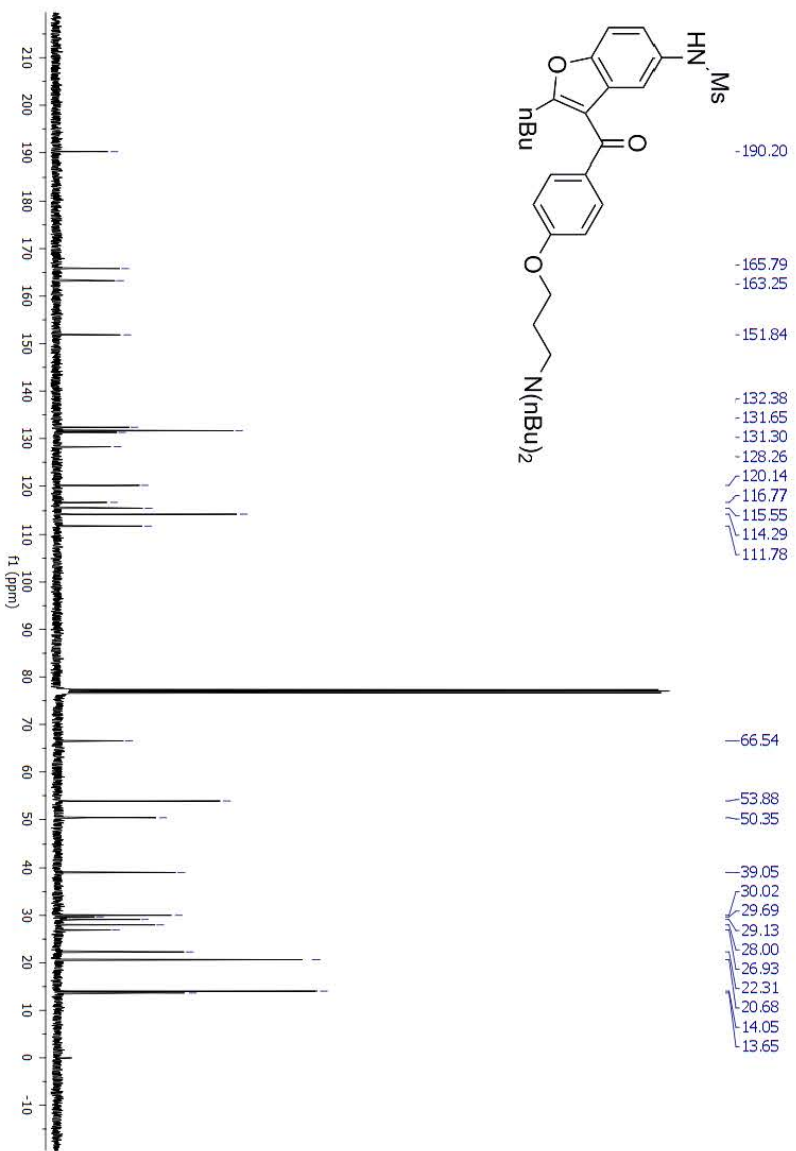
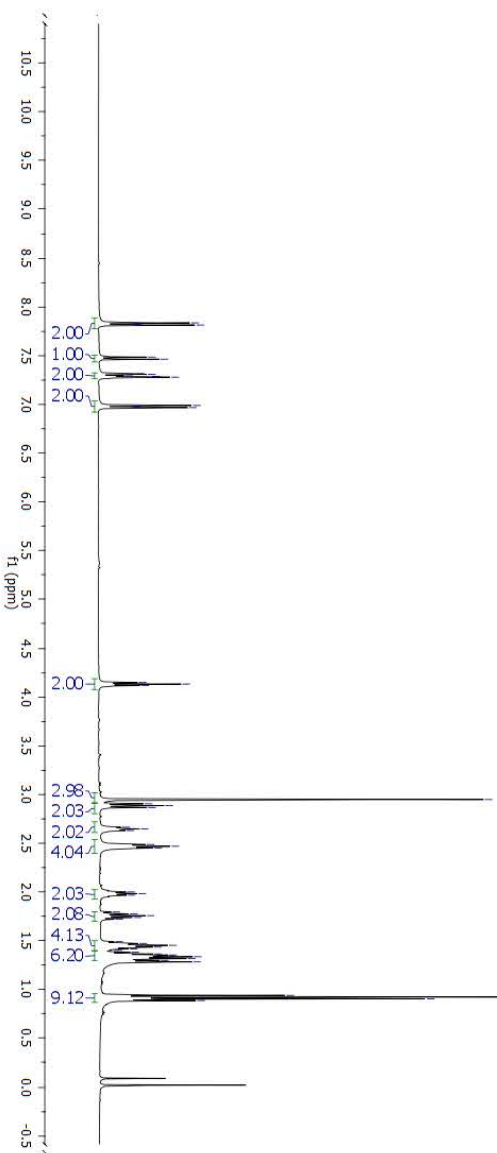
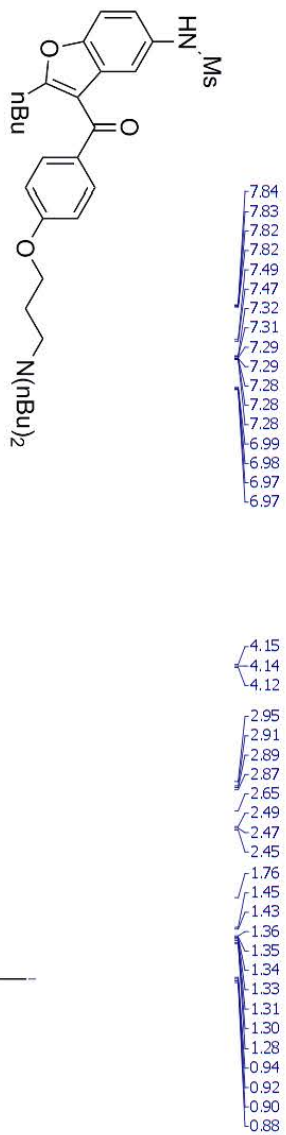
Ryc. 70. Widma ^1H NMR (górn), ^{13}C NMR (dół), dla związku 4.



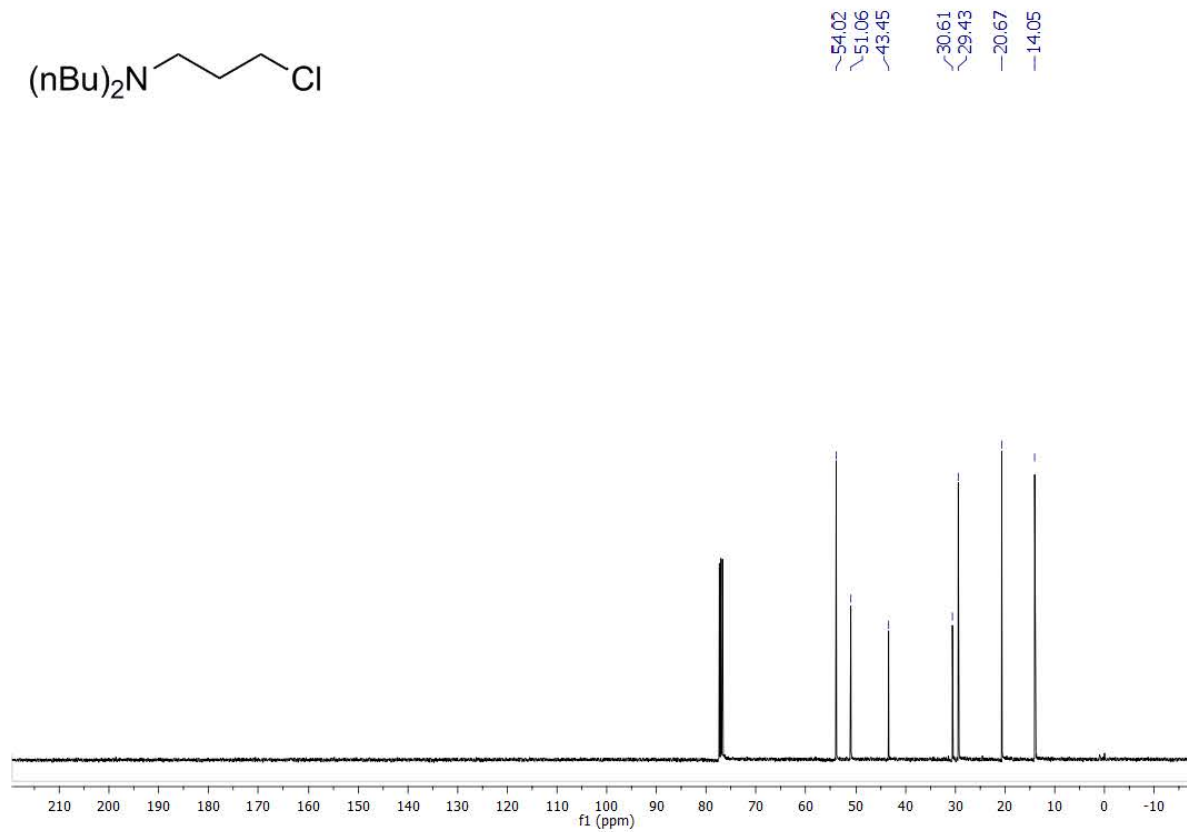
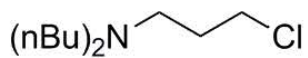
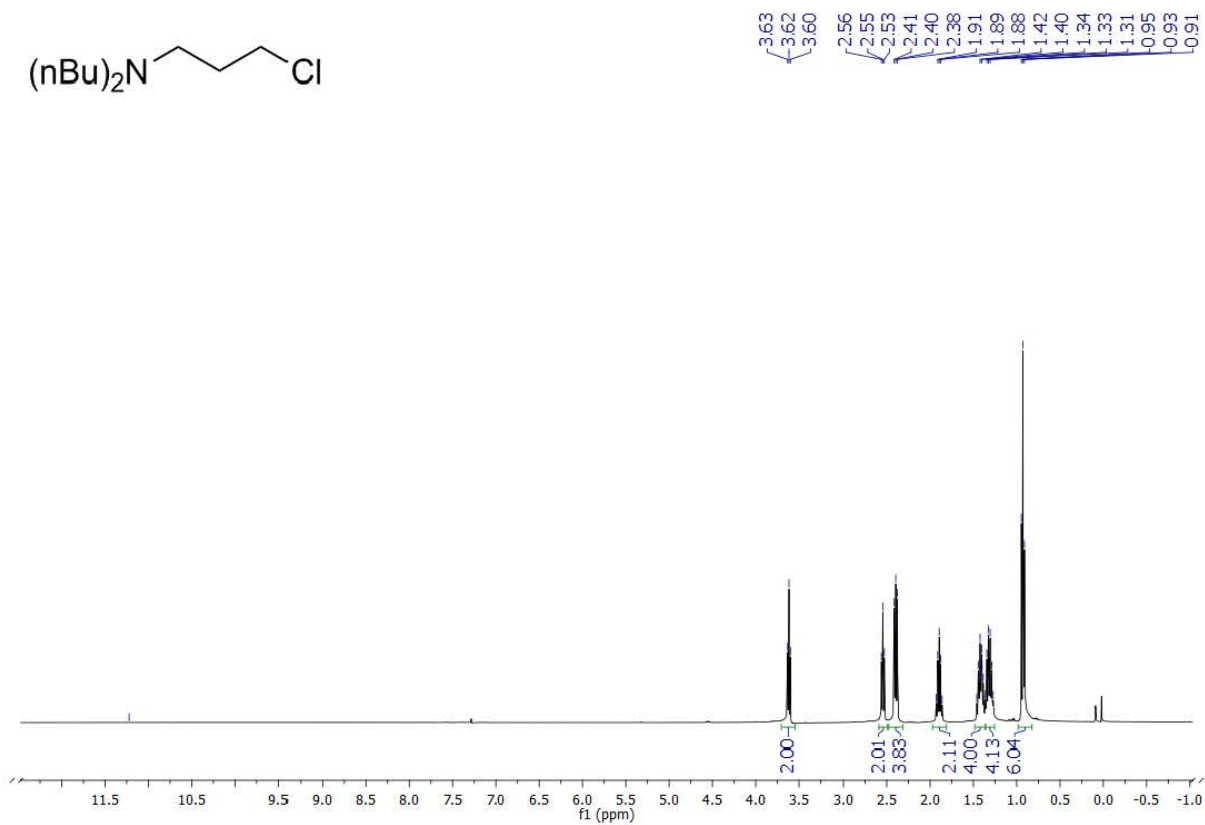
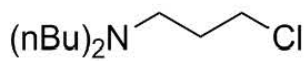
Ryc. 71. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla związku 6.



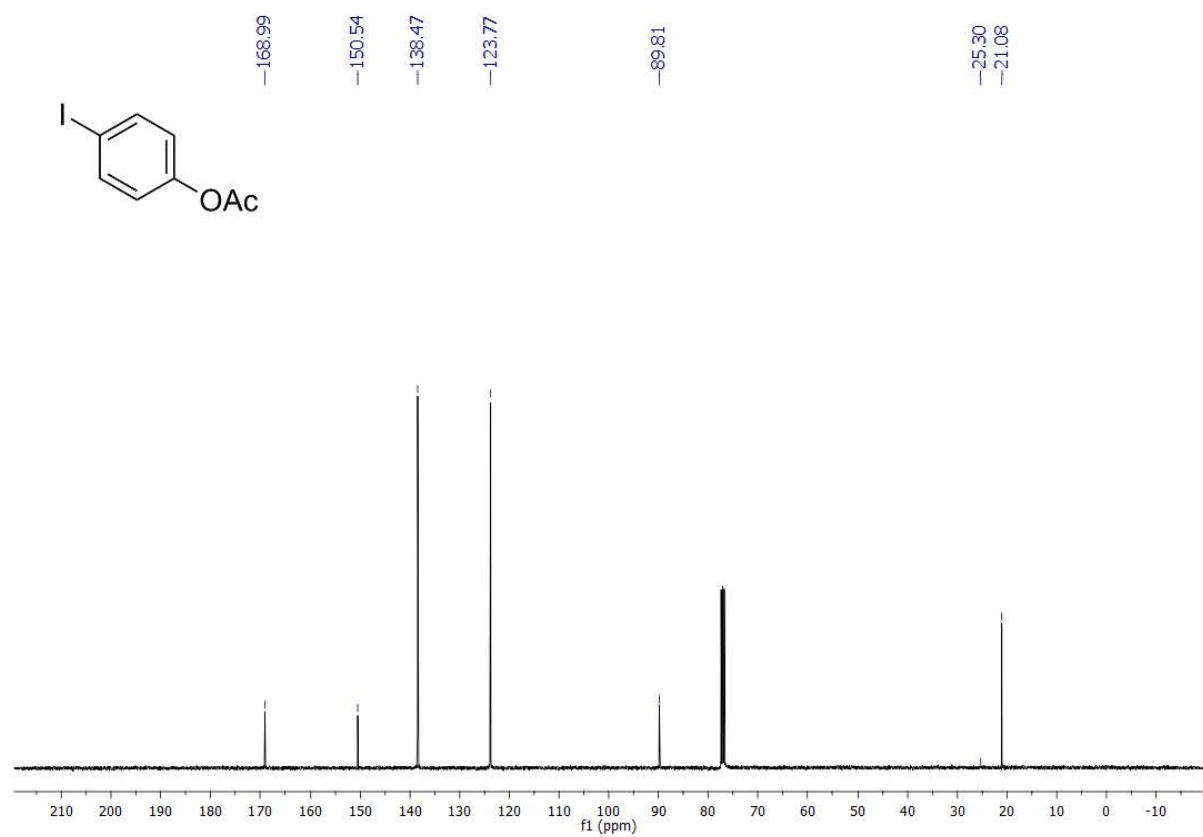
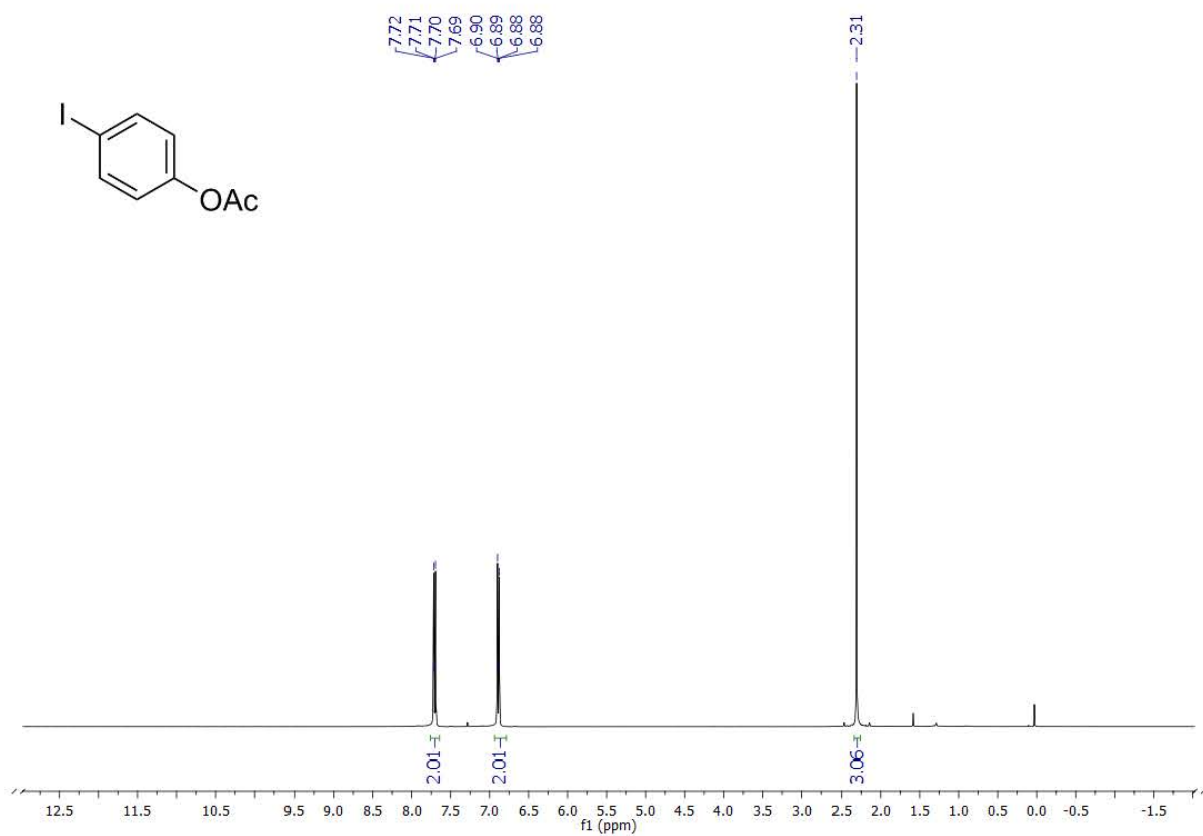
Ryc. 72. Widma ¹H NMR (górn), ¹³C NMR (dół), dla związku 7.



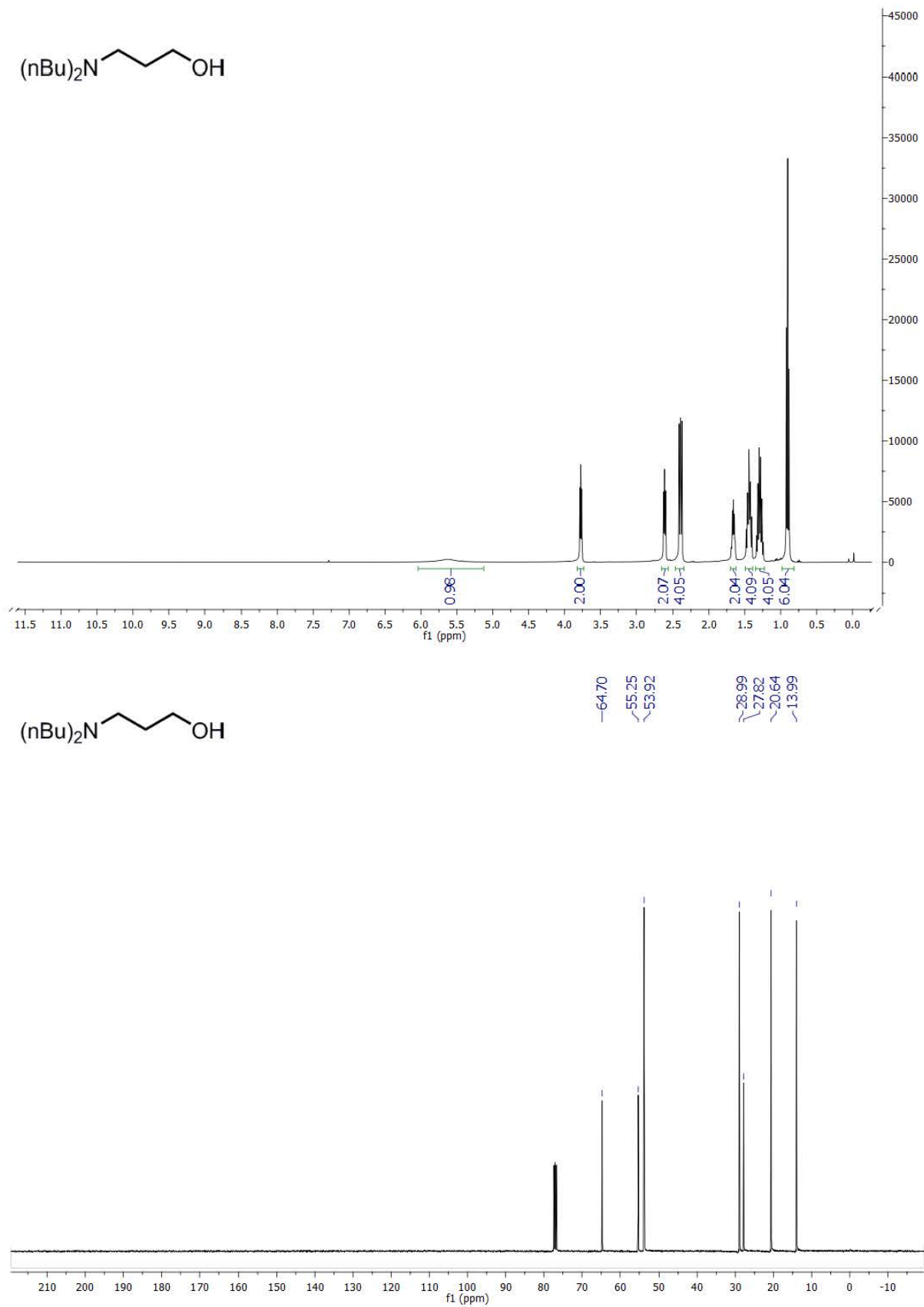
Ryc. 73. Widma ¹H NMR (górn), ¹³C NMR (dół), dla związku 1.



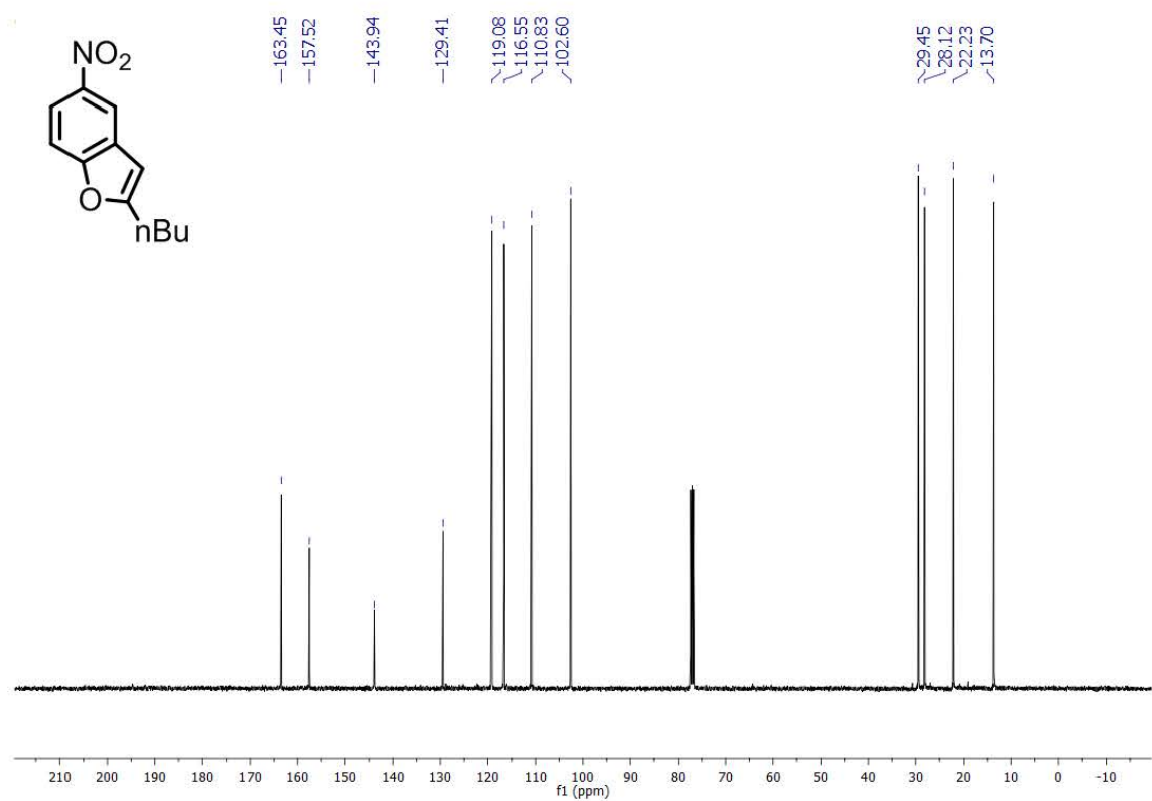
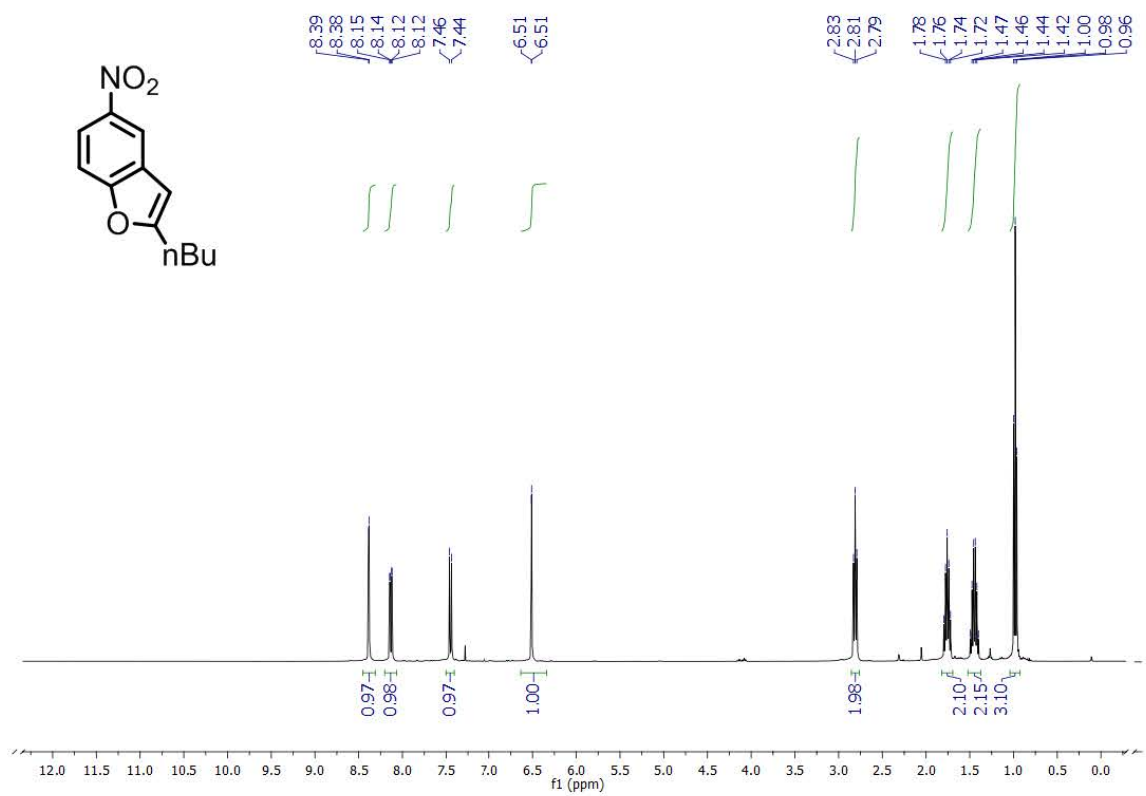
Ryc.74. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla związku **9**.



Ryc.75. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla związku 5.

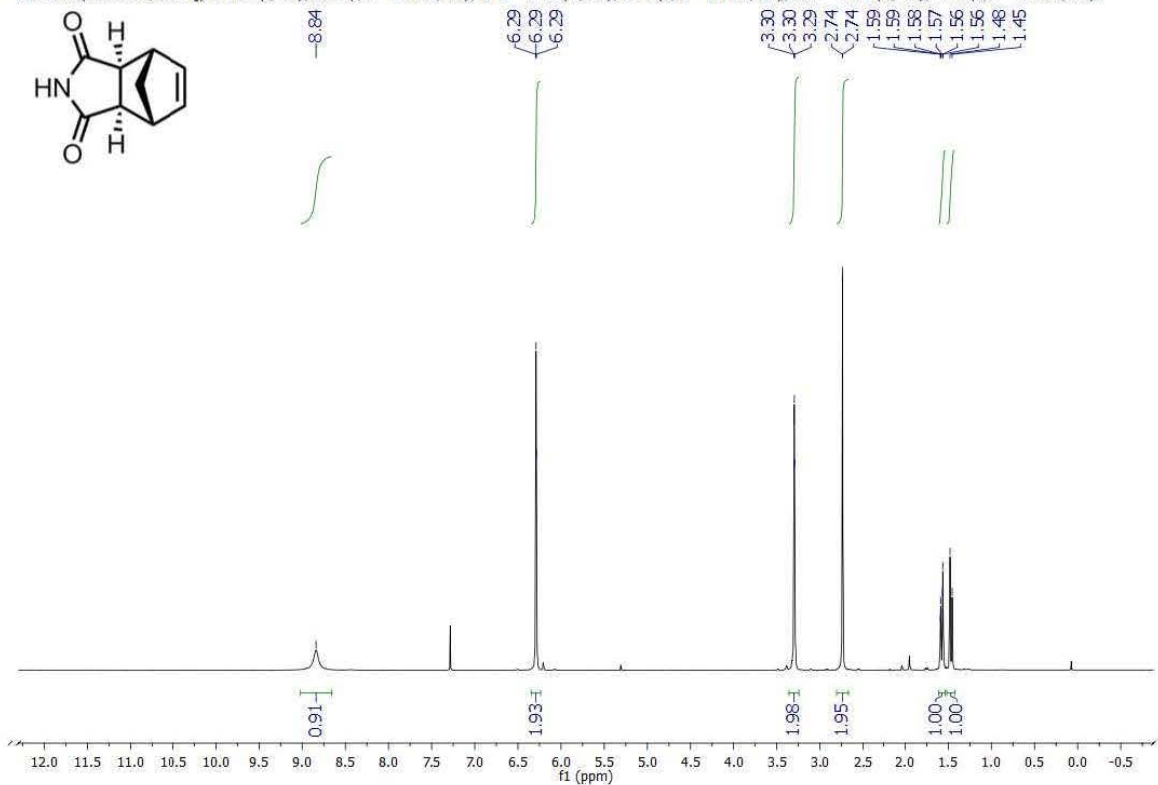


Ryc.76. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla związku **8**.

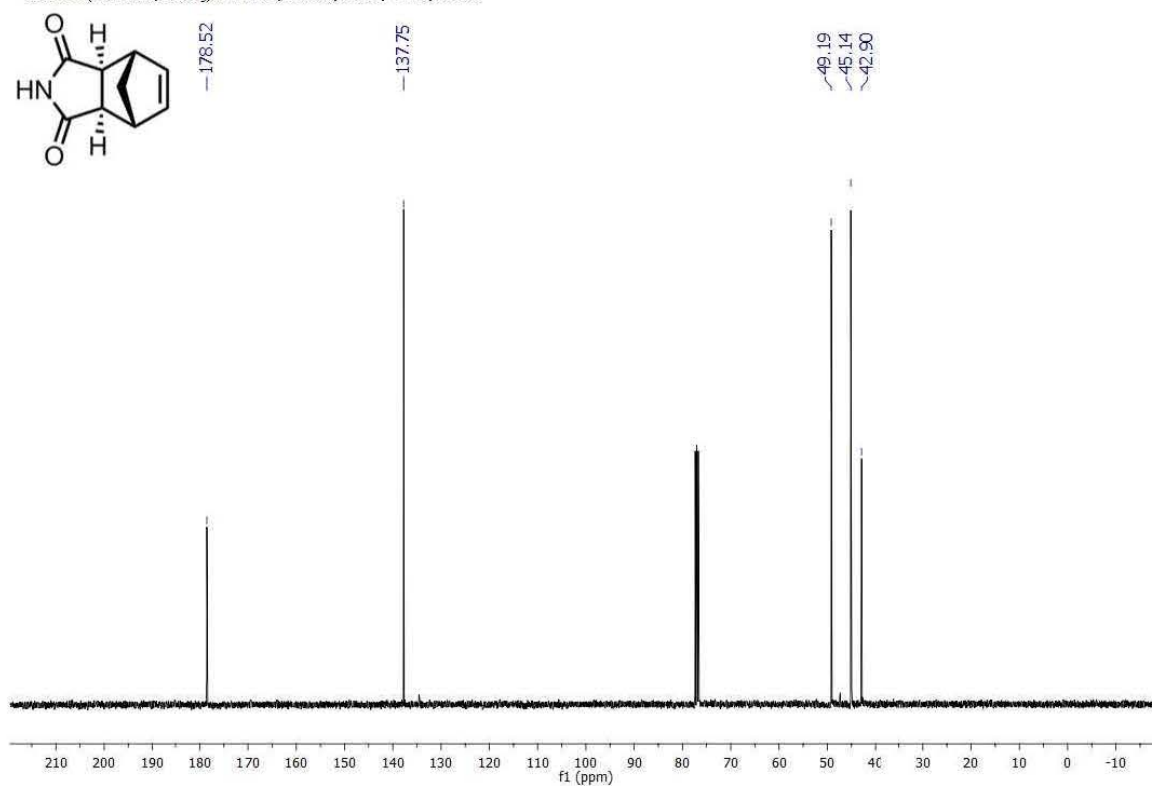


Ryc.77. Widma ¹H NMR (górn), ¹³C NMR (dół), dla związku **11**.

^1H NMR (400 MHz, CDCl_3) δ 8.84 (s, 1H), 6.29 (t, $J = 1.7$ Hz, 2H), 3.48 – 3.12 (m, 2H), 2.74 (d, $J = 1.1$ Hz, 2H), 1.63 – 1.54 (m, 1H), 1.47 (d, $J = 9.9$ Hz, 1H).

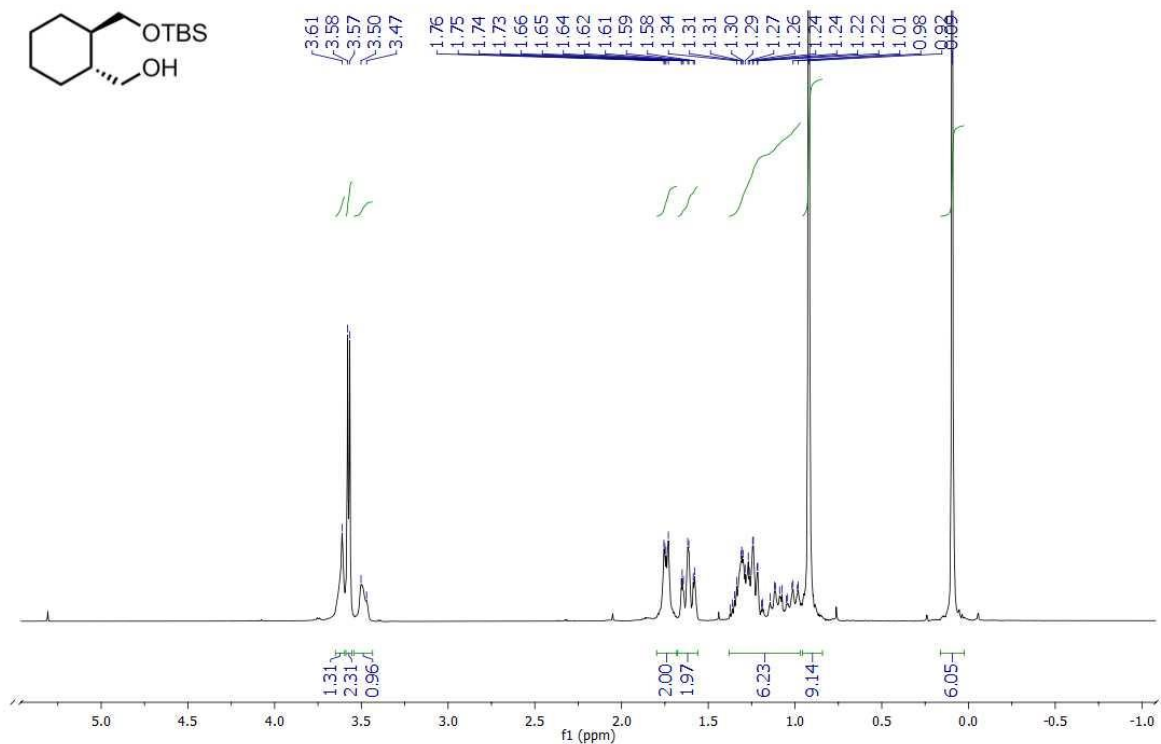


^{13}C NMR (101 MHz, CDCl_3) δ 178.52, 137.75, 49.19, 45.14, 42.90.

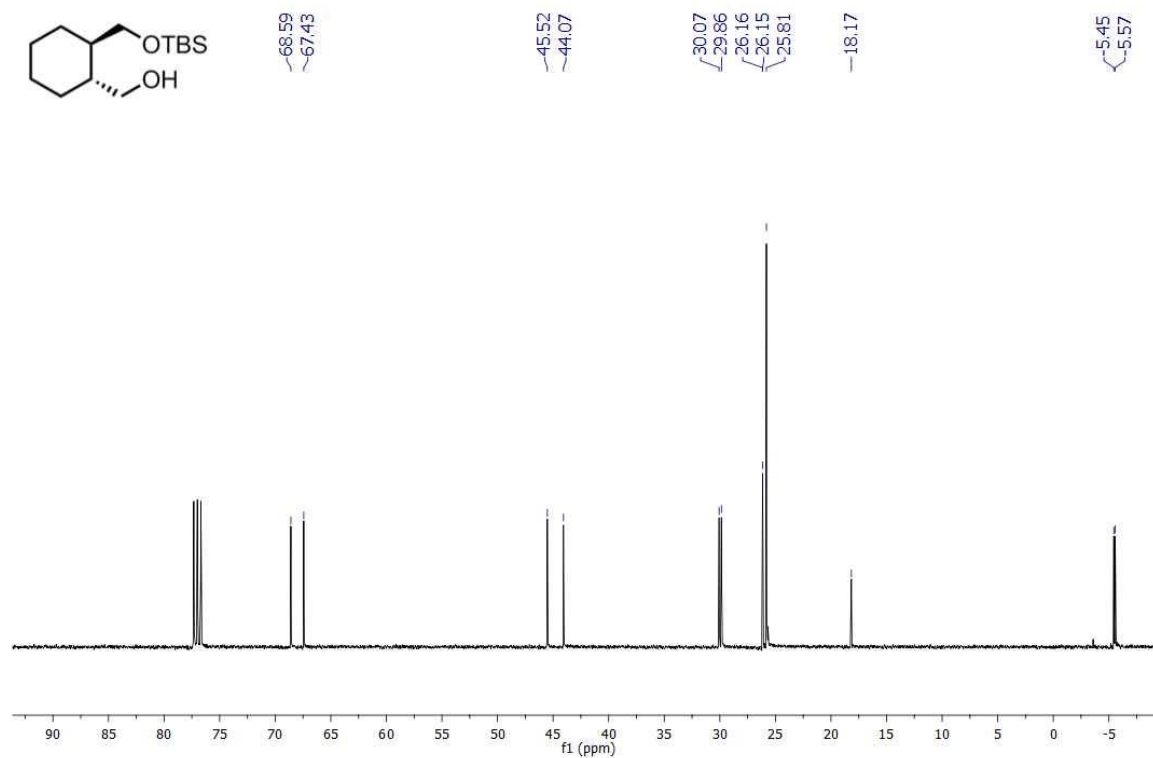


Ryc.78. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla związku 14.

^1H NMR (400 MHz, CDCl_3) δ 3.61 (s, 1H), 3.57 (d, $J = 5.0$ Hz, 2H), 3.52–3.45 (m, 1H), 1.79–1.69 (m, 2H), 1.70–1.62 (m, 2H), 1.38–0.97 (m, 6H), 0.92 (s, 9H), 0.10 (t, $J = 1.3$ Hz, 6H).

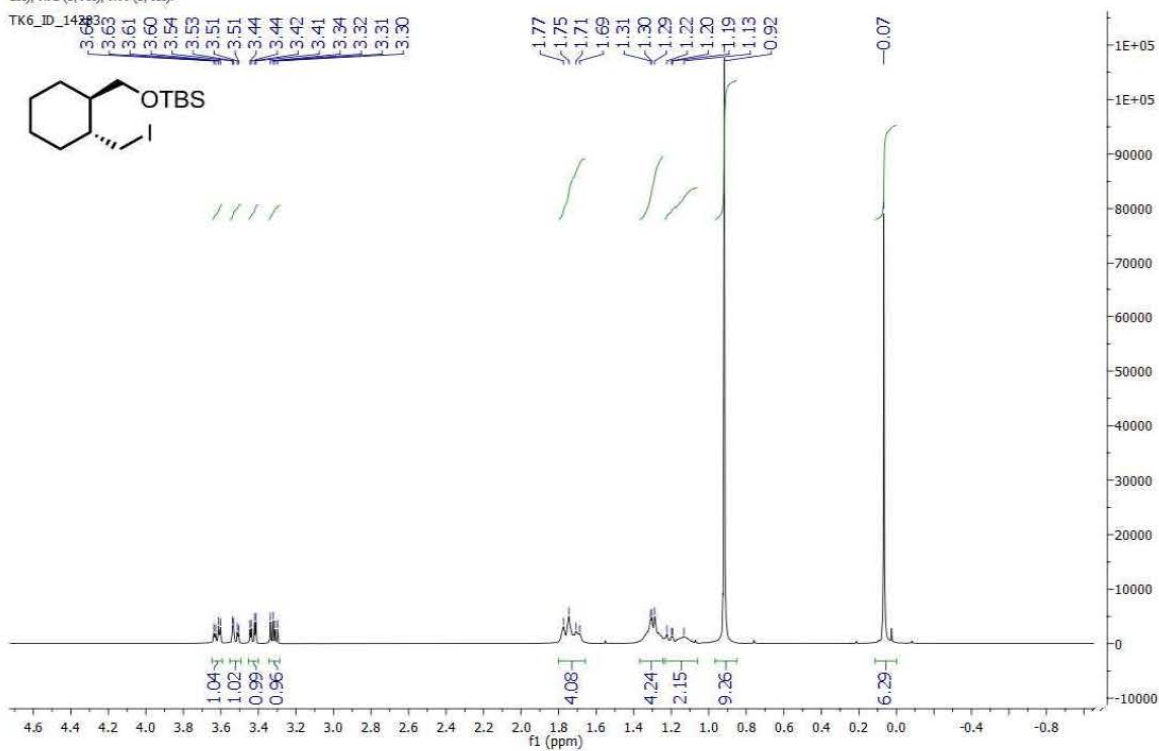


^{13}C NMR (101 MHz, CDCl_3) δ 68.59, 67.43, 45.52, 44.07, 30.07, 29.86, 26.16, 26.15, 25.81, 18.17, -5.45, -5.57.

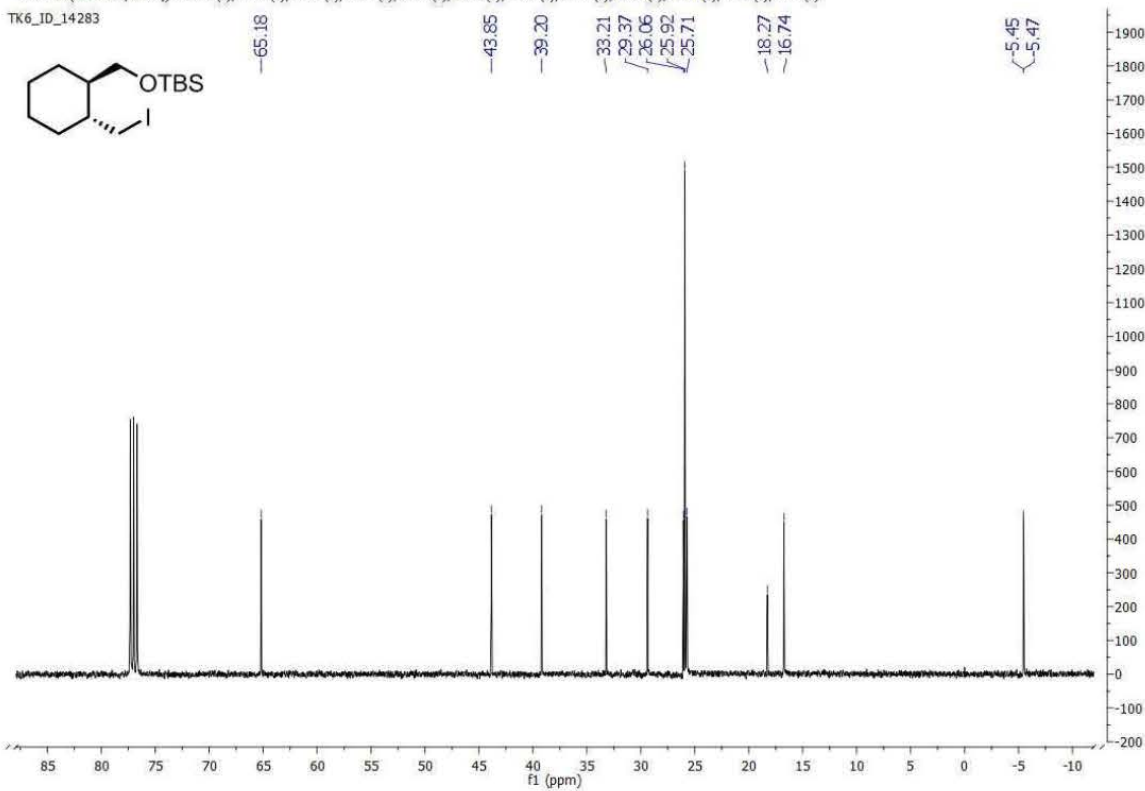


Ryc.79. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla związku **16**.

¹H NMR (400 MHz, CDCl₃) δ 3.62 (dd, *J* = 10.2, 4.2 Hz, 1H), 3.52 (dd, *J* = 10.2, 2.7 Hz, 1H), 3.45 (dd, *J* = 9.8, 2.6 Hz, 1H), 3.32 (dd, *J* = 9.8, 5.7 Hz, 1H), 1.80 – 1.66 (m, 4H), 1.37 – 1.24 (m, 4H), 1.23 – 1.06 (m, 2H), 0.92 (s, 9H), 0.06 (s, 6H).

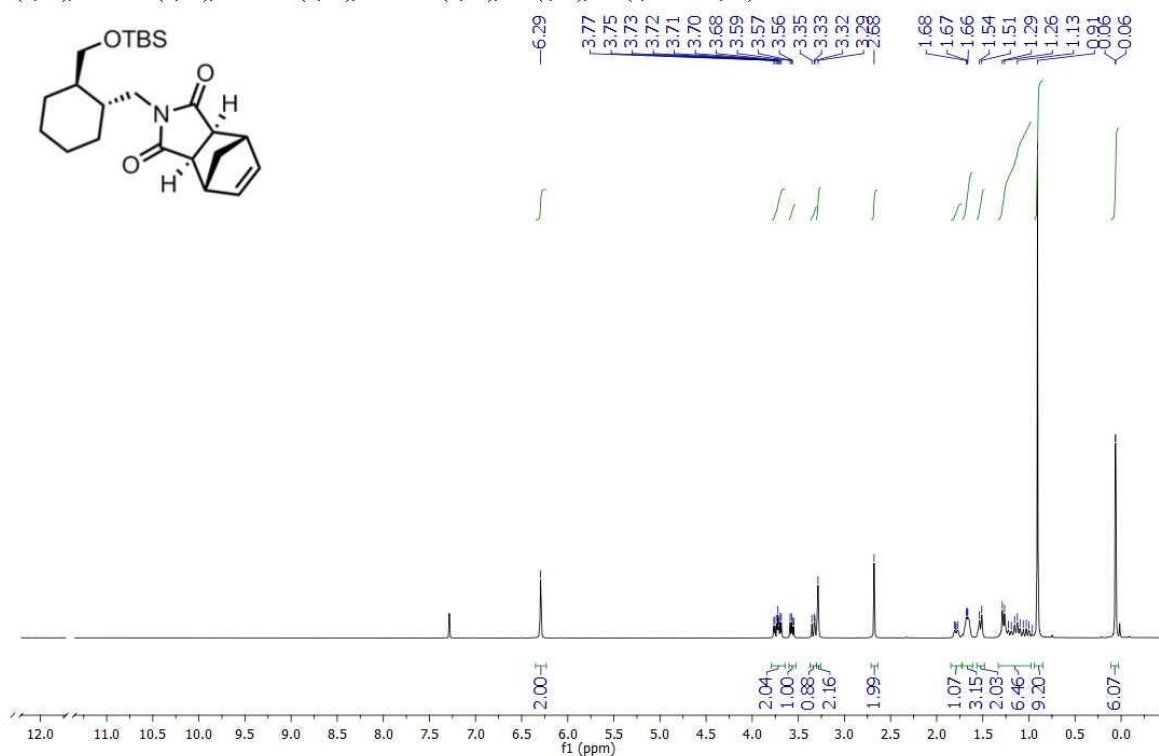


¹³C NMR (101 MHz, CDCl₃) δ 65.18 (s), 43.85 (s), 39.20 (s), 33.21 (s), 29.37 (s), 26.06 (s), 25.92 (s), 25.71 (s), 18.27 (s), 16.74 (s), -5.45 (s), -5.47 (s).

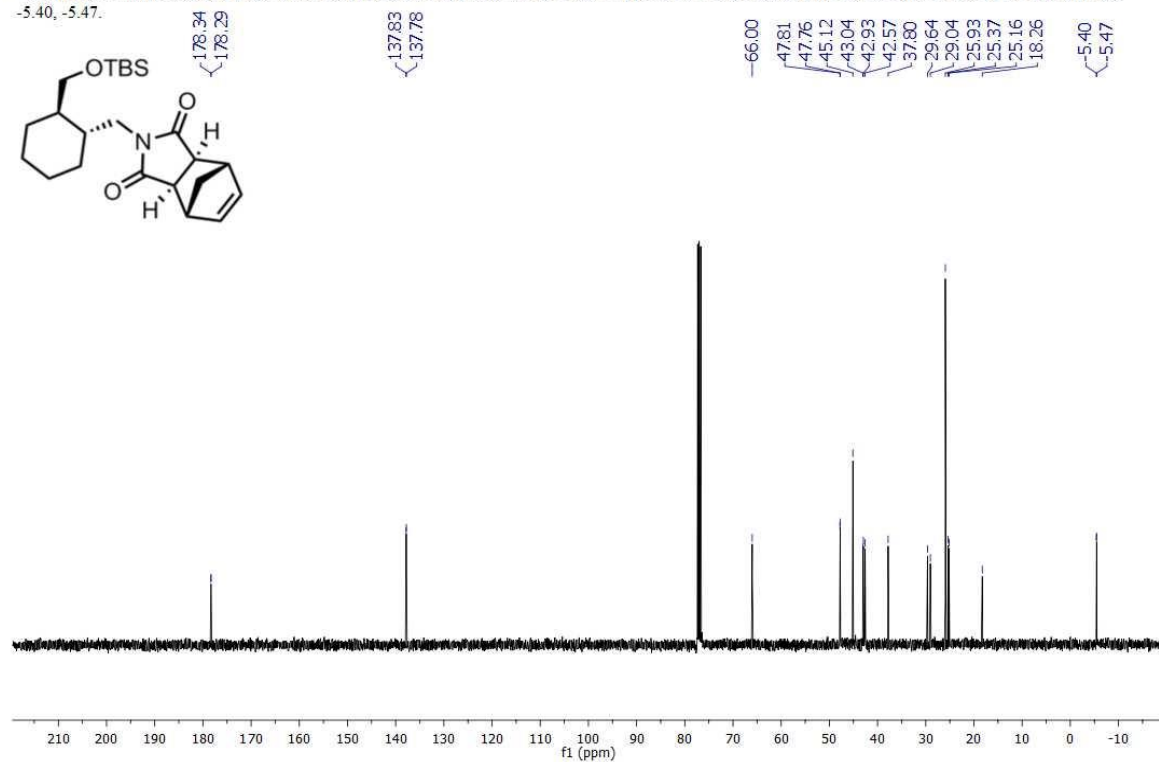


Ryc.80. Widma ¹H NMR (górá), ¹³C NMR (dól), dla zwiázkú 17.

^1H NMR (400 MHz, CDCl_3) δ 6.29 (s, 2H), 3.79 – 3.64 (m, 2H), 3.57 (dd, $J = 10.1, 5.7$ Hz, 1H), 3.33 (dd, $J = 11.9, 8.4$ Hz, 1H), 3.29 (s, 2H), 2.68 (s, 2H), 1.85 – 1.73 (m, 1H), 1.72 – 1.61 (m, 3H), 1.57 – 1.48 (m, 2H), 1.33 – 0.98 (m, 6H), 0.91 (s, 9H), 0.06 (d, $J = 2.0$ Hz, 6H).

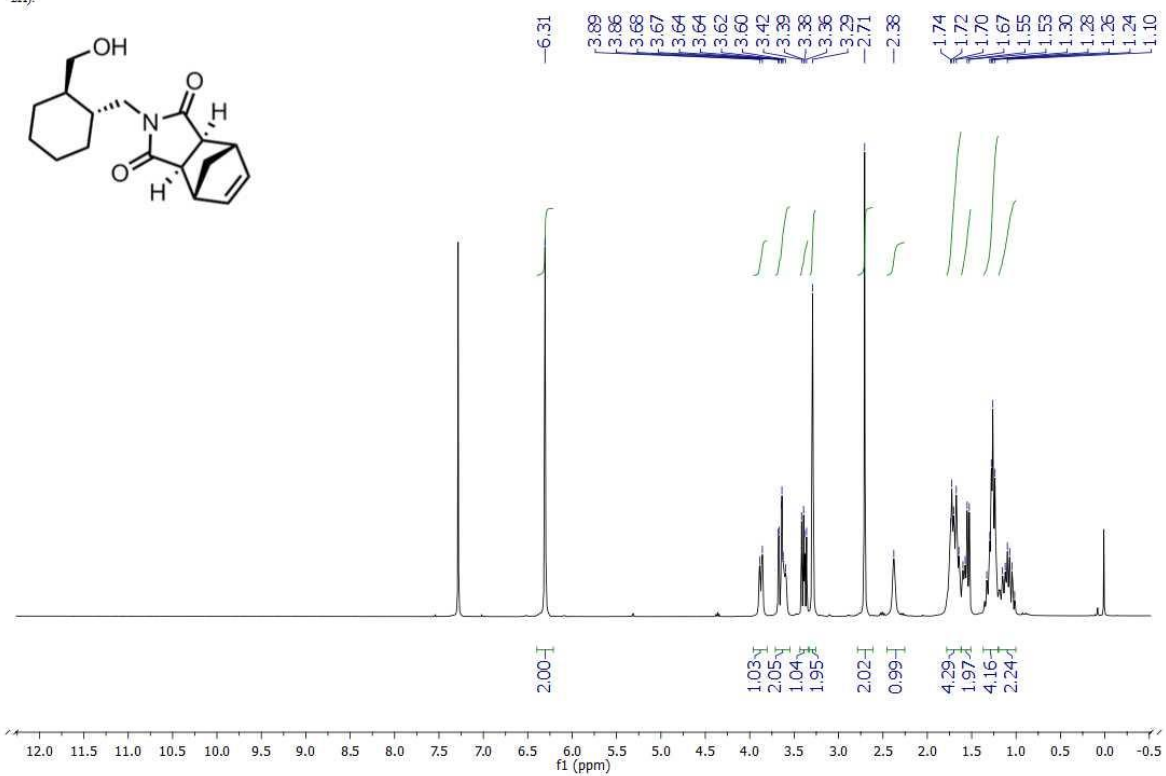


^{13}C NMR (101 MHz, CDCl_3) δ 178.34, 178.29, 137.83, 137.78, 66.00, 47.81, 47.76, 45.12, 43.04, 42.93, 42.57, 37.80, 29.64, 29.04, 25.93, 25.37, 25.16, 18.26, -5.40, -5.47.

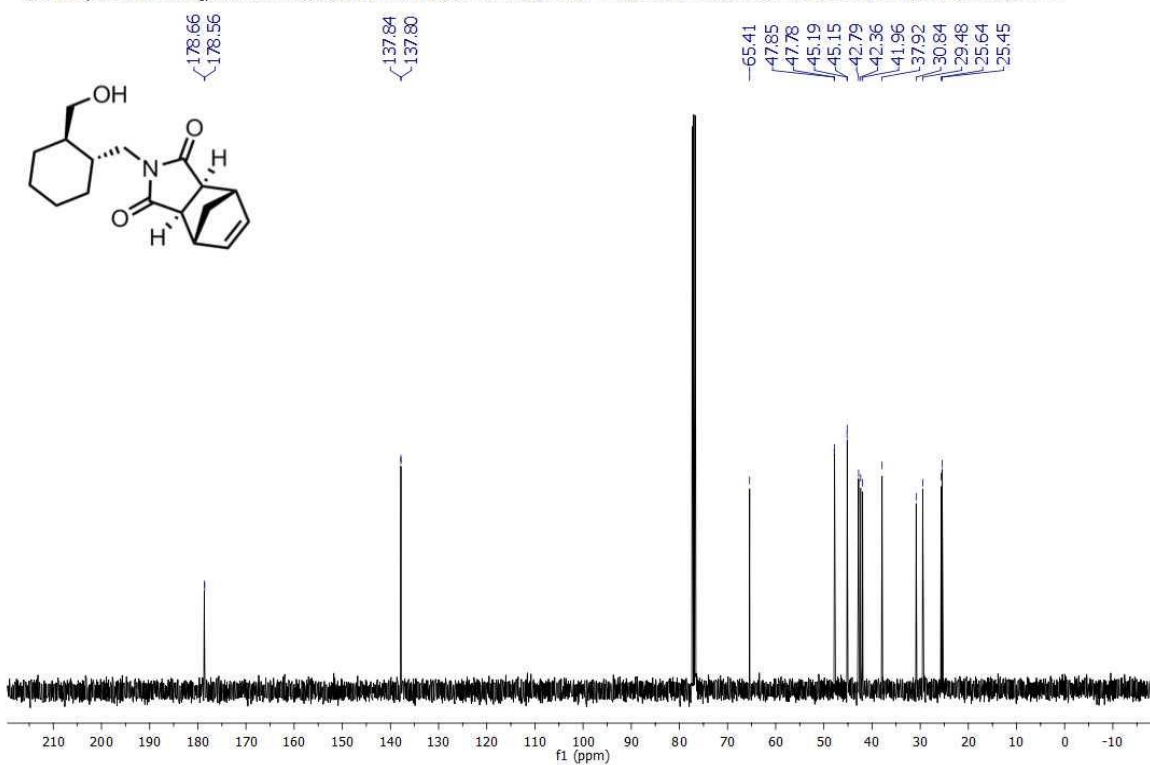


Ryc.81. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla związku **18**.

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 6.31 (s, 2H), 3.96–3.80 (m, 1H), 3.71–3.55 (m, 2H), 3.39 (dd, 1H), 3.29 (s, 2H), 2.71 (s, 2H), 2.45–2.25 (s, 1H), 1.80–1.62 (m, 4H), 1.56 (m, 2H), 1.37–1.20 (m, 4H), 1.08 (m, 2H).

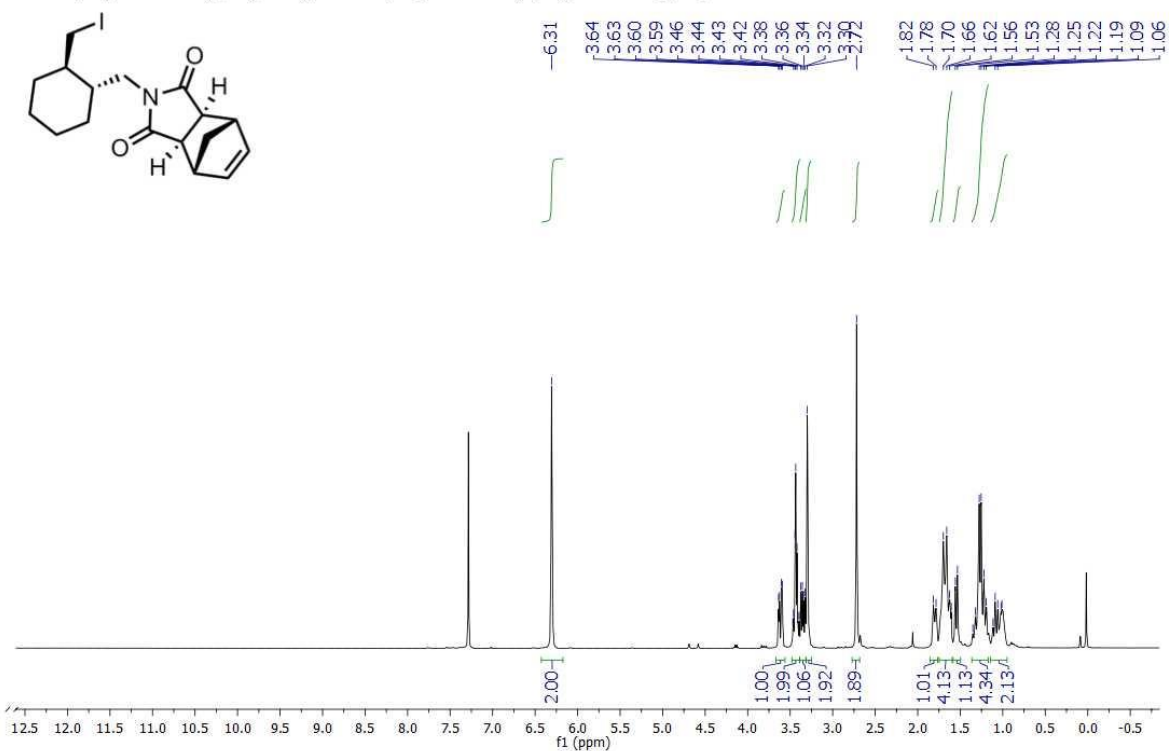


$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 178.66, 178.56, 137.84, 137.80, 65.41, 47.85, 47.78, 45.19, 45.15, 42.79, 42.36, 41.96, 37.92, 30.84, 29.48, 25.64, 25.45.

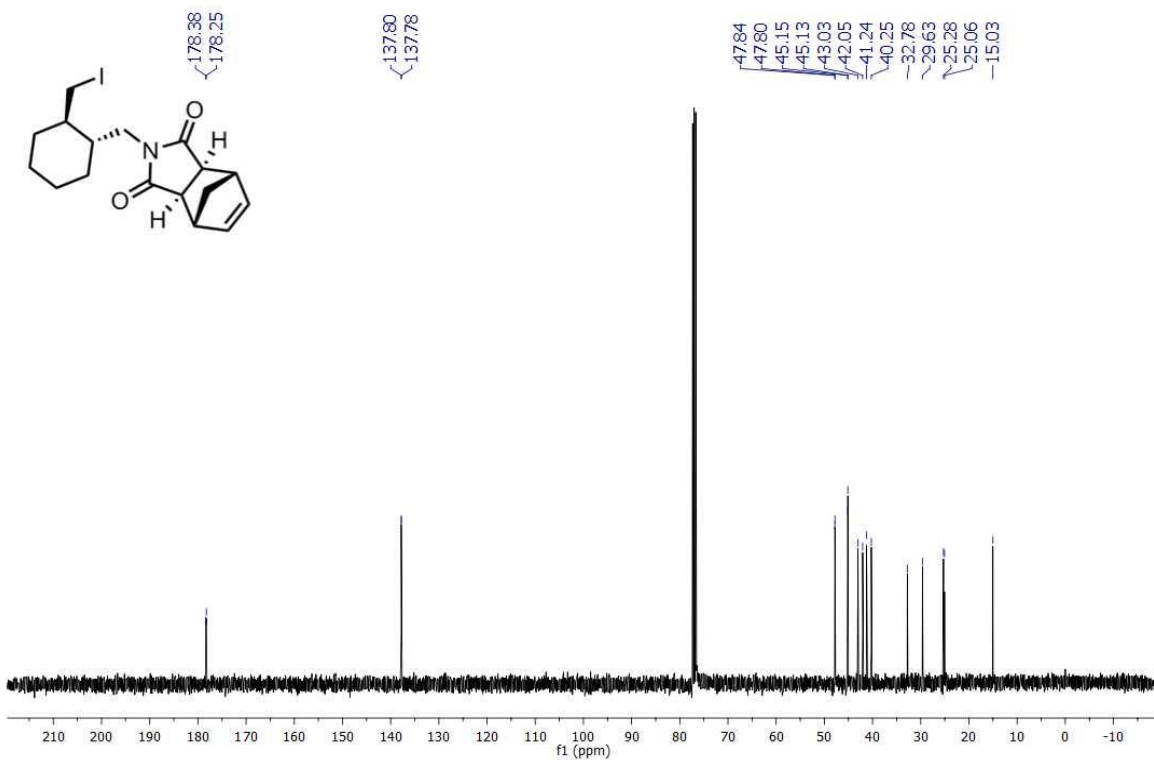


Ryc.82. Widma $^1\text{H NMR}$ (góra), $^{13}\text{C NMR}$ (dół), dla związku 19.

^1H NMR (400 MHz, CDCl_3) δ 6.31 (s, 1H), 3.61 (dd, $J = 13.4, 4.6$ Hz, 1H), 3.48 – 3.39 (m, 1H), 3.38 – 3.31 (m, 1H), 3.30 (s, 1H), 2.72 (s, $J = 17.0$ Hz, 1H), 1.80 (d, $J = 12.3$ Hz, 1H), 1.75 – 1.59 (m, 2H), 1.55 (d, $J = 9.8$ Hz, 1H), 1.37 – 1.17 (m, 2H), 1.13 – 0.95 (m, 1H).

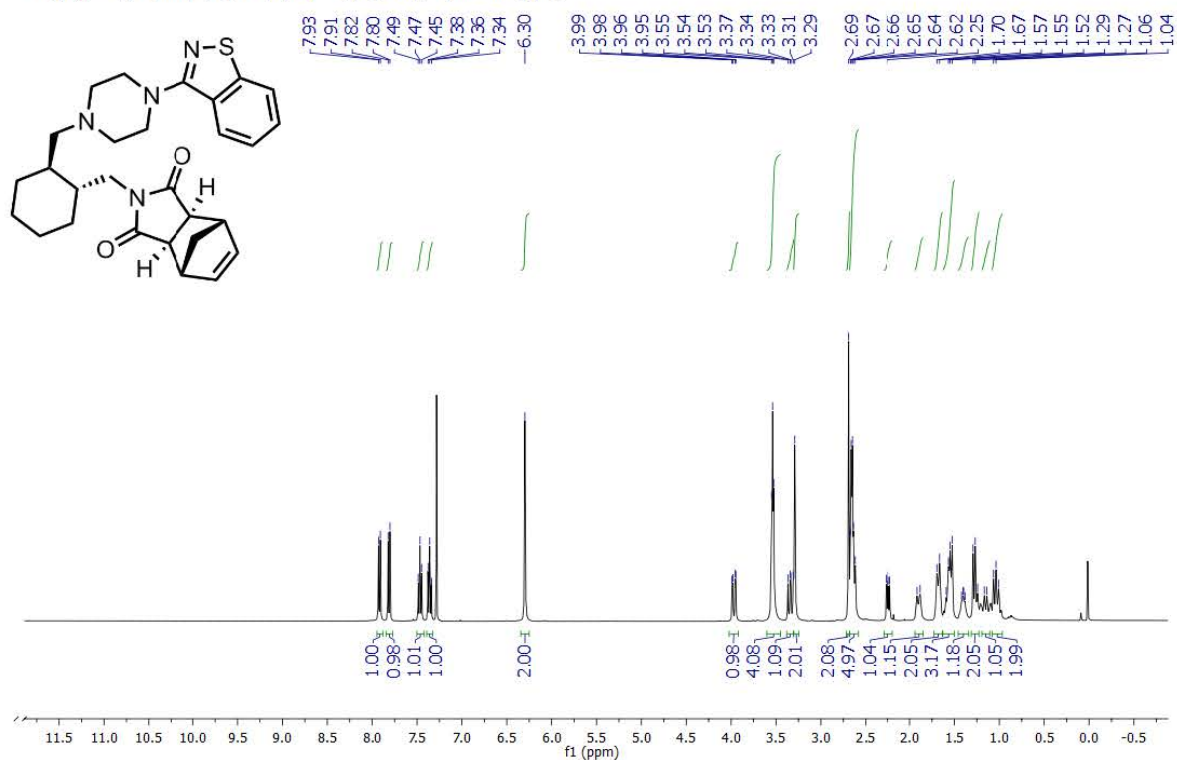


^{13}C NMR (101 MHz, CDCl_3) δ 178.38, 178.25, 137.80, 137.78, 47.84, 47.80, 45.15, 45.13, 43.03, 42.05, 41.24, 40.25, 32.78, 29.63, 25.28, 25.06, 15.03.

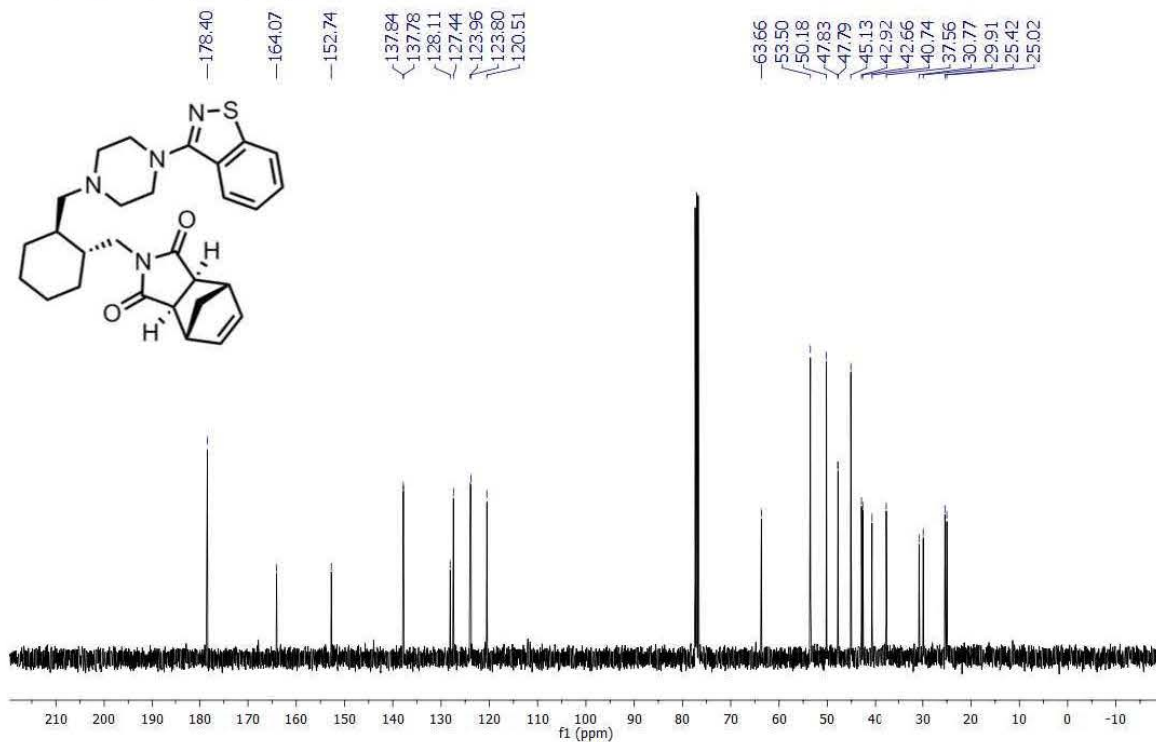


Ryc.83. Widma ^1H NMR (górá), ^{13}C NMR (dóól), dla zwiázku **20**.

¹H NMR (400 MHz, CDCl₃) δ 7.92 (d, 1H), 7.81 (d, 1H), 7.47 (t, 1H), 7.36 (t, 1H), 6.30 (s, 2H), 3.97 (dd, 1H), 3.54 (t, 4H), 3.34 (dd, 1H), 3.29 (s, 2H), 2.69 (s, 2H), 2.68 – 2.58 (m, 5H), 2.25 (dd, J = 12.5, 1H), 1.90 (d, 1H), 1.68 (d, 2H), 1.56 (dd, 3H), 1.40 (dd, 1H), 1.27 (t, 2H), 1.15 (d, 1H), 1.08 – 0.97 (m, 2H).

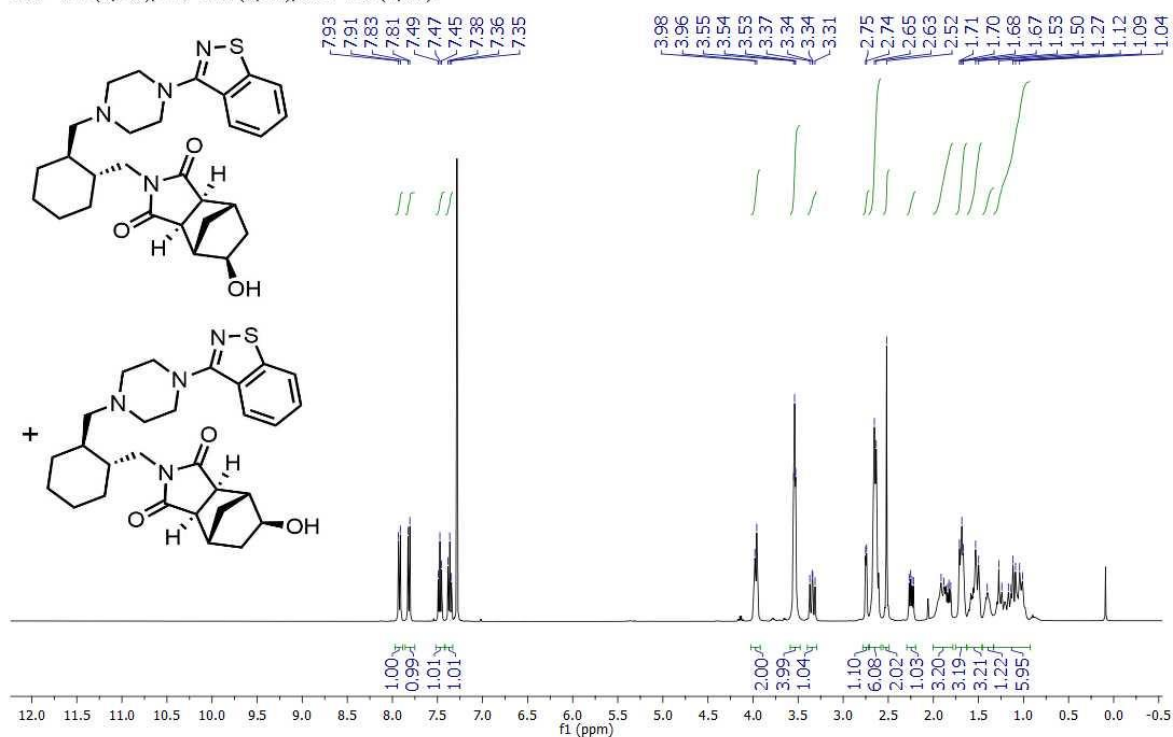


¹³C NMR (101 MHz, CDCl₃) δ 178.40, 164.07, 152.74, 137.84, 137.78, 128.11, 127.44, 123.96, 123.80, 120.51, 63.66, 53.50, 50.18, 47.83, 47.79, 45.13, 42.92, 42.66, 40.74, 37.56, 30.77, 29.91, 25.42, 25.02.

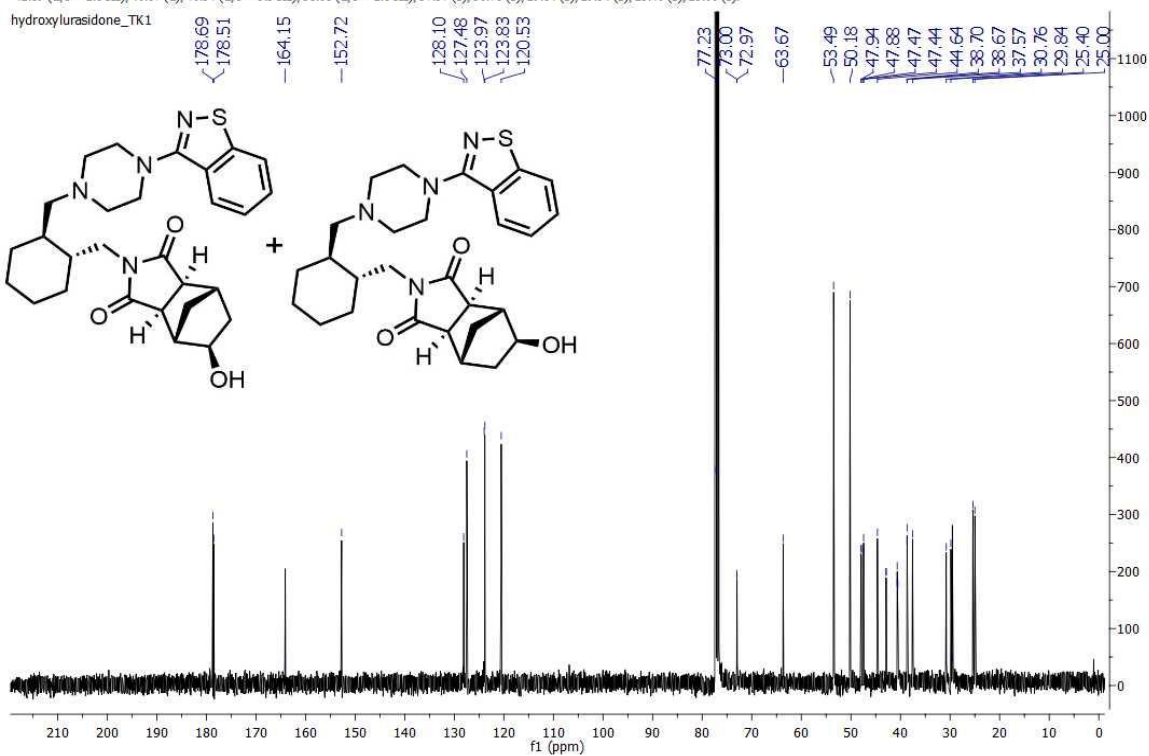


Ryc.84. Widma ¹H NMR (góra), ¹³C NMR (dół), dla związku 22.

^1H NMR (400 MHz, CDCl_3) δ 7.92 (d, $J = 8.2$ Hz, 1H), 7.82 (d, $J = 8.1$ Hz, 1H), 7.47 (t, $J = 7.5$ Hz, 1H), 7.36 (t, $J = 7.6$ Hz, 1H), 3.97 (m, 2H), 3.54 (m, 4H), 3.34 (dd, $J = 13.1, 10.3$ Hz, 1H), 2.75 (d, $J = 4.0$ Hz, 1H), 2.64 (m, 6H), 2.52 (s, 2H), 2.24 (dd, $J = 12.5, 6.5$ Hz, 1H), 1.99–1.77 (m, 3H), 1.75–1.63 (m, 3H), 1.60–1.45 (m, 3H), 1.45–1.35 (m, 1H), 1.32–0.95 (m, 6H).



^{13}C NMR (101 MHz, CDCl_3) δ 178.68, 178.51, 164.09, 152.72, 128.10, 127.48, 123.97, 120.53, 77.23, 72.97 (d), 63.67, 53.49, 50.18, 47.91 (d, $J = 6.1$ Hz), 47.45 (d, $J = 2.9$ Hz), 44.66 (d, $J = 4.8$ Hz), 42.89 (d, $J = 2.8$ Hz), 40.67 (d), 40.54 (d, $J = 5.3$ Hz), 38.68 (d, $J = 2.8$ Hz), 37.57 (s), 30.76 (s), 29.84 (s), 29.54 (s), 25.40 (s), 25.00 (s).



Ryc.85. Widma ^1H NMR (góra), ^{13}C NMR (dół), dla mieszaniny 12,12'.

TK1_ID-14283

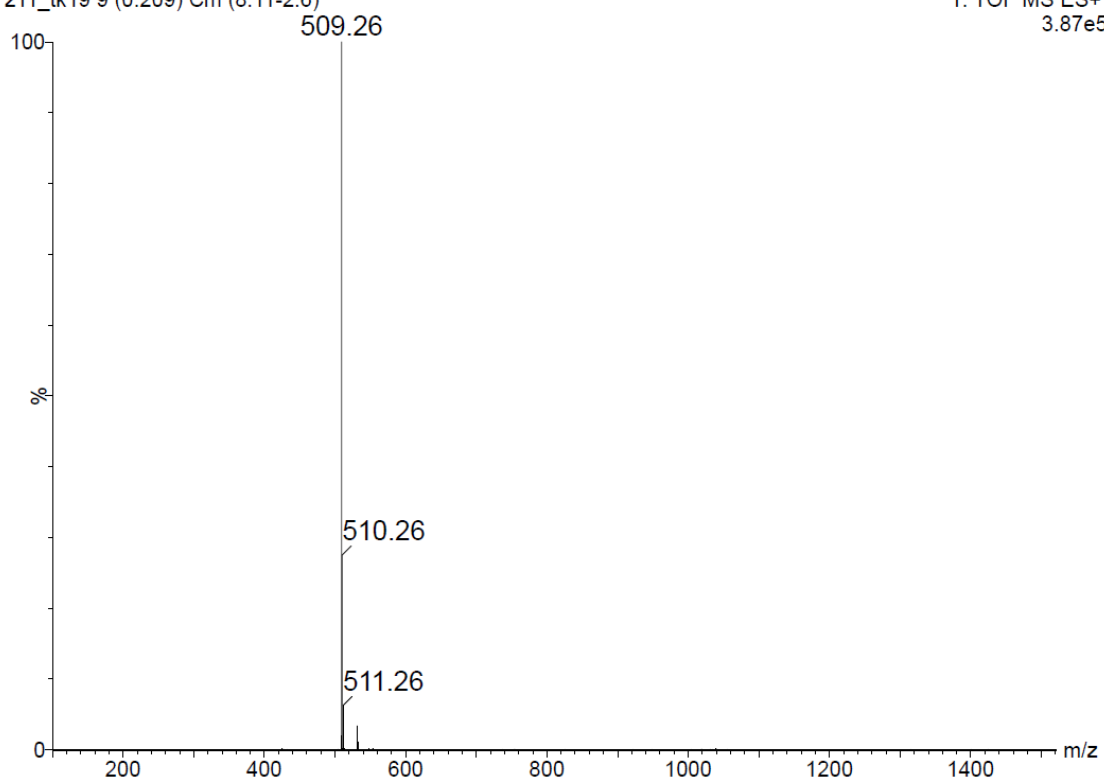
z11_tk19 9 (0.209) Cm (8:11-2:6)

05-Jan-2017

11:03:35

1: TOF MS ES+

3.87e5



Ryc.86. Spektrum masowe uzyskane metodą ESI-MS dla **12,12'**.

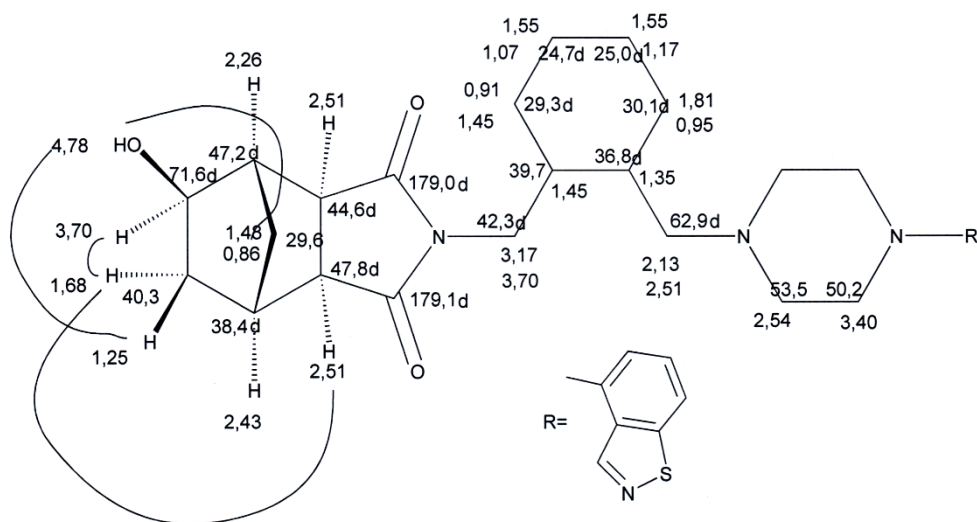
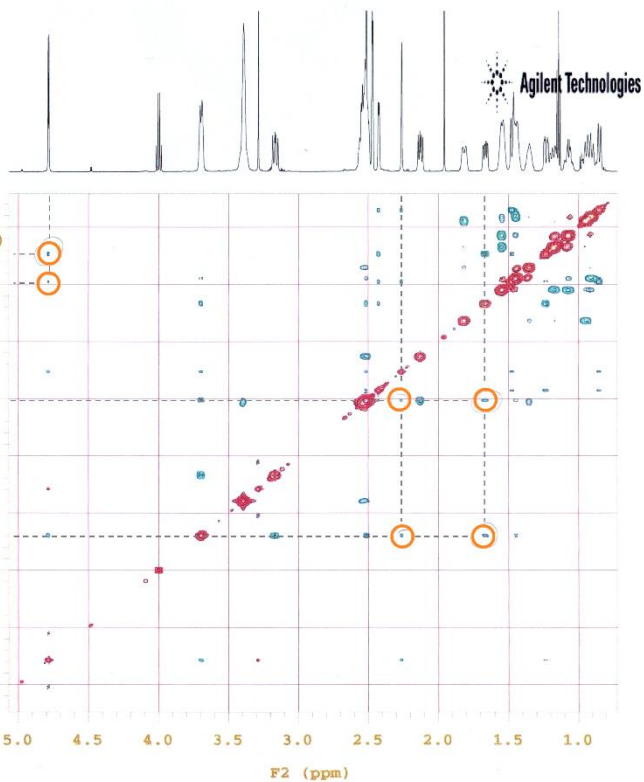
T. Klucznik
resp11/Var600/TK1-ID-14283-DMSO/TK1-ID-14283-DMSO-ROESY

Sample Name:
TK1-ID-14283-DMSO
Data Collected on:
Varian-NMR-vnmrs600
Archive directory:

Sample directory:

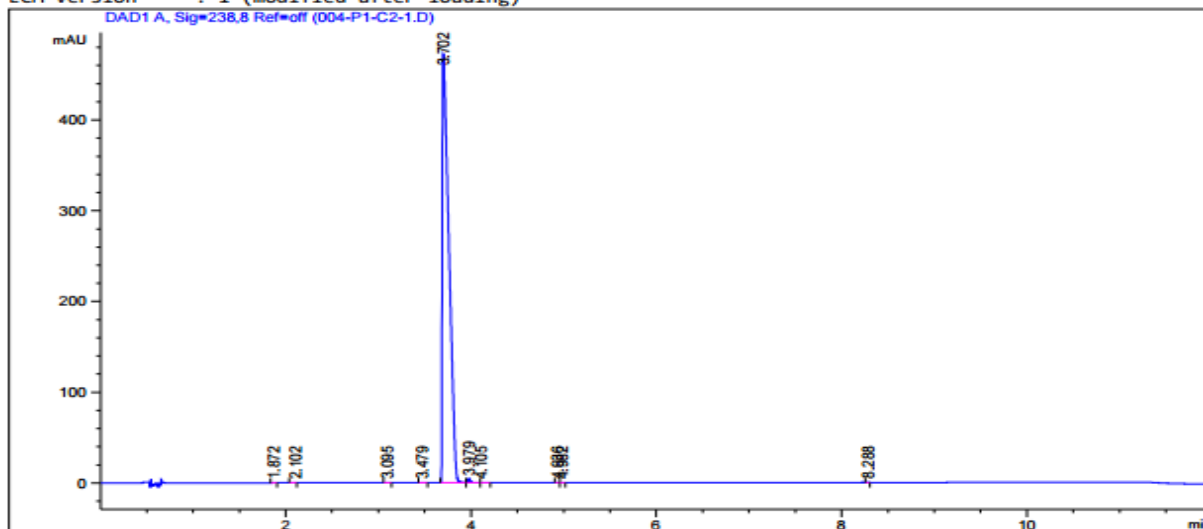
FidFile: TK1-ID-14283-DMSO-ROESY

Pulse Sequence: ROESY
Solvent: dmsc
Data collected on: Jan 12 2017



Ryc.87. Widmo ROESY NMR dla **12,12'**, określające m.in. położenie grupy hydroksylowej.

ECM Path : \LAB\HPLC\LC 20\2017\RMH-240 P LC 20 2017-05-24 01-39-51.SC.SSIzip
 ECM Version : 1 (modified after loading)



=====
 Area Percent Report
 =====

Sorted By : Signal
 Multiplier : 1.0000
 Dilution : 1.0000
 Do not use Multiplier & Dilution Factor with ISTDs

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	1.872	BV	0.0292	2.11181e-1	1.16434e-1	9.043e-3
2	2.102	BV	0.0343	6.32651e-2	2.28509e-2	2.709e-3
3	3.095	BV	0.0281	7.55542e-2	3.43683e-2	3.235e-3
4	3.479	VB	0.0349	5.95381e-2	2.07777e-2	2.550e-3
5	3.702	BV	0.0710	2326.64355	472.24716	99.6308
6	3.979	W	0.0246	7.76837	4.62779	0.3327
7	4.105	VB	0.0330	1.48829e-1	5.78808e-2	6.373e-3
8	4.936	W	0.0258	7.41235e-2	4.05160e-2	3.174e-3
9	4.982	VB	0.0290	6.78387e-2	3.37477e-2	2.905e-3
10	8.288	BBA	0.0220	1.54068e-1	1.10750e-1	6.597e-3

Signal 1: DAD1 A, Sig=238,8 Ref=off

Column: Ascentis Express Phenyl Hexyl, 2.7 µm, 3.0 x 100 mm
 Mobile Phase: A Acetonitrile
 B 0.1% Ammonium acetate
 Gradient: Time (min) %A %B
 0.0 40 60
 8.0 80 20
 10.0 80 20
 10.1 40 60
 15.0 40 60
 Flow Rate: 0.7 mL/min
 Wavelength: 238 nm
 Temperature: 35 °C

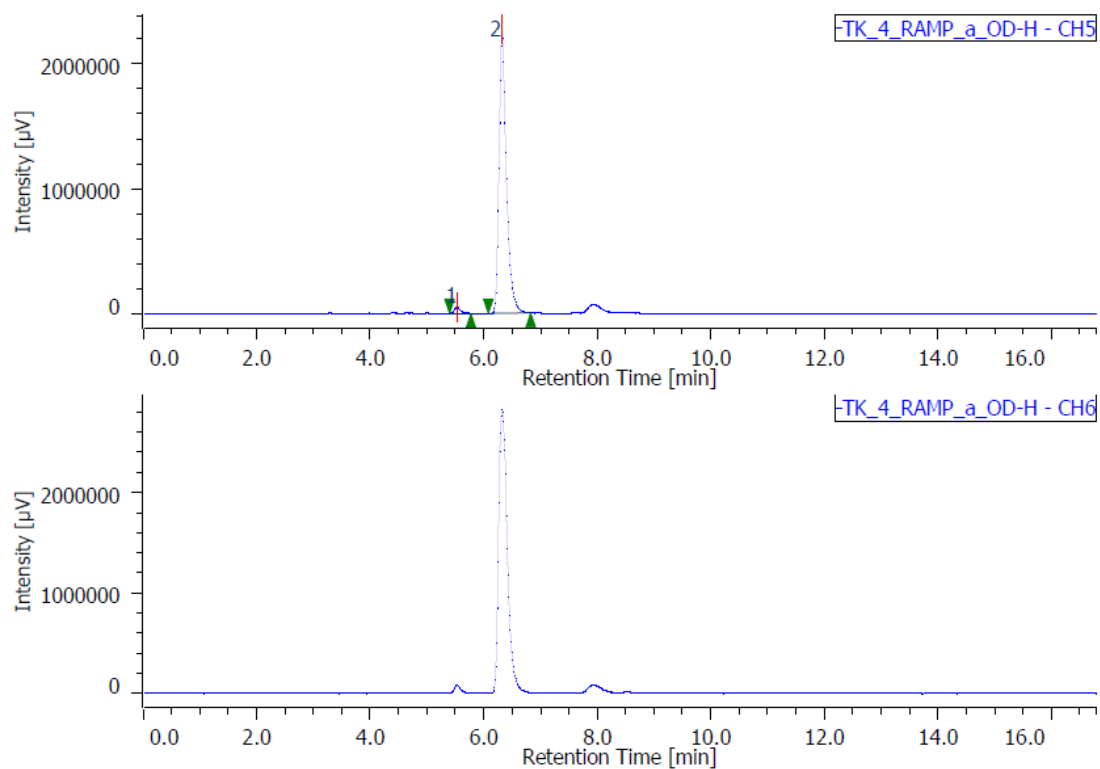
Totals : 2335.26633 477.31228

=====
 *** End of Report ***

Ryc.88. Chromatogram i wynik analizy HPLC, określający czystość uzyskanego **12,12'** na 99,64%.

TK_4_RAMP_a_OD-H-CH1

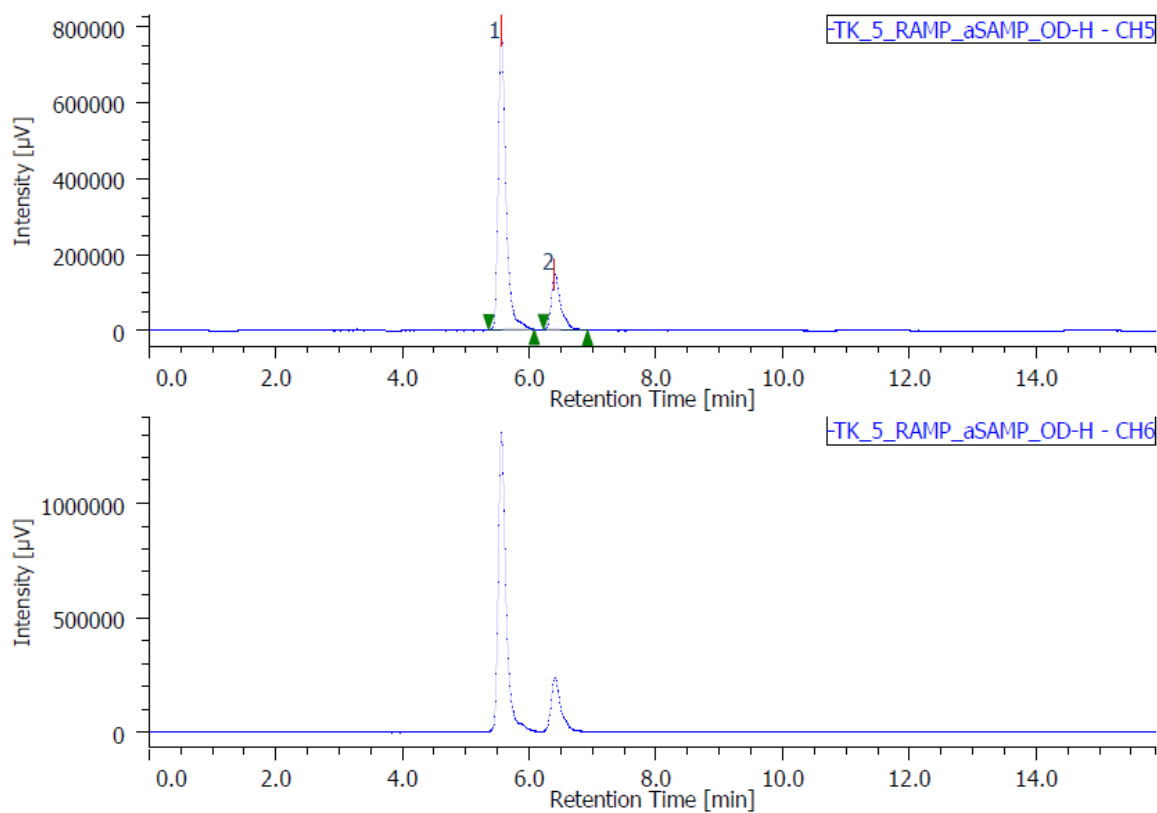
Chromatogram



#	Peak Name	CH	tR [min]	Area [µV·sec]	Height [µV]	Area%	Height%	Quantity	NTP	Resolution	Symmetry Factor	Warning
1	Unknown	5	5.533	370571	47193	1.785	2.037	N/A	12342	3.745	1.317	
2	Unknown	5	6.320	20389798	2269572	98.215	97.963	N/A	12967	N/A	1.468	

Ryc.89. Chromatogram związku **23(R,R)** (kolumna: OD-H Daicel Chiralcel, eluent: 2% *i*PrOH + 98% heksan, λ (detektora): 230 nm, przepływ: 1.0ml/min.).

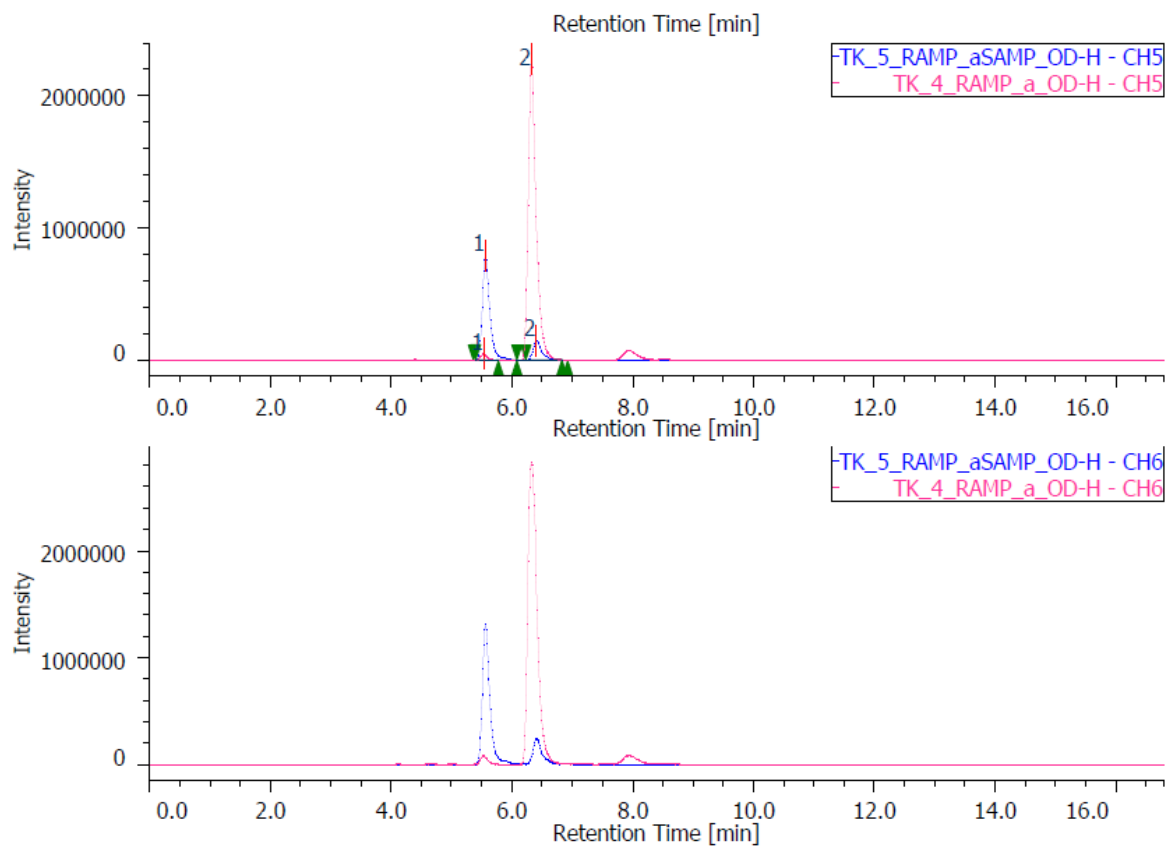
Chromatogram



#	Peak Name	CH	tR [min]	Area [μV·sec]	Height [μV]	Area%	Height%	Quantity	NTP	Resolution	Symmetry Factor	Warning
1	Unknown	5	5.560	6610629	786119	81.366	84.373	N/A	12002	3.745	1.463	
2	Unknown	5	6.400	1513969	145605	18.634	15.627	N/A	10756	N/A	1.577	

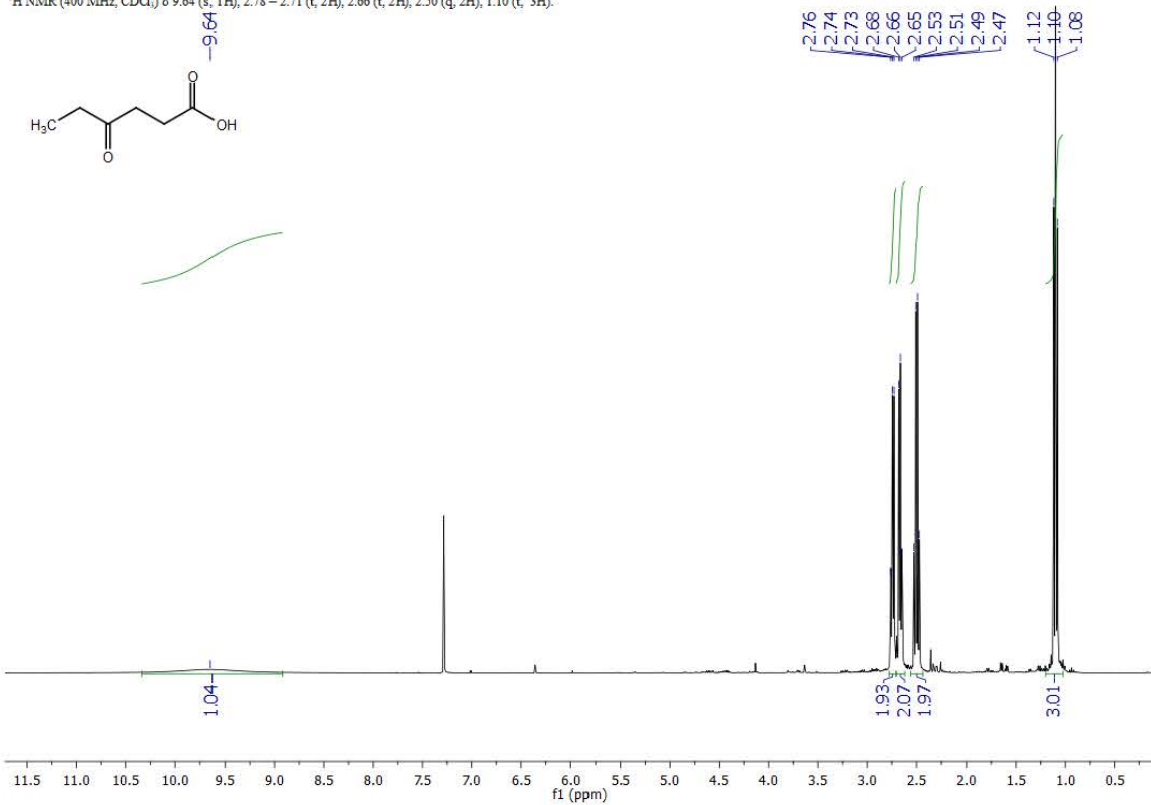
Ryc.90. Chromatogram związku **23(S,R)** (kolumna: OD-H Daicel Chiralcel, eluent: 2% *i*PrOH + 98% heksan, λ (detektora): 230 nm, przepływ: 1.0ml/min.).

Chromatogram

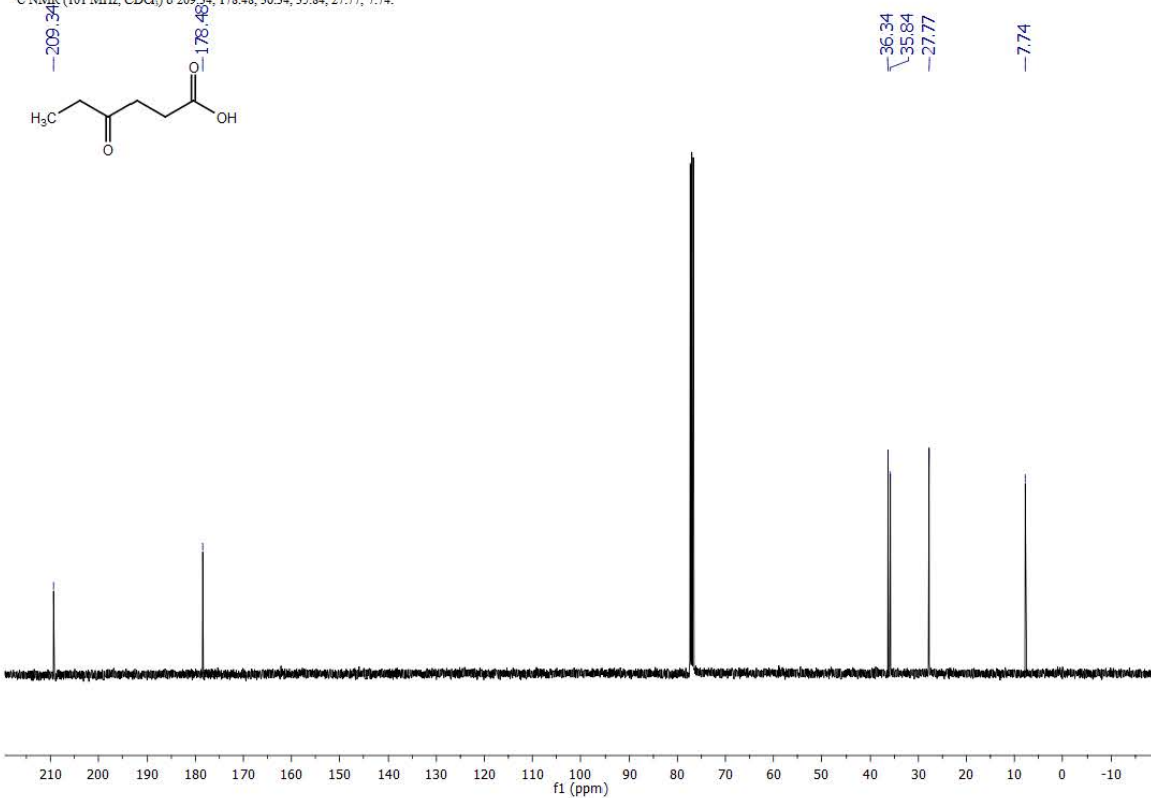


Ryc.91. Porównanie chromatogramów dla związków **23(R,R)**-(czerwony) i **23(S,R)**-(niebieski).

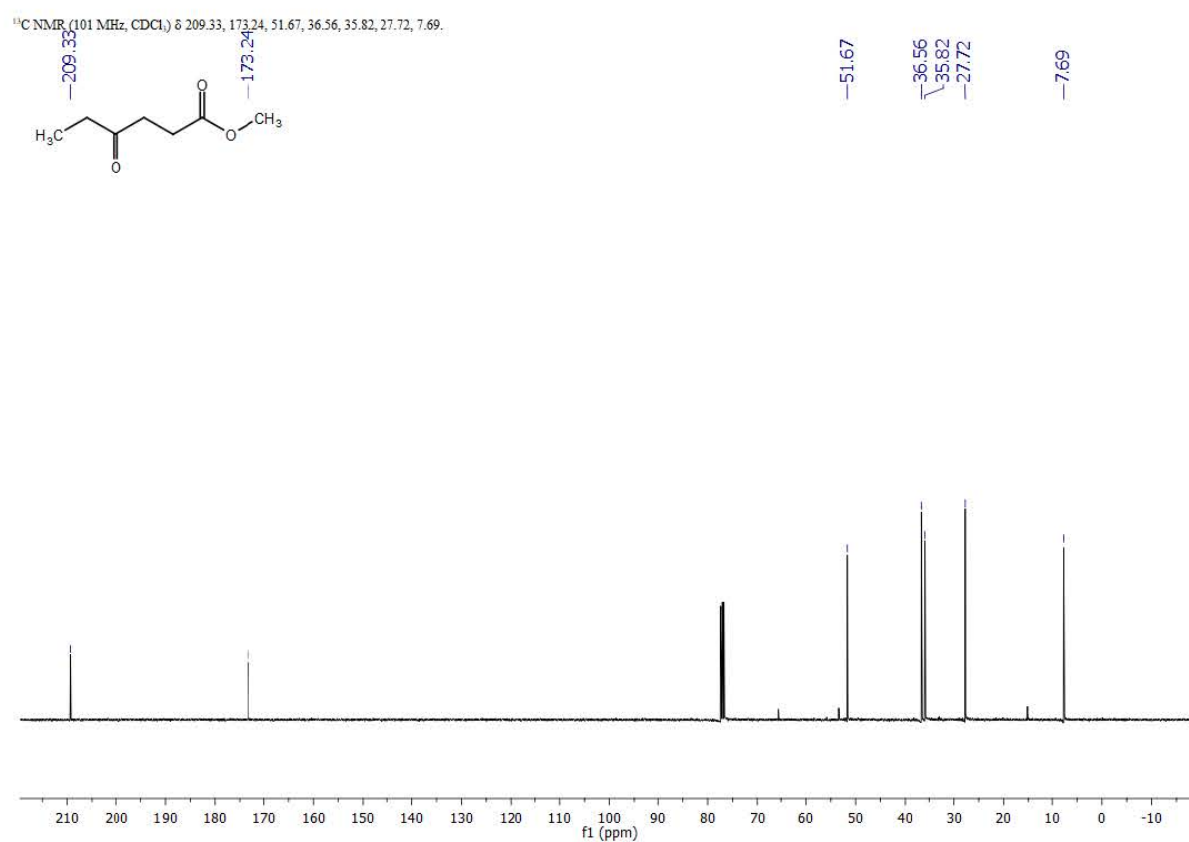
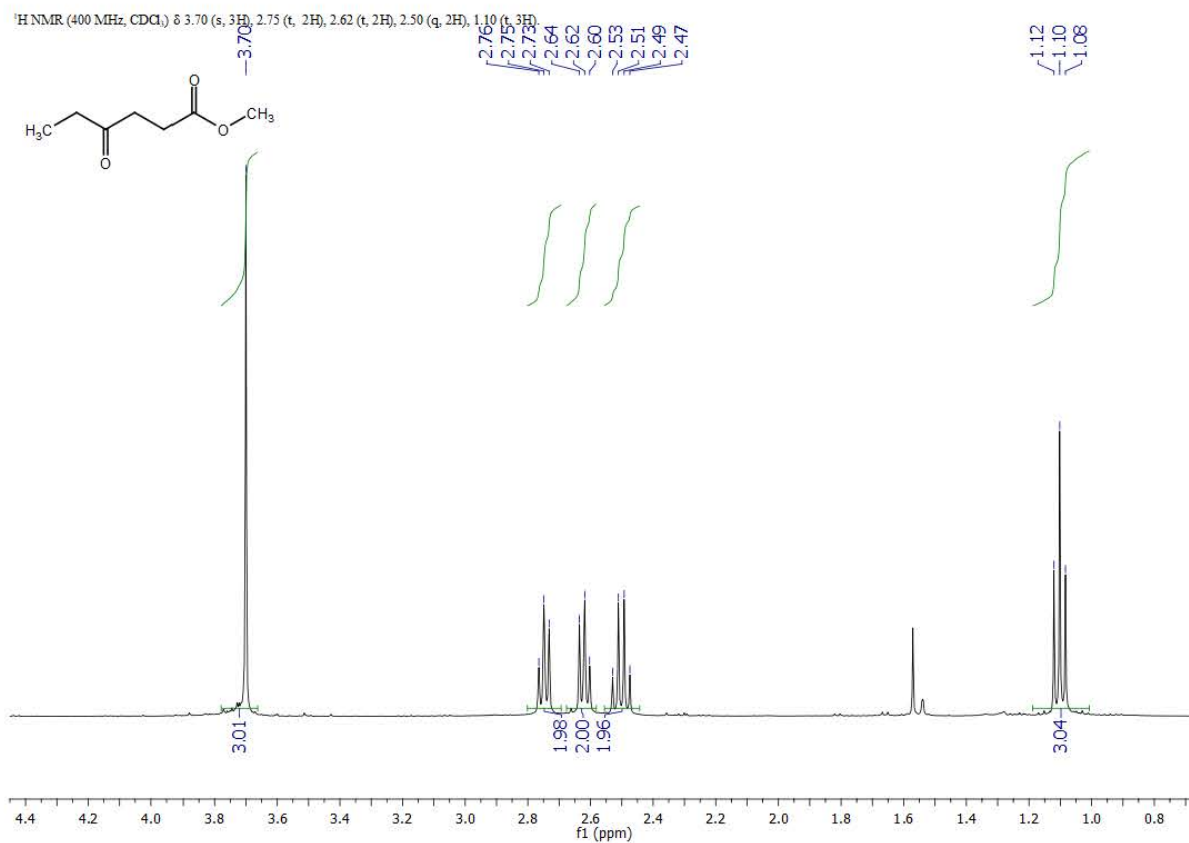
$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 9.64 (s, 1H), 2.78–2.71 (t, 2H), 2.66 (t, 2H), 2.50 (q, 2H), 1.10 (t, 3H).



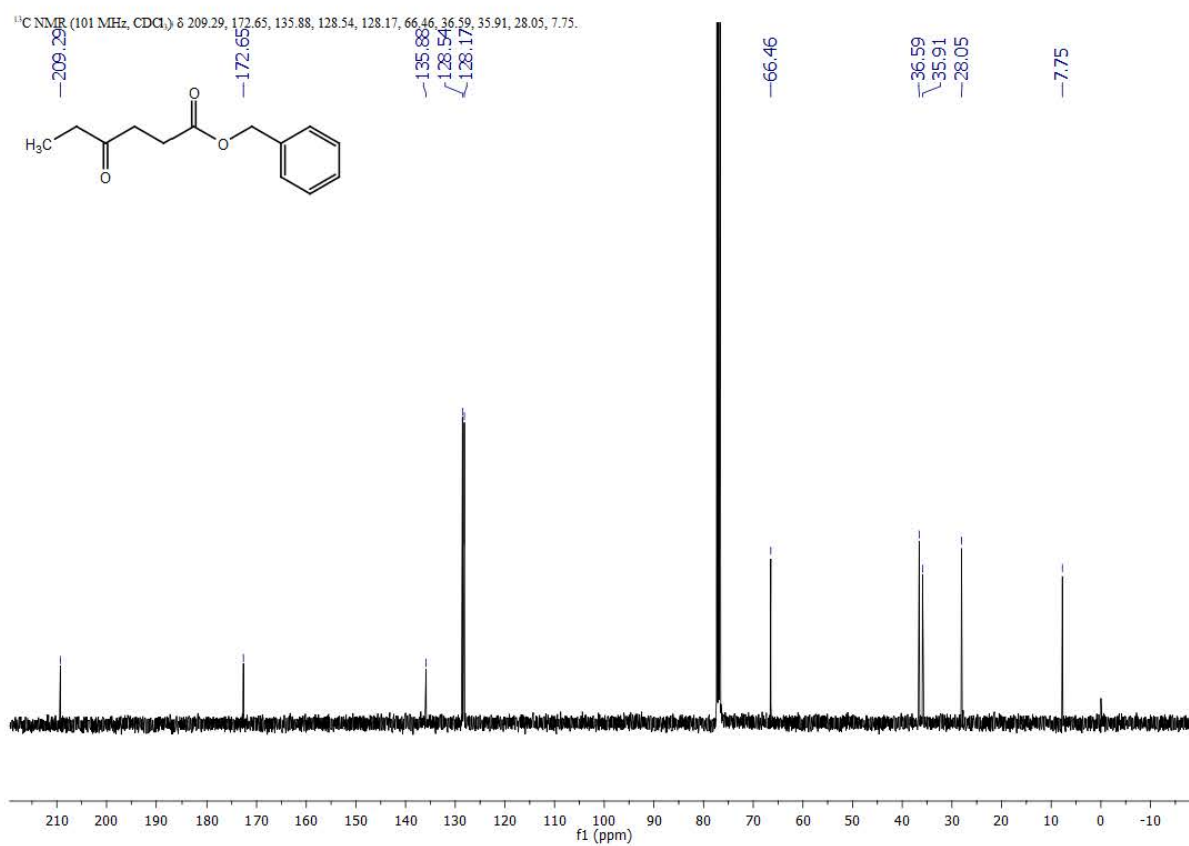
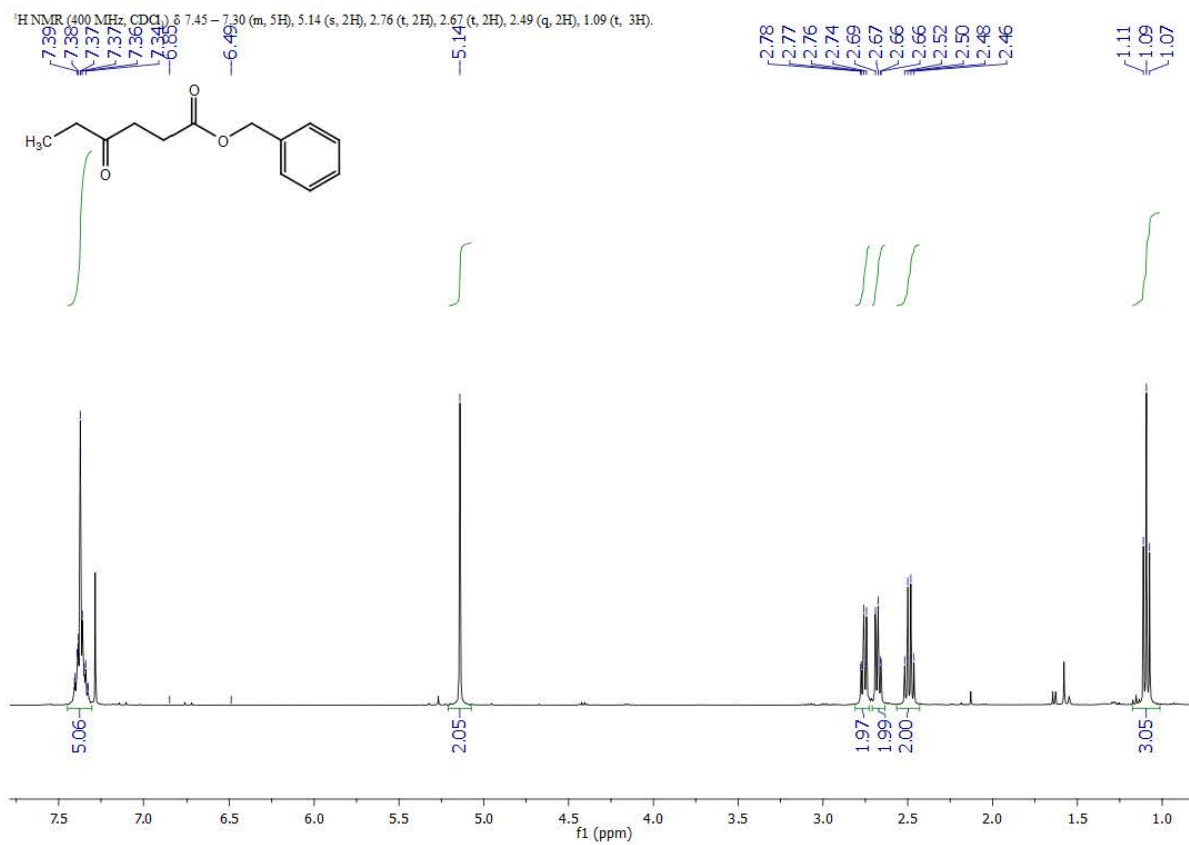
$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 209.34, 178.48, 36.34, 35.84, 27.77, 7.74.



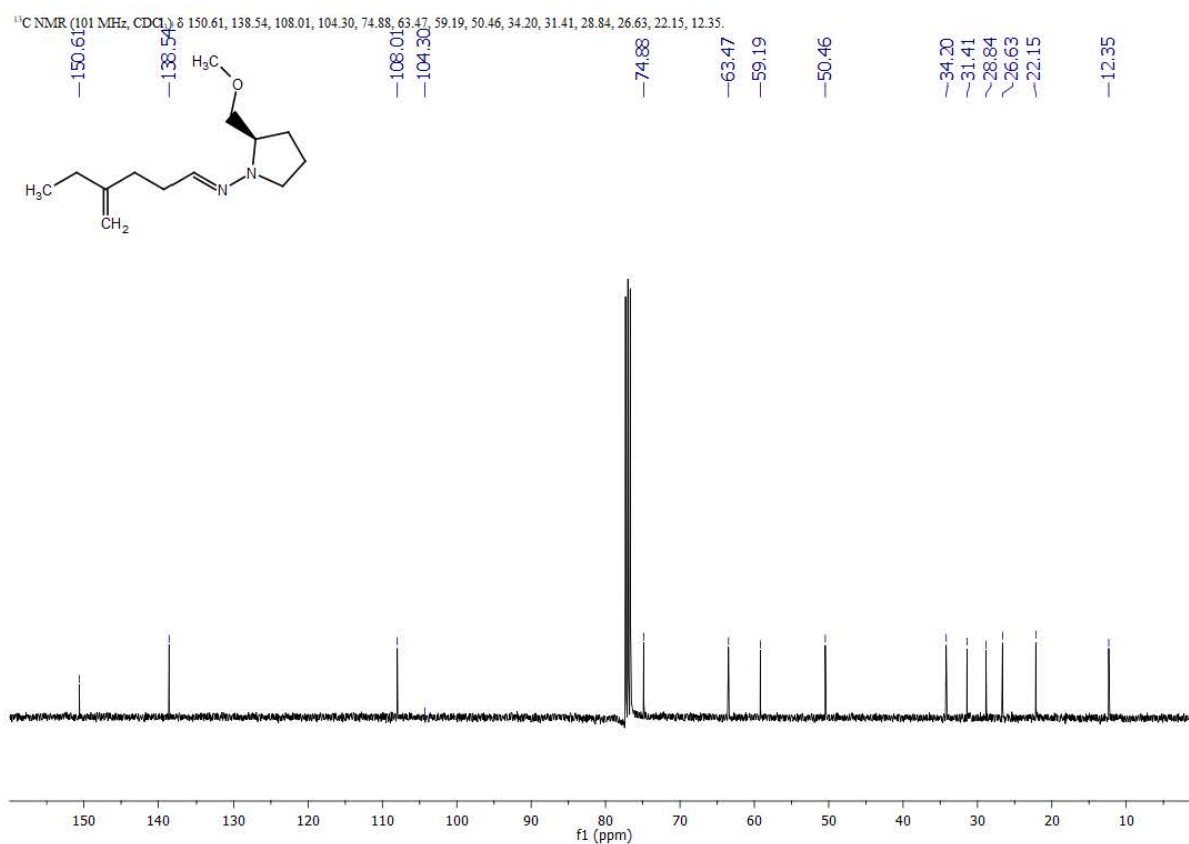
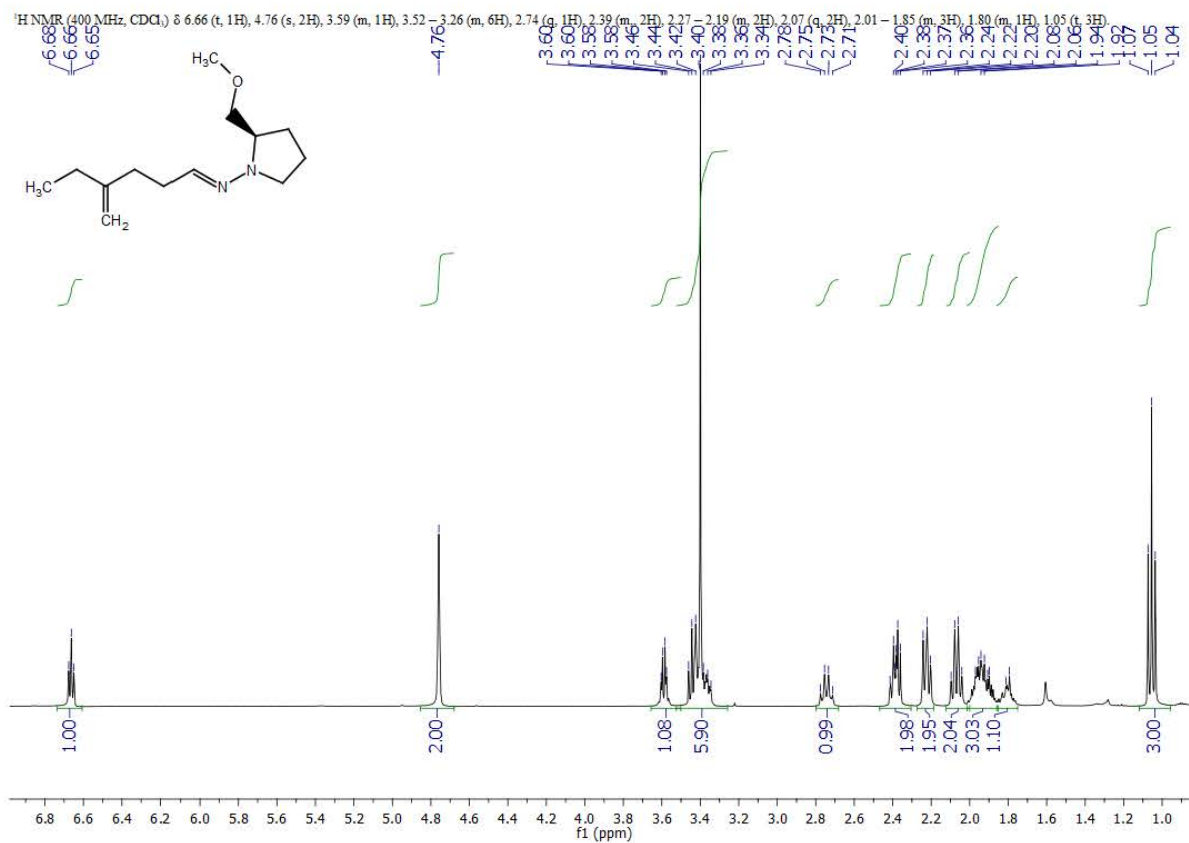
Ryc.92. Widma $^1\text{H NMR}$ (góra), $^{13}\text{C NMR}$ (dół), dla związku 25.



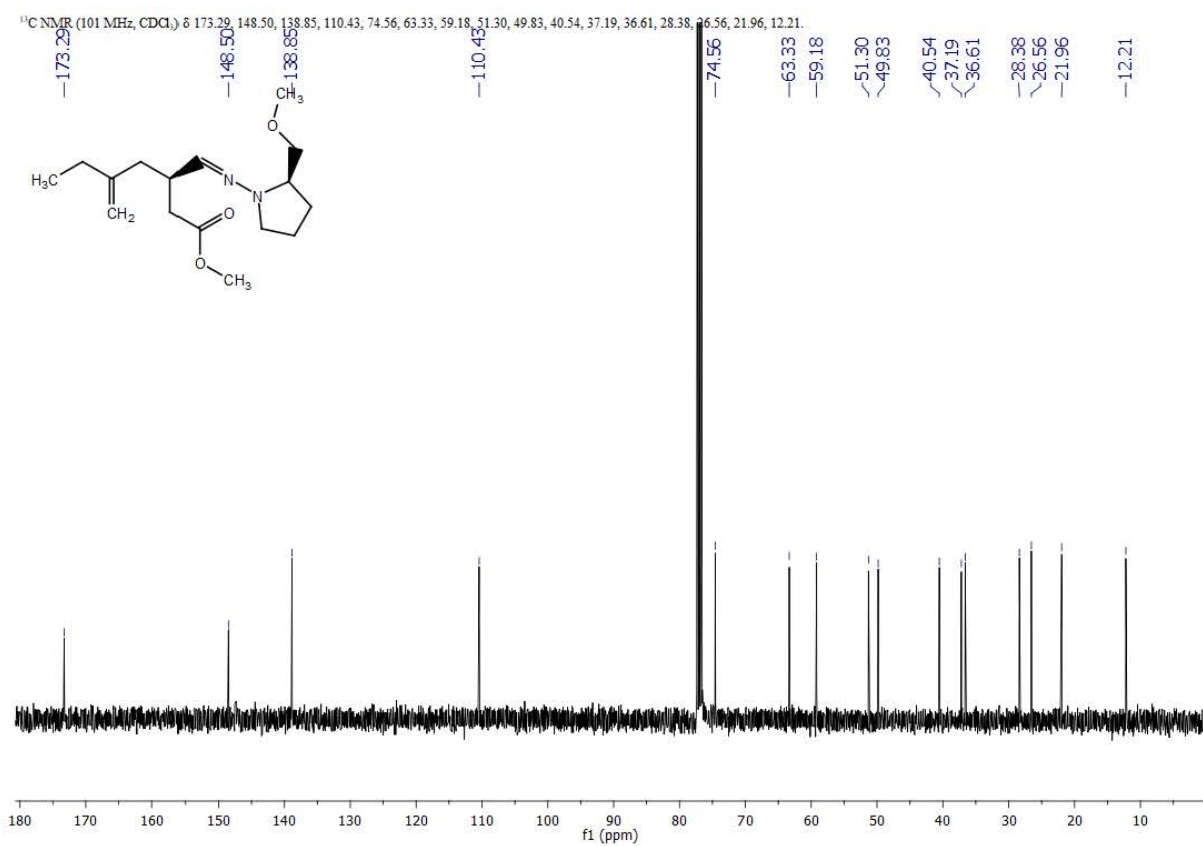
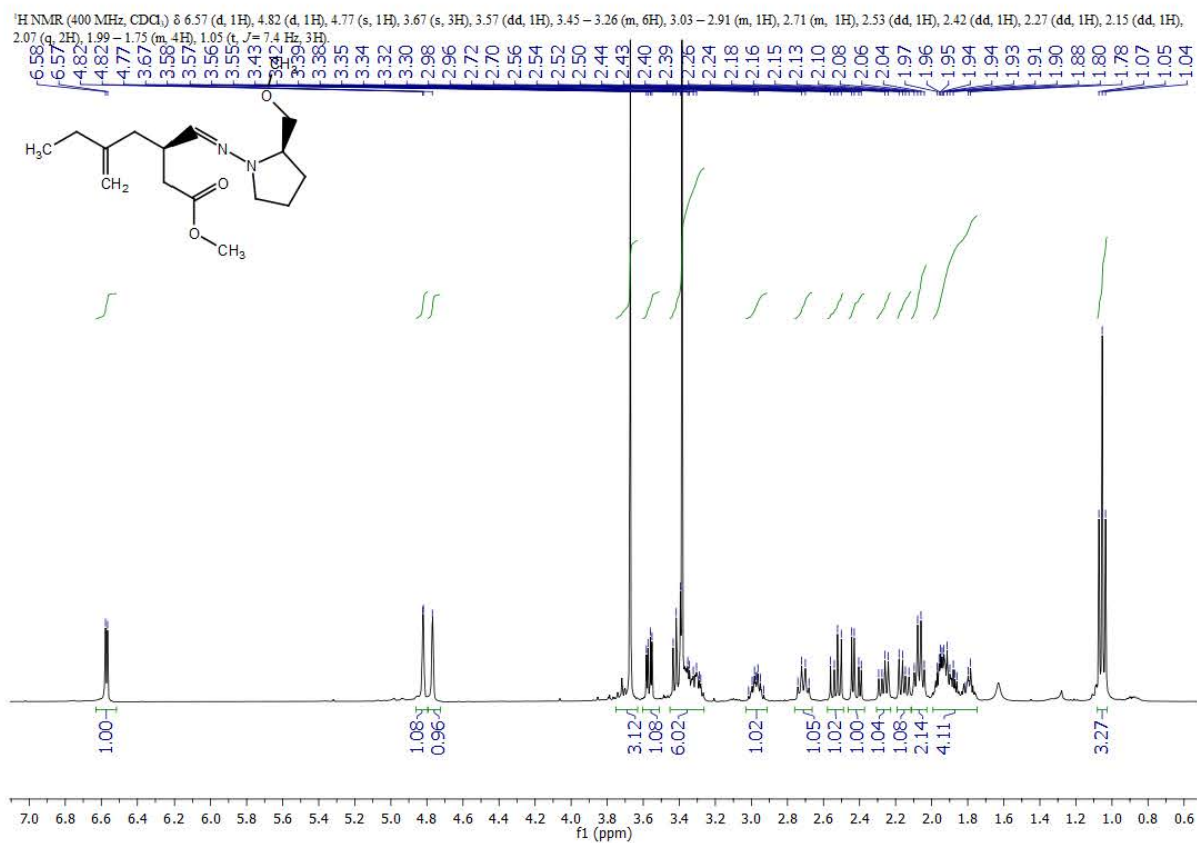
Ryc.93. Widma ¹H NMR (góra), ¹³C NMR (dół), dla związku 26.



Ryc.94. Widma ¹H NMR (górá), ¹³C NMR (dóř), dla zwizku **29**.

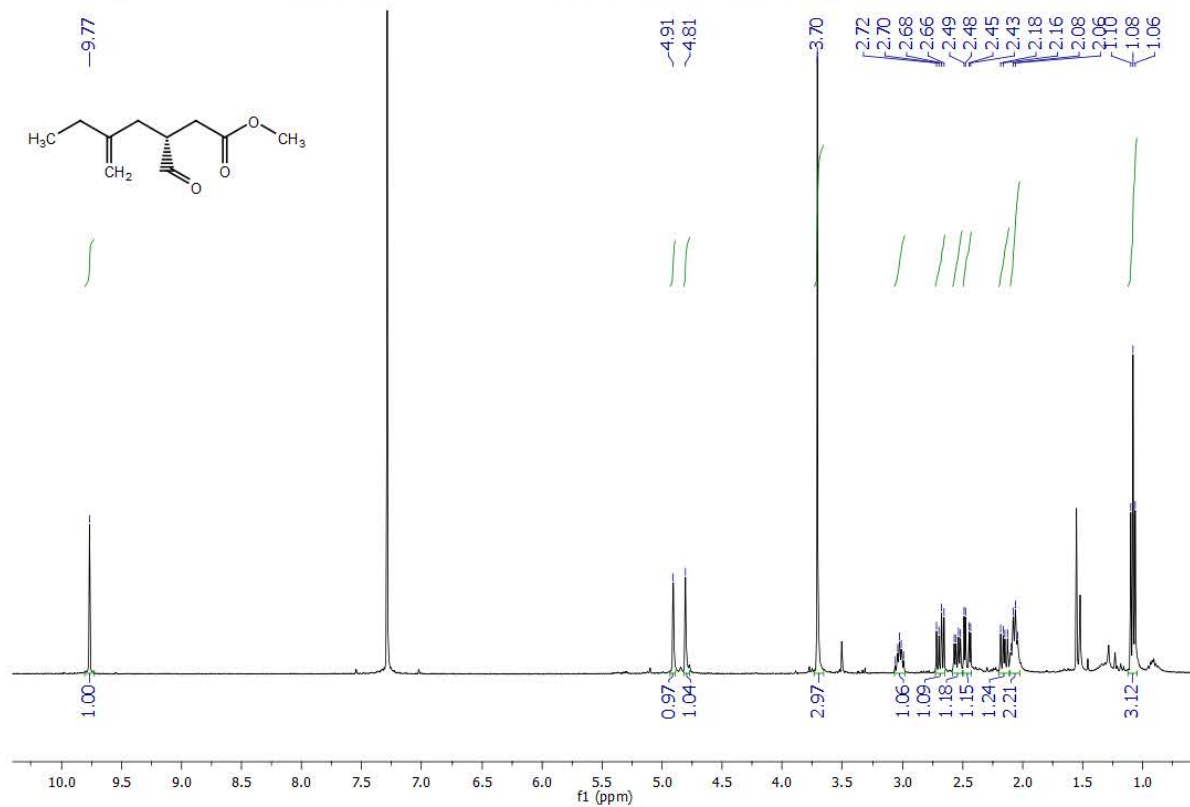


Ryc.95. Widma ¹H NMR (górá), ¹³C NMR (dóř), dla związku **32(R)**.

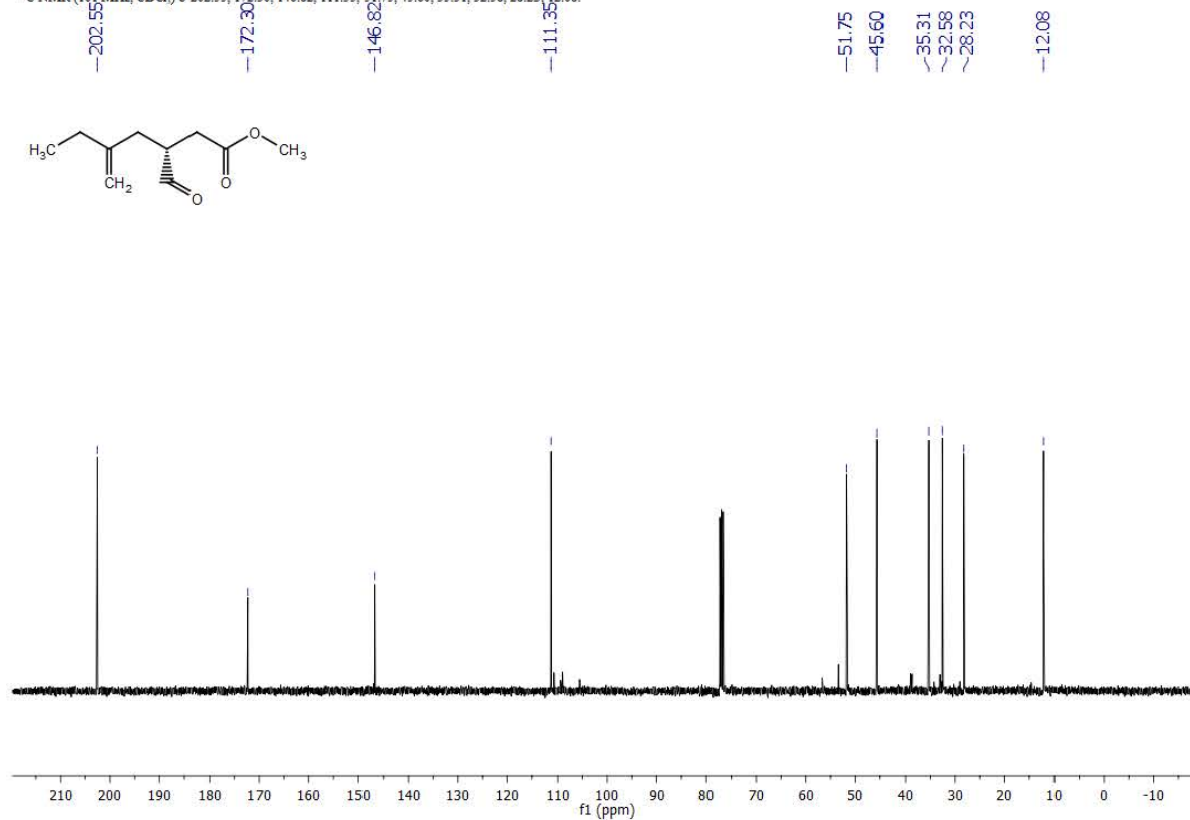


Ryc.96. Widma ¹H NMR (góra), ¹³C NMR (dół), dla związku **23(R,R)**.

$^1\text{H NMR}$ (400 MHz, CDCl_3) δ 9.77 (s, 1H), 4.91 (s, 1H), 4.81 (s, 1H), 3.70 (s, 3H), 3.07–2.98 (m, 1H), 2.69 (dd, 1H), 2.55 (dd, 1H), 2.46 (dd Hz, 1H), 2.16 (dd, 1H), 2.07 (bq, 2H), 1.08 (t, 3H).



$^{13}\text{C NMR}$ (101 MHz, CDCl_3) δ 202.55, 172.30, 146.82, 111.35, 51.75, 45.60, 35.31, 32.58, 28.23, 12.08.

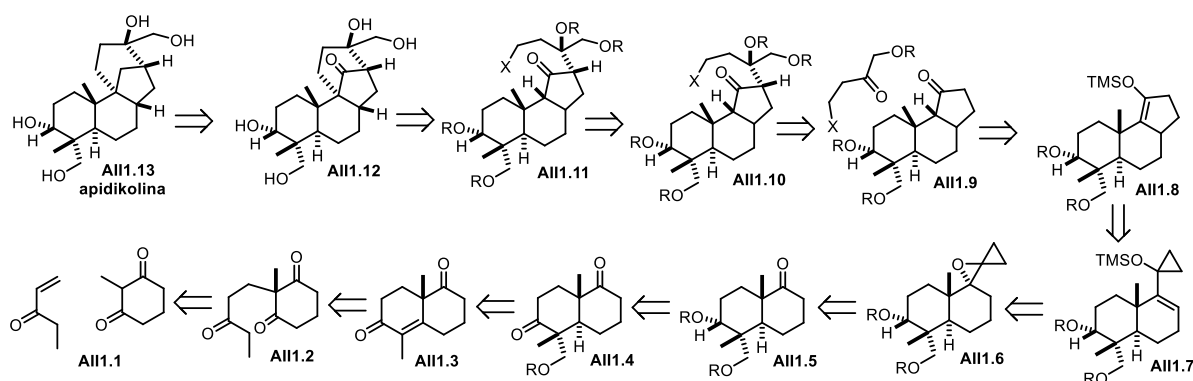


Ryc.97. Widma $^1\text{H NMR}$ (góra), $^{13}\text{C NMR}$ (dół), dla związku **23(S)**.

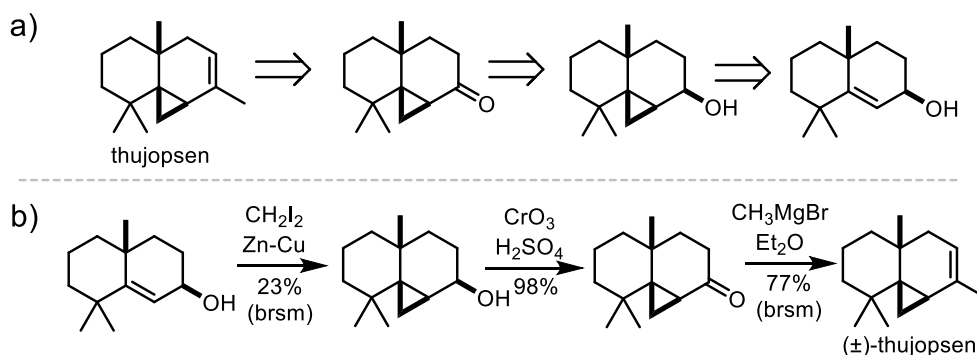
Appendiks II. Schematy pozostałych analiz retrosyntetycznych z LHASA, SYNCHEM i SYNSUP

Analizy retrosyntetyczne z LHASA

Apidikolina (*uzupełnienie*) – analiza retrosyntetyczna antybiotyku apidikoliny¹⁶⁰ (Ryc.98.). Dzięki kombinacjom taktycznym uzyskano przejścia **AIII.13**→**AIII.9** i **AIII.9**→**AIII.5**, zastosowane grupy ochronne nie zostały wyspecyfikowane. Pomimo szeregu znanych syntez totalnych apidikoliny **AIII.13**⁷¹⁻⁹¹ ten plan syntetyczny nigdy nie został w całości potwierdzony; co najwyżej α -metylowy keton Wielanda-Mieschera **AIII.3** uzyskany w reakcji Hajosa-Parrisha, został wykorzystany w syntezach totalnych apidikoliny, przez McMurry'ego⁸⁹ i Trosta⁷¹. Co ciekawe, obecność pierścieni sześciocłonowych nie była w tym przypadku dla LHASA przesłanką do zastosowania reakcji Dielsa-Aldera, którą z powodzeniem stosowano w syntezach Deslongchamps⁷⁷⁻⁷⁸ i Ihary⁸⁴. Pomimo obecności wielu centrów stereogenicznych w cząsteczce, autorzy nie komentują, w jaki sposób (oprócz oczywistego wpływu wewnętrznej indukcji) należałoby przeprowadzić te reakcje w wariacie asymetrycznym. Przykład apidikoliny, oprócz wygenerowania planu syntetycznego, trudno uznać za walidację programu, gdyż autorzy nie dokonali porównania z istniejącym stanem wiedzy (oprócz dostrzeżenia analogii początku ścieżki z syntezą Trosta⁷¹), ani nie odnieśli się do problemu grup ochronnych, czy stereoselektywności reakcji.

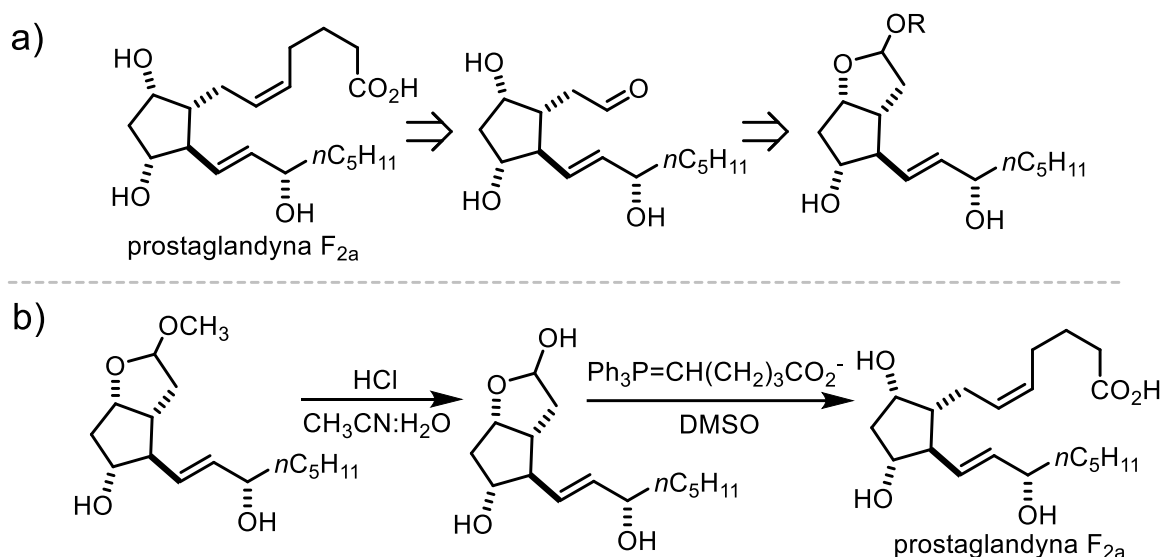


Ryc.98. Wygenerowana przy użyciu LHASA analiza retrosyntetyczna apidikoliny.



Ryc.99. a) Ilustracja analizy retrosyntetycznej thujopsenu, uzyskanej przy użyciu LHASA¹² w 1976 roku; b) Schemat ostatnich 3 etapów, wykonanej przez Daubena i Ashcrafta syntezy totalnej thujopsenu⁴⁵ z roku 1963.

Thujopsen (*uzupełnienie*) – trzypierścieniowy, seskwiterpen występujący w olejkach eterycznych, stanowił kolejny przykład walidacji LHASA przez porównanie wygenerowanej ścieżki (Ryc.99a) z syntezą znaną z literatury (Ryc.99b). W pracy¹² z roku 1976, opisującej generowanie sekwencji reakcji zmieniających tylko grupy funkcyjne, wygenerowano dla thujopsenu 3 etapową sekwencję. Sekwencja ta odpowiada 3 ostatnim etapom syntezy totalnej⁴⁵ (±)-thujopsenu, opublikowanej w 1963 roku przez Daubena i Ashcrafta. Autorzy zwracają tu szczególną uwagę na wygenerowanie przez LHASA, stereoisomeru alkoholu allilowego i wykorzystanie tego centrum stereogenicznego w stereoindukowanej reakcji Simmonsa-Smitha.



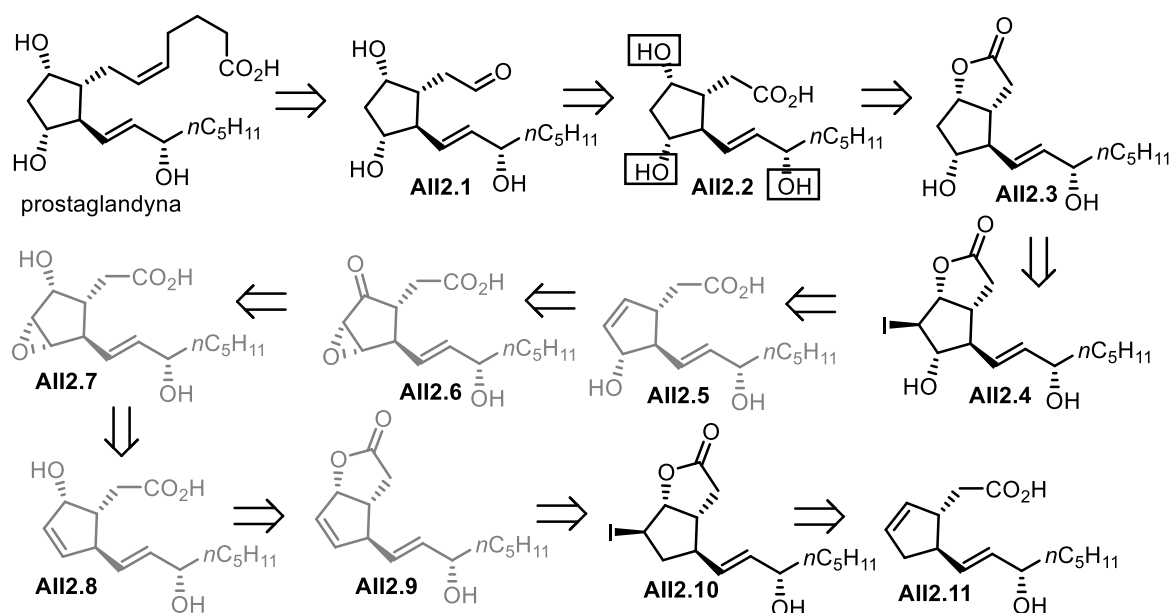
Ryc.100. a) Ilustracja analizy retrosyntetycznej prostaglandyny F_{2a}, uzyskanej przy użyciu LHASA¹² w 1976 roku; b) Schemat ostatnich 2 etapów, odpowiada wcześniejszym syntezom Corey'a^{46,47} z 1970 roku i Woodward'a⁴⁸ z 1973 roku.

Prostaglandyna F_{2a} (*uzupełnienie*) – była kolejnym związkiem występującym w pracy¹², ilustrującym sekwencję zmian grup funkcyjnych. Sekwencja podobnie jak w przypadku thujopsenu, dotyczyła analizy ostatnich etapów syntezy. Zaproponowana sekwencja (Ryc.100a) pokrywająca się z końcówkami syntez totalnych, wykonanych wcześniej przez Corey'a^{46,47} i Woodwarda⁴⁸ (Ryc.100b). Jest to sekwencja kwaśnej hydrolizy niesymetrycznego, cyklicznego acetalu i następnie podziałanie nań ylidem fosforowym, który reaguje z aldehydem pozostającym w równowadze z hemiacetalem. Odtworzenie istniejących syntez nie budzi wątpliwości, jednak zapis niesymetrycznego acetalu w postaci ogólnej grupy OR, w przypadku LHASA jest zastanawiający. Program ten operował wyłącznie na konkretnych związkach chemicznych, tymczasem zapis OR na dowolną pochodną grupy hydroksylowej, nie mógł pochodzić z programu i musiał zostać dopisany później.

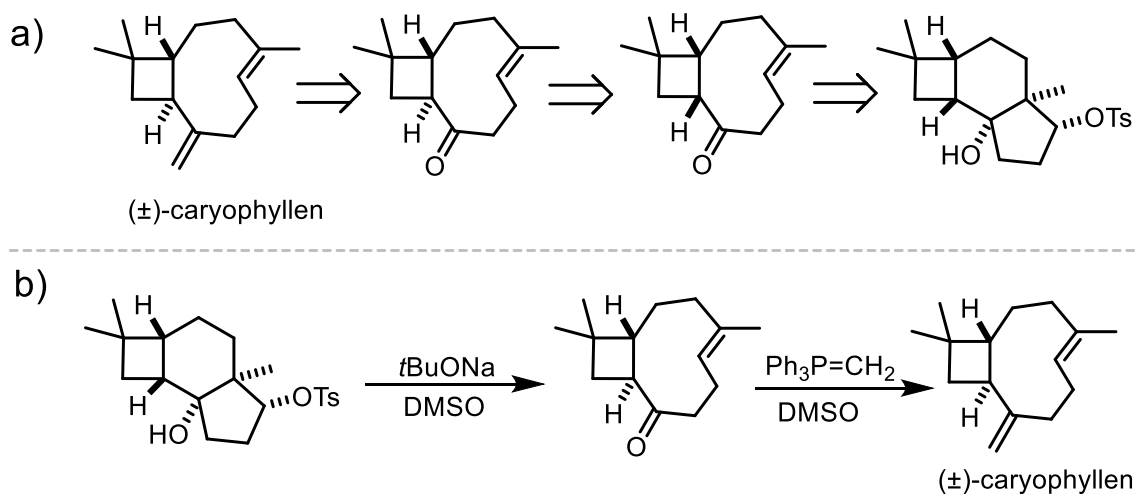
Cztery lata później, w 1980 roku, ukazała się praca dotycząca implementacji halolaktonizacji¹⁴, gdzie ponownie pojawia się plan syntetyczny prostaglandyny F_{2a} (Ryc.101.). Tym razem jest to plan pełniejszy, uwzględniający ochronę grup funkcyjnych (np. związek: **AII2.2**), jednak plan w dalszym ciągu nie kończący się na dostępnym prekursorze. Końcówka ścieżki powtarza znaną z wcześniejszej pracy sekwencję hydrolizy i reakcji Wittiga, następnie, w celu późniejszej zmiany stopnia utleniania między związkami **AII2.2**→**AII2.1**, stosowana jest hydroliza laktonu **AII2.3** uzyskanego z **AII2.5** na drodze jodolaktonizacji. Podobny (po uwzględnieniu grup ochronnych) do **AII2.4**, prekursor prostaglandyny, w reakcji jodolaktonizacji, uzyskał Sakai³⁵⁶,

jednak do jego uzyskania wykorzystał inny substrat. Prekursor typu **AII2.5**, nigdy nie został wykorzystany w reakcji jodolaktonizacji, zbliżone do niego związki, zawierające grupę aldehydową zamiast karboksylowej, wykorzystywane były w syntezach innych prostaglandyn³⁵⁷⁻³⁶², jednak wykorzystywano wówczas inne reakcje niż jodolaktonizacja. Przejście **AII2.11**→**AII2.10** zostało wykonane przez zespół Kobayashi'ego³⁶³⁻³⁶⁴ jako kontynuacja prac nad przyłączeniem bocznego łańcucha z motywem alkoholu allilowego. Strategia ta dobrze uzupełniałaby poniższą ścieżkę, o brakujące etapy syntezy **AII2.11** z prostych substratów. Związków od **AII2.10** do **AII2.6** nigdy nie uzyskano.

Zaproponowana przez LHASA analiza retrosyntetyczna prostaglandyny $F_{2\alpha}$ z 1980 roku, nigdy nie została wykonana w przedstawionym kształcie, jedynie fragmenty planu odpowiadają istniejącym wcześniej syntezom, zwłaszcza końcowe przyłączenie łańcucha kwasowego i wykorzystanie halolaktonizacji, podobne jak w podejściu Corey'a^{46,47}.



Ryc.101. Ilustracja analizy retrosyntetycznej prostaglandyny $F_{2\alpha}$, uzyskanej przy użyciu LHASA¹⁴ w 1980 roku, szarym kolorem zaznaczone są związki nigdy nie uzyskane.



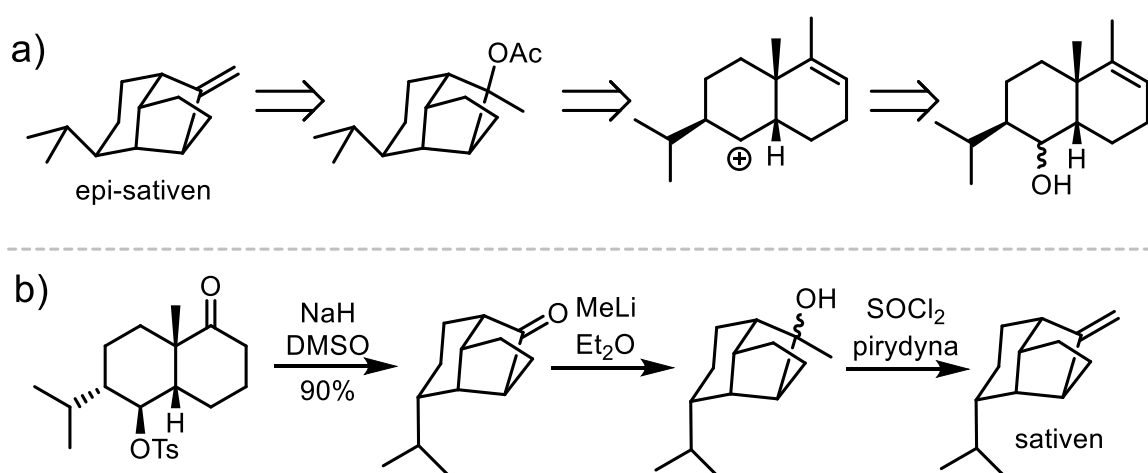
Ryc.102. a) Ilustracja analizy retrosyntetycznej (±)-caryophyllenu, uzyskanej przy użyciu LHASA¹² w 1976 roku; b) Schemat ostatnich 2 etapów, odpowiada wcześniejszej syntezie Corey'a⁵¹ z 1964 roku.

Caryophyllen (*uzupełnienie*) – w pracy¹² z roku 1976, przedstawiono krótką sekwencję ostatnich etapów syntezy caryophyllenu (Ryc.102a). Odpowiada ona ostatnim etapom wcześniejszej syntezy totalnej⁵¹ Corey'a z 1964 roku (Ryc.102b). Na szczególną uwagę zasługuje tu, zastosowanie przez LHASA, nieintuicyjnej fragmentacji Groba. Jest to reakcja nieczęsto wykorzystywana w syntezie, i trudno oprzeć się wrażeniu, że wprowadzono ją do LHASA tylko po to by móc za pomocą programu wygenerować ścieżkę, wykonaną wcześniej przez Corey'a. W żadnym innym przykładzie ścieżki z LHASA, ta reakcja nie jest stosowana.

Analiza retrosyntetyczna jest w tym wypadku dłuższa od zaprezentowanego fragmentu syntezy, bowiem LHASA osobno potraktowała etap izomeryzacji centrum stereogenicznego w pozycji α do ketonu, faktycznie reakcja ta zachodzi w zasadowych warunkach stosowanych w poprzedzającym etapie fragmentacji Groba.

Jest to kolejny przykład walidacji przez odtworzenie istniejącej syntezy.

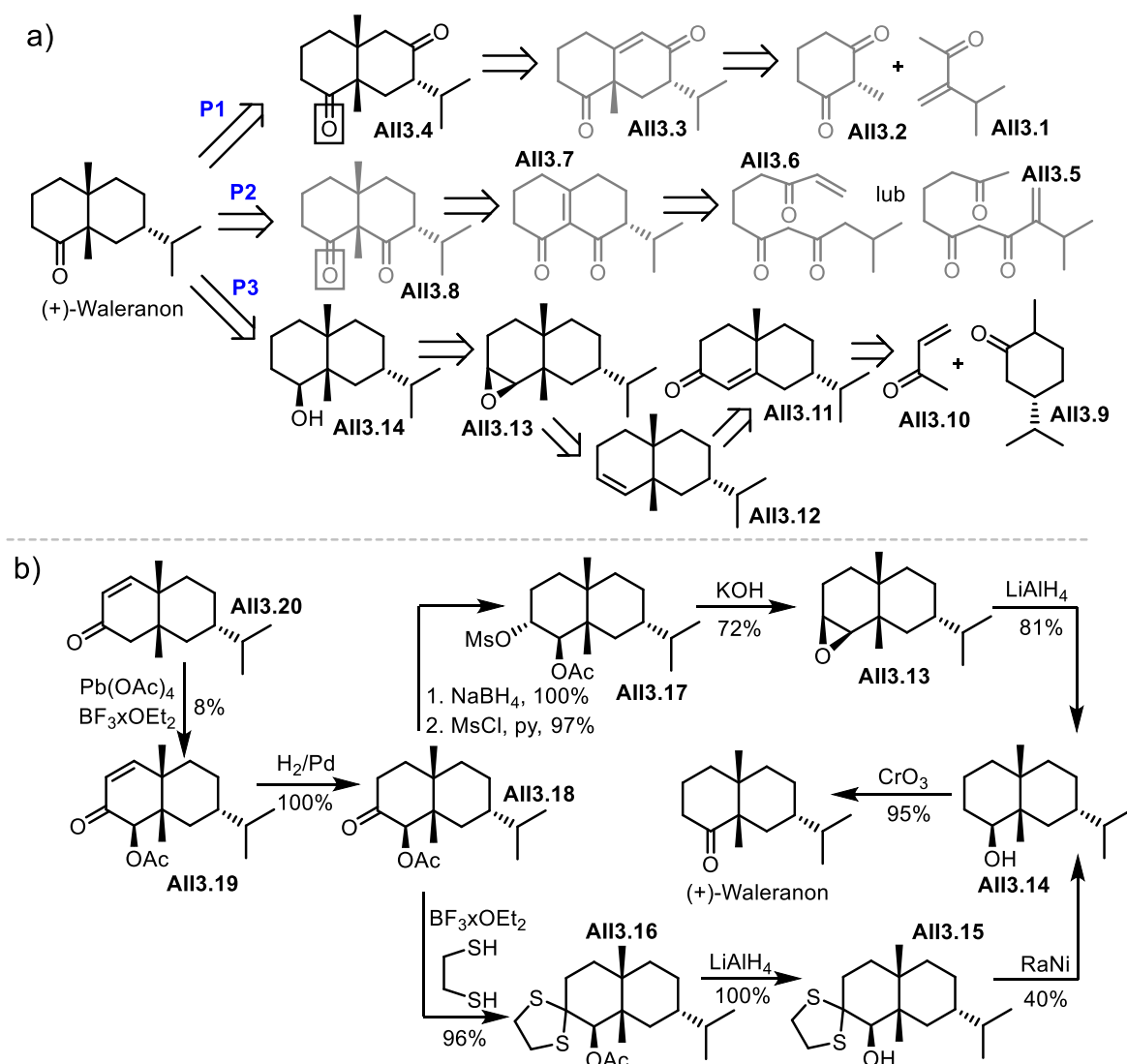
Sativen (*uzupełnienie*) – kolejnym przykładem z pracy¹² z 1976 roku o sekwencjach reakcji, jest analiza ostatnich etapów prowadzących do tego związku. Tutaj jednak różnice, po zestawieniu z literaturą są większe niż w poprzednich przykładach. Po pierwsze zwraca uwagę odwrotna konfiguracja centrum stereogenicznego przy grupie izopropylowej w planie z LHASA i w związku uzyskanym przez McMurry'ego w 1968 roku. Wygląda na to, że w skutek pomyłki operatora, LHASA wygenerowała plan dla innego związku, można go określić jako epi-sativen. Drugą i poważniejszą różnicą, jest rodzaj reakcji użyty do syntezy trójpierścieniowego szkieletu sativenu. LHASA wygenerowała przejście prowadzące przez atak wiązania podwójnego na karbokation. Wcześniej problem trudności planowania reakcji z użyciem karbokationów, omawiałem przy okazji planu dla alkoholu paczulowego. Dla tego przypadku, nigdy tej reakcji nie wykonano. Nie dziwi to, gdyż w razie wygenerowania drugorzędowego karbokationu, najpewniej przegrupuje się on na jeden z sąsiadujących atomów węgla, tworząc stabilniejszy karbokation trzeciorzędowy. Tymczasem w syntezie totalnej McMurry'ego³⁶⁵, połączeniu ulegają te same co w planie LHASA atomy węgla, tylko w przeciwieństwie do analizy komputerowej, u McMurry'ego nie występuje metylenolefina, lecz grupa ketonowa, która stabilizując karboanion, pozwala na wewnątrzcząsteczkową substytucję nukleofilową, budującą szkielet sativenu.



Ryc.103. a) Ilustracja analizy retrosyntetycznej dla epi-sativenu, uzyskanej przy użyciu LHASA¹² w 1976 roku; b) Schemat ostatnich 3 etapów syntezy epi-sativenu, przypomina wcześniejszą syntezę McMurry'ego³⁶⁵ z 1968 roku.

Zatem, pomijając nawet błąd we wprowadzeniu struktury do programu, propozycja LHASA, już na pierwszy rzut oka okazuje się niemożliwa do wykonania, ze względu na niebranie pod uwagę konkurencyjnych dróg przegrupowania zaproponowanego karbokationu. Jest to o tyle dziwne, że sam Corey wykorzystał podobne zamknięcie trójpierścieniowego szkieletu węglowodoru, przy pomocy karboanionu stabilizowanego grupą ketonową, w swojej syntezie (+)-longifolenu³⁶⁶. Można by zatem oczekiwać, że podobnie jak w poprzednich przykładach, strategia przezeń wykorzystana (a cechuje ją dużo większa doza uniwersalności niż choćby przegrupowanie Groba), znajdzie się również w tworzonym pod jego kierunkiem programie.

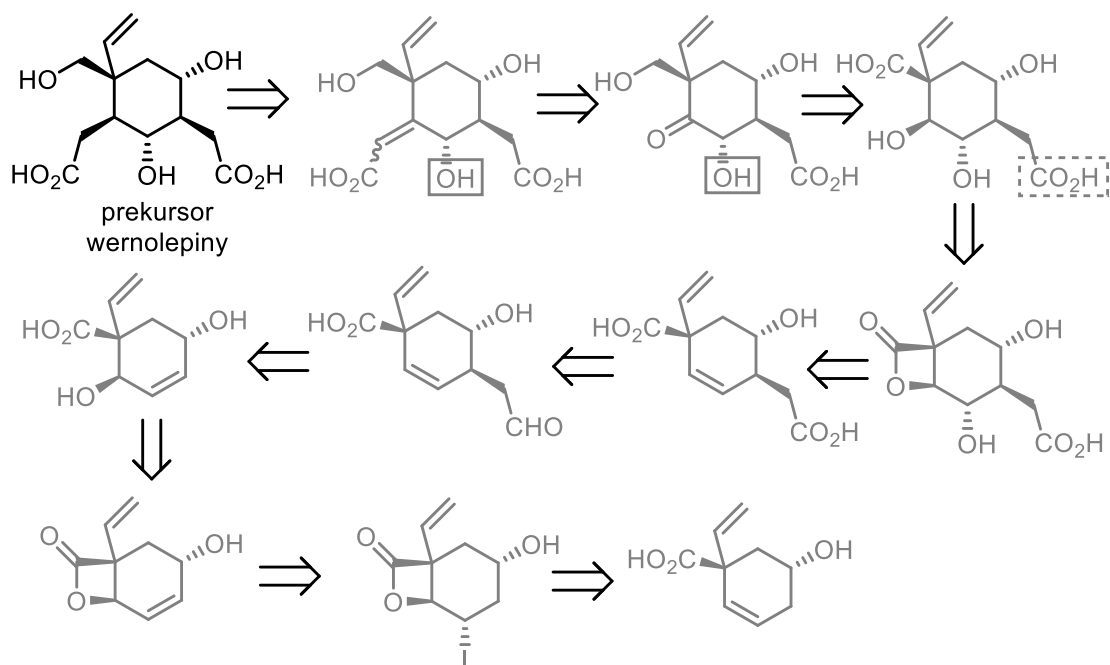
Waleranon (*uzupełnienie*) – był związkiem ilustrującym możliwości jakie dawała LHASA podczas stosowania anulacji Robinsona. Pierwszy raz analiza zawierająca 3 alternatywne plany dla waleranonu (Ryc.104a) pojawiła się w pracy¹³ z 1980 roku, następnie powtórzono ją w pracy¹⁶ z 1985 roku. Autorzy podkreślają, że plan P1 znakomicie ilustruje możliwości znajdowania



Ryc.104. a) Ilustracja 3 analiz retrosyntetycznych dla (+)-waleranonu, uzyskanych przy użyciu LHASA¹³ w 1980 roku; b) Schemat ostatnich etapów planu P3 dla (+)-waleranonu, przypominająca wcześniejszą syntezę Marshall'a³⁶⁷ z 1968 roku.

zastosowania dla anulacji Robinsona. Jednak do dziś nikt nie uzyskał związku **AII3.3** na drodze anulacji Robinsona z **AII3.1**+ **AII3.2**. Zaproponowane przejście **AII3.4**→(+)-waleranon, zostało wykonane³⁶⁸, jednak substraty uzyskano na innej drodze. Plan P2, ani żadne z jego poszczególnych przejść, nigdy nie zostało zrealizowane. Plan P3 przypomina wcześniejszą syntezę totalną Marshall'a³⁶⁷ (Ryc.104b) z 1968 roku. W porównaniu z zaproponowanym (Ryc.104a, P3) planem syntetycznym, istnieją pewne rozbieżności. Jedynie ostatnie etapy **AII3.14**→(+)-waleranon oraz redukcja epoksydu **AII3.13** do alkoholu **AII3.14** zgadzają się dokładnie. Epoksyd **AII3.13** nie został uzyskany na drodze utlenienia wiązania podwójnego (takie utlenianie trudno byłoby ówczas wykonać w sposób stereokontrolowany), lecz poprzez wewnątrzcząsteczkową substytucję nukleofilową **AII3.17**→**AII3.13**, w której kontrola stereochemiczna powstającego epoksydu jest pełna. Można jeszcze doszukiwać się podobieństw między związkami **AII3.20** z syntezy Marshall'a i **AII3.11** z planu LHASA, różniących się jedynie obecnością grupy metylowej i innym położeniem wiązania podwójnego. Jednak związku **AII3.20** autorzy nie uzyskali na drodze zaczynającej się anulacją Robinsona, przyczynę tego opisali w swojej wcześniejszej pracy³⁶⁹ z 1966 roku. Próbowali wówczas podejścia przypominającego plan LHASA: anulacji Robinsona **AII3.9**+**AII3.10** do związku **AII3.11**, jednak próba dalszej 1,4-addycji grupy metylowej w obecności soli miedzi do związku **AII3.11**, skończyła się niepowodzeniem, stąd wybór innej strategii w syntezie z roku 1968. Zastanawia zatem fakt, że autorzy późniejszej pracy o LHASA, jako walidację programu, przedstawiają plan P3, odrzucony w połowie lat 60-tych przez Marshalla. Dwa pozostałe plany P1 i P2, nie zostały nigdy wykonane.

Na pierwszy rzut oka wygenerowana przez LHASA plany P1-3, wydają się teoretycznie poprawne, jednak zaproponowane związki wyjściowe, takiej jak **AII3.5** i **AII3.6** budzą wątpliwości, czy ulegną anulacji Robinsona, czy też akrylowy fragment ulegnie reakcji z najlepiej stabilizowanym karboanionem w pozycji 2. 1,3-diketonu. Ponadto w żadnej z 2 prac przytaczających przedstawione plany, nie znalazł się komentarz dotyczący stereokontrolowanego wprowadzania grup metylowych.



Ryc.105. Ilustracja planu syntetycznego dla prekursora wernolepiny, wygenerowanego¹⁴ przy użyciu LHASA i ogłoszonego w 1980 roku.

Można założyć, że pewna indukcja wynikać będzie z ataku od strony przeciwnej do grupy izopropylowej, jednak to zmienia tylko pytanie do postaci: w jaki stereokontrolowany sposób, wprowadzić centrum stereogeniczne przy grupie izopropylowej – tego typu wyjaśnień, nie podano. Reasumując, przytoczone przez LHASA plany syntetyczne (+)-waleranonu w konfrontacji z literaturą budzą wątpliwości.

Prekursor wernolepiny (*uzupełnienie*) – żadna z istniejących syntez totalnych wernolepiny nie przypomina tej zaproponowanej przez LHASA Ryc.105. Żaden z pozostałych zaproponowanych związków (również w postaci z grupami ochronnymi), nigdy nie został uzyskany.

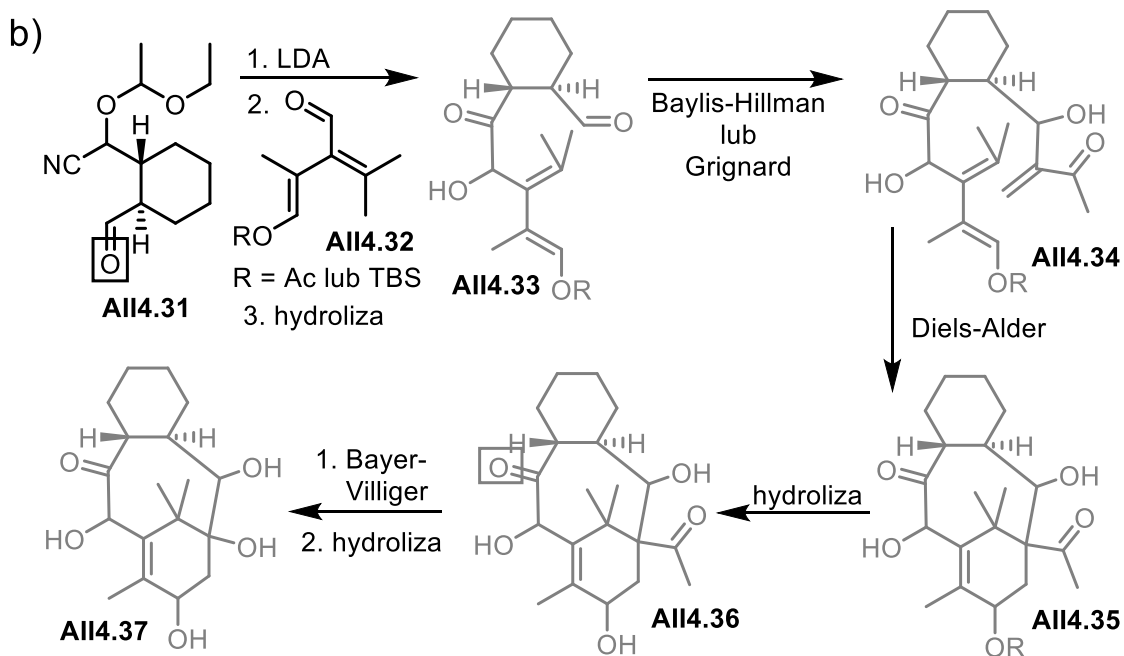
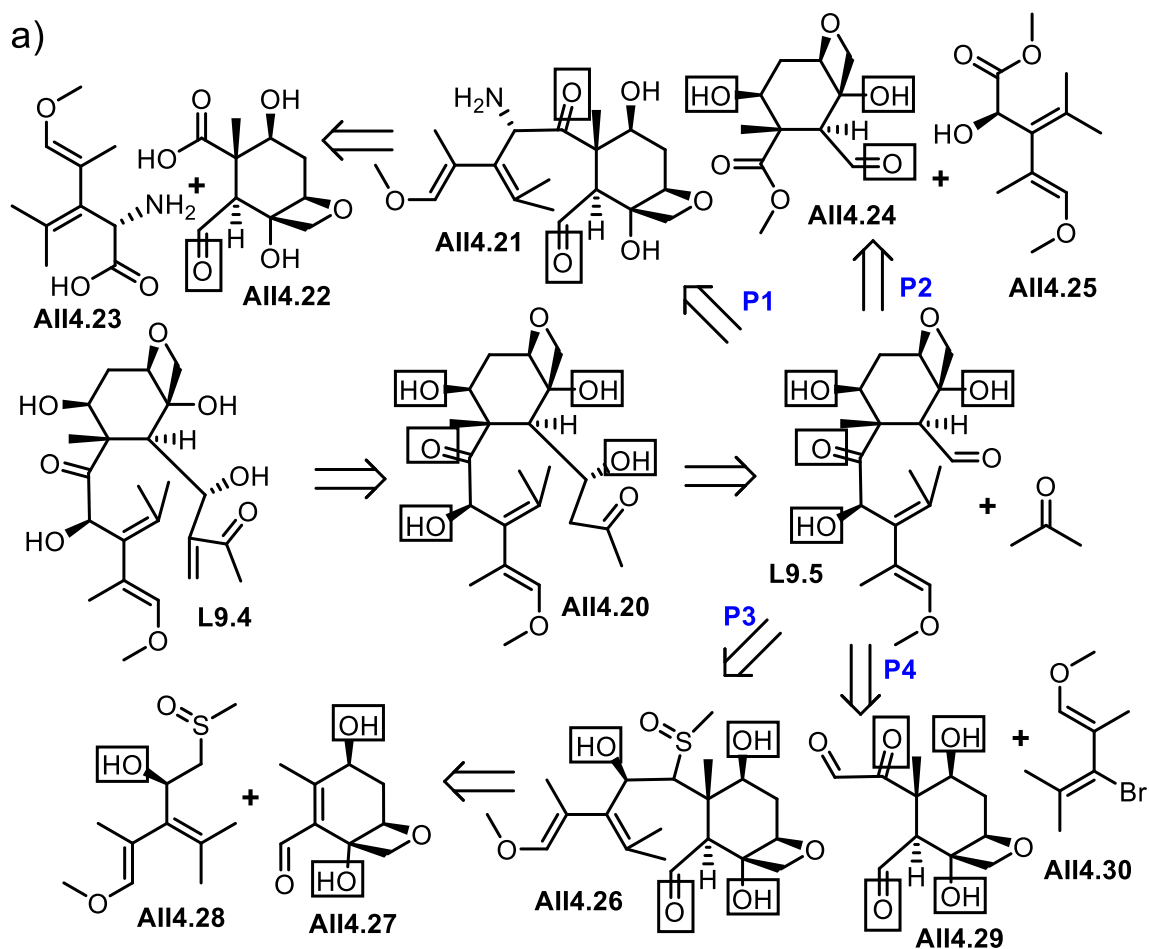
Taksol (*uzupełnienie*) – autorzy pracy²² oprócz zaprezentowanej na Ryc.9. analizy retrosyntetycznej, przedstawili 4 alternatywne (P1-P4) plany syntezy **L9.5** (Ryc.106a) omówione skrótowo w głównym tekście. Inspirując się analizami pochodzącymi z LHASA (Ryc.106b), autorzy zaproponowali plan syntetyczny dla związku **AII4.37**, który jest jeszcze bardziej uproszczoną wersją związku **L9.1** (Ryc.9.), pozbawioną 3 centrów stereogenicznych i wrażliwego pierścienia oksetanowego.

W pracy²² z 1994 roku opisano 5 różnych dróg uzyskania **L9.5** (Ryc.106a). Celem tych reakcji jest przyłączenie motywu dienu do sześcioczłonowego rdzenia, co w dalszych etapach umożliwi reakcję Dielsa-Aldera **L9.4** → **L9.3**. Mamy tu okazję zapoznać się z reakcjami tworzenia wiązania węgiel-węgiel, wytypowanymi przez LHASA.

Plan P1 (Ryc.106a) zakładał wykorzystanie reakcji Dakina-Westa. Problem jednak w tym, że zaproponowana reakcja, musiałaby zajść pomiędzy bezwodnikiem, pochodzącym z **AII4.22** a aminokwasem **AII4.23**. W reakcji tej stosuje się symetryczny bezwodnik, ponieważ, jak wynika z mechanizmu reakcji¹¹⁵ aż trzy ekwiwalenty bezwodnika reagują z aminokwasem. Zatem potrzeba by było aż 6 ekwiwalentów, złożonego strukturalnie **AII4.22**. Wykorzystanie tej reakcji nie jest dobrą propozycją utworzenia wiązania węgiel-węgiel ze złożonym strukturalnie kwasem karboksylowym.

Plan P2 proponował wykorzystanie reakcji Claisena i następczej dekarboksylacji. Na uwagę zasługuje jednak fakt, że substrat **AII4.24** nie posiada protonów w pozycji α , zatem nukleofilem w tej reakcji byłby karbanion pochodzący z **AII4.25**. Pomijając już fakt, że LHASA nie zaproponowała ochrony wolnej grupy hydroksylowej w tym hydroksyestrze, to wykorzystanie **AII4.25** jako nukleofila, skutkowałoby racemizacją centrum stereogenicznego przy grupie hydroksylowej, tymczasem LHASA nie zauważyła problemu racemizacji w warunkach Claisena i jako produkt proponuje **L9.5** z niezmienną konfiguracją centrum stereogenicznego.

Plan P3 zakłada użycie reakcji Michaela, akroleiny z karbanionem w pozycji α do sulfotlenku i następnie zamianę sulfotlenku na keton w przegrupowaniu Pummerera. Propozycja takiej reakcji pojawiła się, gdyż pozwoliły na to reguły zapisane w bazie transformacji w LHASA, jednak istnieje szereg zastrzeżeń wobec tego przekształcenia. Po pierwsze, karbanion na sulfotlenku musiałby zostać wygenerowany w sposób selektywny na dłuższym podstawniku **AII4.28**. Po drugie, twardy nukleofil węglowy wykazuje tendencję do bezpośredniego ataku na aldehyd, zamiast sprzężonego ataku 1,4. O ile dla związków Grignarda i litoorganicznych, daje się sterować tą selektywnością za pomocą dodatku związków miedzi, o tyle dla akrolein i nukleofili sulfotlenkowych, nie jest znana metoda takiej kontroli. Zatem chemoselektywność proponowanej przemiany nie koresponduje z rzeczywistą reaktywnością substratów. Po trzecie, w substracie **AII4.27** występuje wrażliwy na nukleofile pierścień oksetanowy¹¹⁶. Konkurencyjna reakcja jego otwarcia przez karbanion w pozycji α do sulfotlenku, nie została uwzględniona. Po czwarte, reakcja generuje czwartorzędowe centrum stereogeniczne w reakcji Michaela, nawet jeżeliby ta reakcja mogła zajść, to brak komentarza o źródle indukcji asymetrycznej w reakcji. Plan P3 proponuje zatem



Ryc.106 a). Ilustracja wariantów A-D planu syntetycznego dla związku przejściowego L9.5 występującego w analizie retrosyntetycznej rdzenia taksolu L9.1 (Ryc.9.), wygenerowanego przy użyciu LHASA²² w 1994 roku.

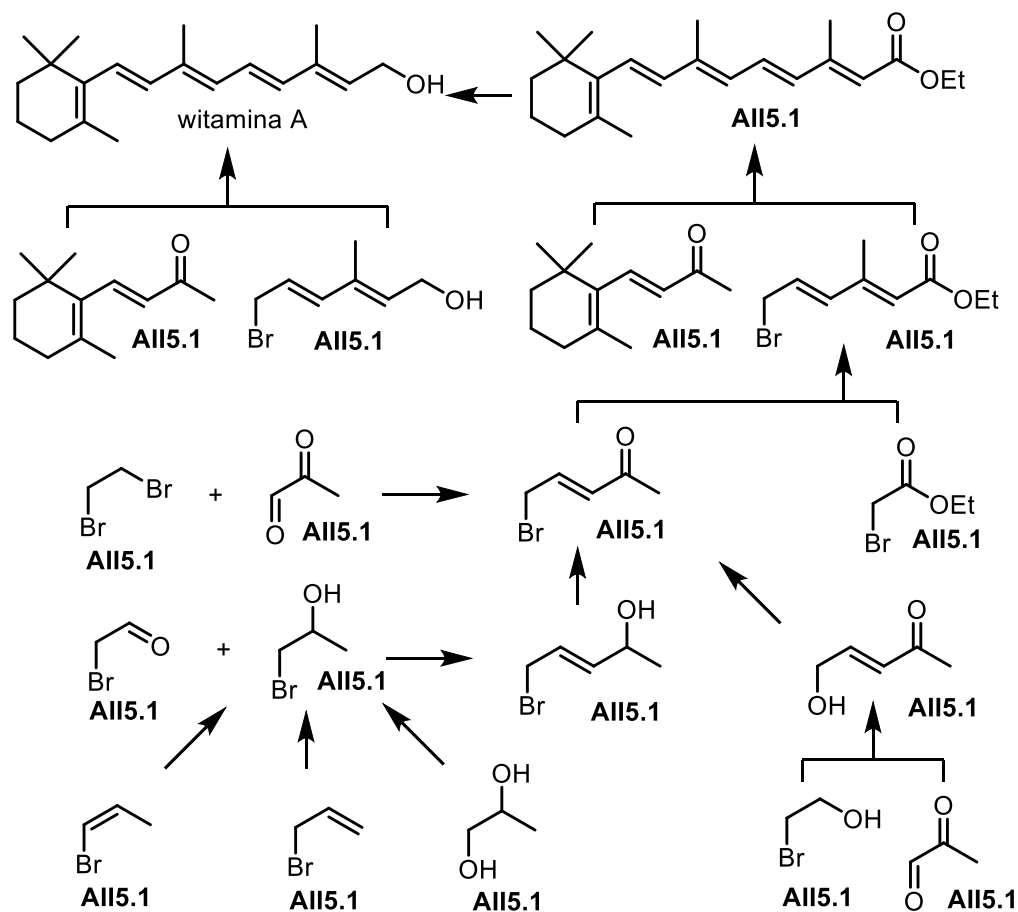
b) Plan syntetyczny dla AII4.37 będącego uproszczoną wersją związku L9.1 (Ryc.9.).

wykorzystanie budzącej wątpliwości metody tworzenia wiązania węgiel-węgiel, która dla nieskomplikowanych strukturalnie substratów, mogłaby mieć szanse powodzenia, jednak dla rdzenia taksolu, brak percepcji, kluczowego kontekstu chemicznego (chemoselektywność, stereoselektywność, niekompatybilne grupy funkcyjne), nie pozwala poważnie brać pod uwagę zastosowania tej reakcji.

Plan P4, proponujący reakcję uzyskanego z **AII4.30** związku Grignarda, z aldehydem **AII4.29**, budzi podobne wątpliwości co plan P3. Znana jest wszak reaktywność nukleofili magnezoorganicznych wobec pierścieni oksetanowych¹¹⁶. Tu niekompatybilność tej grupy funkcyjnej, wobec proponowanych warunków reakcji, nie została odnotowana.

Analizy retrosyntetyczne z SYNCHEM

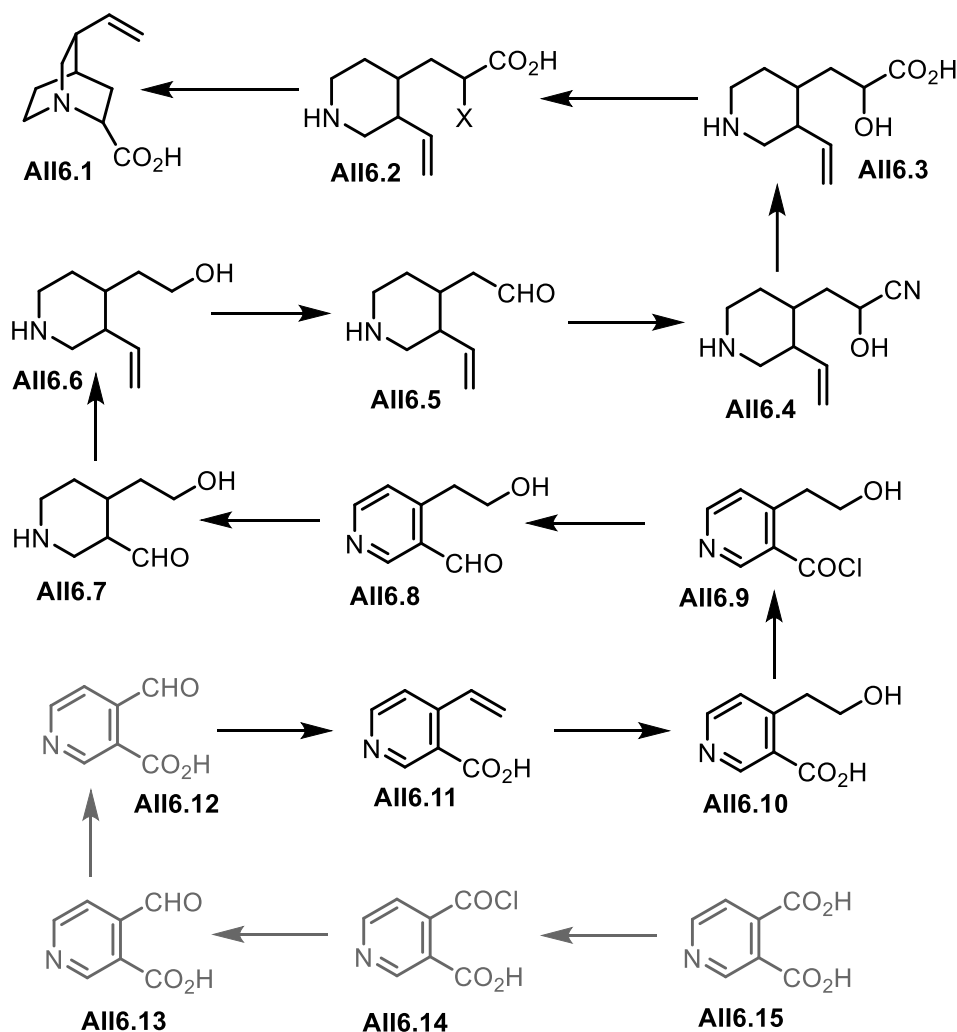
Witamina A – w pracy¹²⁴ z 1973 roku, znajdujemy plan syntetyczny dla witaminy A (Ryc.19), o którym twórcy w kolejnych pracach już nie wspominają. Opiera się on na wielokrotnym stosowaniu reakcji Wittiga, w celu wydłużania szkieletu związku za pomocą tworzenia kolejnych wiązań podwójnych. Interesujące jest, że program „tnie” retrosyntetycznie cząsteczkę w połowie (**AII5.1** + **AII5.17** → **AII5.2** i **AII5.1** + **AII5.3** → **AII5.2**), wybierając do dalszej analizy fragmenty o zbliżonej wielkości. Brak jest komentarza, na temat różnicy pomiędzy proponowanymi substratami **AII5.3** i **AII5.17** wybranymi na prekursorzy ylidu fosforowego. Obecność elektronoakceptorowej grupy estrowej w **AII5.3**, umożliwia powstanie winylogowej stabilizacji generowanego ylidu, czego



Ryc.107. Ilustracja planów syntetycznych dla witaminy A, wygenerowanych przez program SYNCHEM¹²⁴.

grupa hydroksylowa w **AII5.17** nie zapewnia. Również następny etap opiera się na wykorzystaniu stabilizowanego ylidu, pochodzącego z **AII5.4**, który w reakcji z ketonem **AII5.8**, prowadzi do estru **AII5.3**. Do tego etapu, proponowany plan, wygląda chemicznie przekonująco, jednak przedstawione propozycje syntezy **AII5.8** ujawniają ograniczony repertuar stosowanych reakcji. Przedstawiony plan zakłada przejście **AII5.13** + **AII5.12** → **AII5.11**, które grozi jednak zajęciem wewnątrzcząsteczkowego epoksydowania bromohydryny **AII5.12**, na etapie tworzenia ylidu, ochrona grupy hydroksylowej nie jest tu wzmiankowana. Podobnie reakcja Wittiga **AII5.9** + **AII5.10** → **AII5.8**, nigdy nie została zastosowana (popularniejsze jest bromowanie pozycji allilowej). Podsumowując, opublikowany w 1973 roku plan syntetyczny dla witaminy A, można określić jako chemicznie poprawny choć podejrzanie monotony.

Pochodna chinuklidyny (*uzupełnienie*) – praca z 1977 roku¹²⁵ podaje 13-etapową ścieżkę prowadzącą do podjednostki strukturalnej **AII6.1**, występującej w alkaloidach kory chinowca. Należy tu jednak zaznaczyć, że plan automatycznie wygenerowany przez SYNCHEM, kończy się na związku **AII6.11**. Ze względu na zapis reakcji Wittiga, wskazujący na niekompatybilność z nią grupy karboksylowej, etapy **AII6.12**- **AII6.15** nie mogły zostać wygenerowane, dopisano je później.



Ryc.108. Ilustracja planu syntetycznego dla **AII6.1**, wygenerowanego przez program SYNCHEM¹²⁵.

Przedstawiony plan wydaje się wykonalny, jednak kilka z zaproponowanych przejść budzi wątpliwości. Już pierwsze przejście **АИІ6.15** → **АИІ6.14** różnicujące grupy karboksylowe, nie wydaje się dobrą propozycją. Takiej reakcji nigdy nie przeprowadzono, co nie dziwi, gdyż selektywne wytworzenie chlorku kwasowego na tylko jednej z grup karboksylowych, nie jest w żaden sposób promowane. Co więcej, taki związek w obecności pierścieni pirydynowych, ulegać może wewnątrzcząsteczkowemu tworzeniu bezwodnika, w którym na powrót obie grupy karbonylowe przestają być wyraźnie rozróżniane. Następnie wątpliwości budzi przejście **АИІ6.11** → **АИІ6.10** wymagające antymarkownikowskiego hydroksylowania grupy winylowej. Przemianę tą można zrealizować przy użyciu BH_3 , jednak odczynnik ten reaguje również z kwasami karboksylowymi. Tego typu reakcja dla związków aromatycznych nie jest znana, można próbować zastąpić BH_3 za pomocą 9-BBN, jednak obawiam się, że w zapisie reakcji, informacja o niekompatybilności została pominięta. Z następną niekompatybilnością spotykamy się w przejściu **АИІ6.10** → **АИІ6.9**, brak tu informacji o ochronie grupy hydroksylowej. Nawet jeżeli udałoby się wygenerować chlorek kwasowy bez zamienienia grupy hydroksylowej na chlorek, to zaproponowany związek **АИІ6.9** będzie wchodził w wewnątrzcząsteczkową reakcję tworzenia 6-członowego laktonu, co wydaje się procesem faworyzowanym. Pewne obawy budzić też może jednoczesna obecność grupy aldehydowej i aminowej w związkach **АИІ6.7** i **АИІ6.5**. O ile wewnątrzcząsteczkowe zamknięcie pierścienia w związku **АИІ6.7** jest wykluczone ze względu na jego zbyt mały rozmiar i idące za tym naprężenie struktury, a jedyna groźba płynie ze strony reakcji międzycząsteczkowych. To już dla związku **АИІ6.5** taki scenariusz wydaje się prawdopodobny. Nawet jeżeli nie doszłoby do utworzenia 6-członowego pierścienia iminy, to powstanie hemiaminalu jest już bardziej prawdopodobne. Związek w postaci stabilnego hemiaminalu, nie będzie wchodził w konkurencyjną reakcję tworzenia cyjanohydryny. Problemów tych można by uniknąć, poprzez zabezpieczenie grupy aminowej – jednak informacje o konieczności jej ochrony w celu unikania konfliktów reaktywności nie zostały podane.

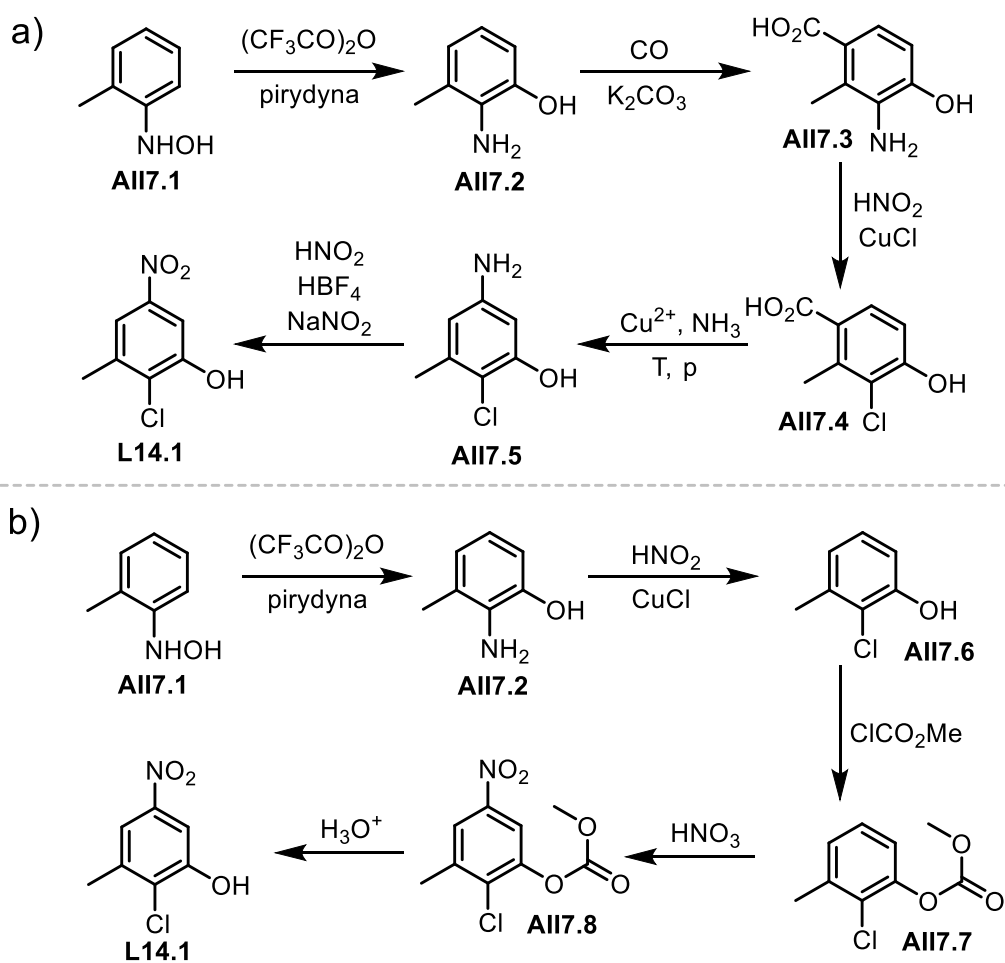
Podobnie jak w przypadku pochodnej slafradiolu, na uwagę zasługuje wykorzystanie pierścienia pirydiny w celu funkcjonalizacji łańcuchów bocznych i następująca po tym dearomatyzacja w celu uzyskania pożądanej alifatycznej aminy. Zaprezentowana sekwencja (po uwzględnieniu kilku powyższych uwag) wydaje się planem poprawnym choć nieco zbyt długim w porównaniu ze złożonością celu syntetycznego. W przypadku szkieletu alkaloidów chinowych, wykorzystywanych przede wszystkim jako ligandy zapewniające stereokontrolę reakcji, brak uwzględnienia właściwości stereochemicznych cząsteczki w planie syntetycznym jest szczególnie dojmujący.

Analizy retrosyntetyczne z SYNSUP

Pierwsze rezultaty pracy SYNSUP nie są skomplikowane strukturalnie, możliwość zaproponowania syntezy związków aromatycznych albo posiadających centra stereogeniczne ilustruje jedynie możliwości, których dotychczasowe programy nie posiadały. Przedstawienie w głównym tekście wszystkich zaprezentowanych przez autorów przykładów nie było niezbędne dla prowadzonej dyskusji. Tutaj umieszczam nieomówione przykłady pracy SYNSUP.

W pracy o SYNSUP z 1990 roku¹³⁷ przedstawiono pierwszą serię rezultatów pracy programu. Na Ryc.109. zaprezentowano dwa plany syntetyczne, będące ilustracją pracy z SYNSUP ze strukturami aromatycznymi. Cel syntetyczny **L14.1** nie posiada szczególnych właściwości, ani nie stanowi bloku budulcowego dla żadnego ważnego związku. Przy warunkach początkowych w postaci: 1) związek wyjściowy może mieć nie więcej niż 1 grupę funkcyjną i nie więcej niż 2 pierścienie aromatyczne, 2) wszystkie reakcje muszą być regiospecyficzne, by uniknąć powstawania i rozdziału izomerów orto/para, 3) synteza nie może być dłuższa niż 4 etapy. Przy tak

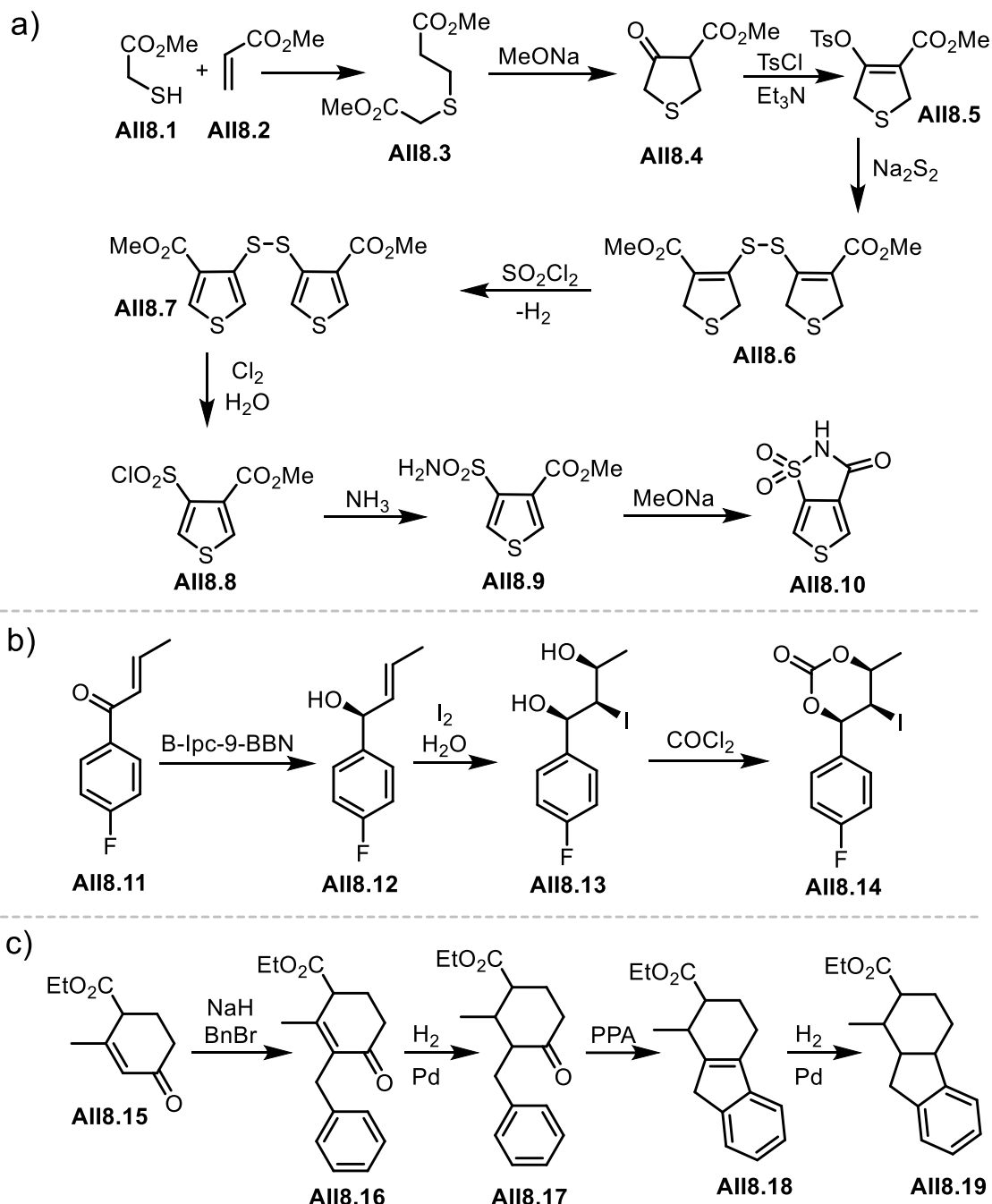
założonych warunkach nie znaleziono żadnej drogi syntezy, po zwiększeniu limitu kroków o 1, SYNSUP wygenerował dwa plany: Ryc.109a) i Ryc.109b). W obu z nich związek wyjściowy jest dwupodstawioną pochodną benzenu. Pierwsza reakcja **AII7.1** → **AII7.2** dla obu planów jest identyczna i zgodnie z nałożonym wymogiem – *ortho*-specyficzna. Druga reakcja **AII7.2** → **AII7.3** z planu a) budzi większe wątpliwości. Autorzy powołują się tu, na *para*-specyficzne karboksylowanie Yasuhary³⁷⁰, jednak w oryginalnej pracy reakcji poddawany jest jedynie niepodstawiony fenol, wpływ podstawników nie został tam zbadany, a w literaturze nie ma przypadków realizacji tej reakcji w obecności grupy aminowej w pozycji *ortho* do grupy hydroksylowej. Co to nam mówi o regule reakcji w SYNSUP? W tym przypadku została ona stworzona wyjątkowo ogólnie w oparciu o skąpą podstawę literaturową, brak uwzględnienia kontekstu chemicznego w postaci wpływu grup funkcyjnych na reaktywność pierścienia fenyłowego stanowi poważne zastrzeżenie. Jeżeli inne reguły reakcji z podobną łatwością abstrahowania, dopuszczały dowolne podstawniki w miejsca gdzie nie miało prawa ich być, to pomimo imponującej liczby reakcji w bazie danych, reguły te trzeba nazwać bezużytecznymi. Pomijając reakcję Sandmeyera, następny etap budzi wątpliwości podobne do omówionego poprzednio karboksylowania. Selekttywne aminowanie chlorofenolu **AII7.4** wydaje się zachodzić wyłącznie pod kontrolą podstawnika chlorkowego, z pominięciem silnego wpływu kierującego grupy hydroksylowej i słabszego metylowej. Podobnej reakcji również nie znajdziemy w literaturze.



Ryc.109. Ilustracje planów syntetycznych dla czteropodstawionej pochodnej benzenu **L14.1** wygenerowane przez program SYNSUP¹³⁷.

Krótko reasumując w 5-etapowym planie, dla aż 2 reakcji, zaproponowane selektywności nie mają miejsca, w takich warunkach powstaną związki inne niż założono. Przykład ten ujawnia trudności jakie napotyka program już przy pozornie prostym związku.

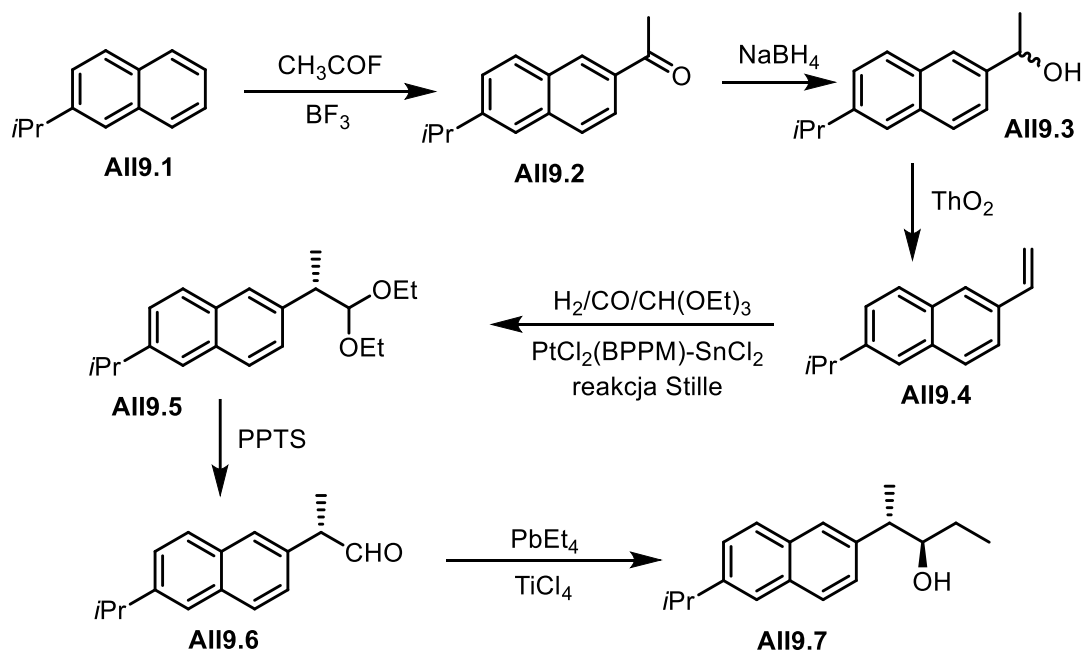
Plan b) jest dużo bardziej przekonujący, tu jedynie założenie selektywności nitrowania **AII7.7** → **AII7.8** budzi wątpliwości. Podobnie można odnieść wrażenie, że uwzględniono tylko wpływ chlorku, z pominięciem wpływu grupy metylowej i węglanowej. W tym przypadku, przynajmniej jednym z głównych produktów reakcji, powinien być założony produkt.



Ryc.110. Ilustracja planów syntetycznych dla: a) tiofenowego odpowiednika sacharyny **AII8.10**, b) chiralnego węglanu **AII8.14**, c) związku **AII8.19**; wygenerowanych przez program SYNSUP i ogłoszonych w 1990 roku¹³⁷.

W tej samej pracy¹³⁷ z pierwszymi wynikami SYNSUP z 1990 roku zaprezentowano jeszcze inne plany syntetyczne. Jedynie plan Ryc.110a) był analizą znanego związku **AII8.10**. Poszukiwano wówczas tańszej drogi syntezy tego słodzika – planu nie osiągnięto, a uzyskano jedynie ścieżkę będącą kopią dotychczas już znanej. Warunki początkowe zakładały: 1) związek wyjściowy może mieć nie więcej niż 4 atomy węgla, 2) ilość kroków nie może być większa niż 8 (dla limitu 7 kroków brak rezultatów). Co ciekawe na ścieżce a) pojawia się dimer **AII8.7**, zgodnie z założeniem rosnącego i jak największego uproszczenia w każdej generacji, związek **AII8.7** nie powinien się pojawić po **AII8.8** – dla dimerów jednak zrobiono wyjątek od tej reguły. Interesujące jest też, że złożenie pierścienia tiofenu rozbito na etapy dearomatyzacji i złożenia pierścienia z alkilowych fragmentów. Można odnieść wrażenie, że autorzy zaimplementowali do SYNSUP nietrywialne transformacje (**AII8.7** → **AII8.8**, **AII8.6** → **AII8.7**, **AII8.5** → **AII8.6**) występujące w znanej ścieżce, w nadziei, że wykorzystując elementy znanego planu SYNSUP zaproponuje lepsze alternatywy generowania związków przejściowych. To przypuszczenie poparte jest jednak oświadczeniem twórców z 2009 roku¹³⁸, że dopiero w ostatniej dekadzie tworzenia programu w sposób systematyczny dodawano reguły reakcji opisujące transformacje pierścieni heterocyklicznych. Trudno mi zatem uwierzyć, że bez wcześniejszego związku z tą cząsteczką, we wczesnej wersji programu i wobec braku tylu innych ważniejszych transformacji, znalazła się tak specyficzna reakcja jak utlenianie dihydrotiofenu do tiofenu.

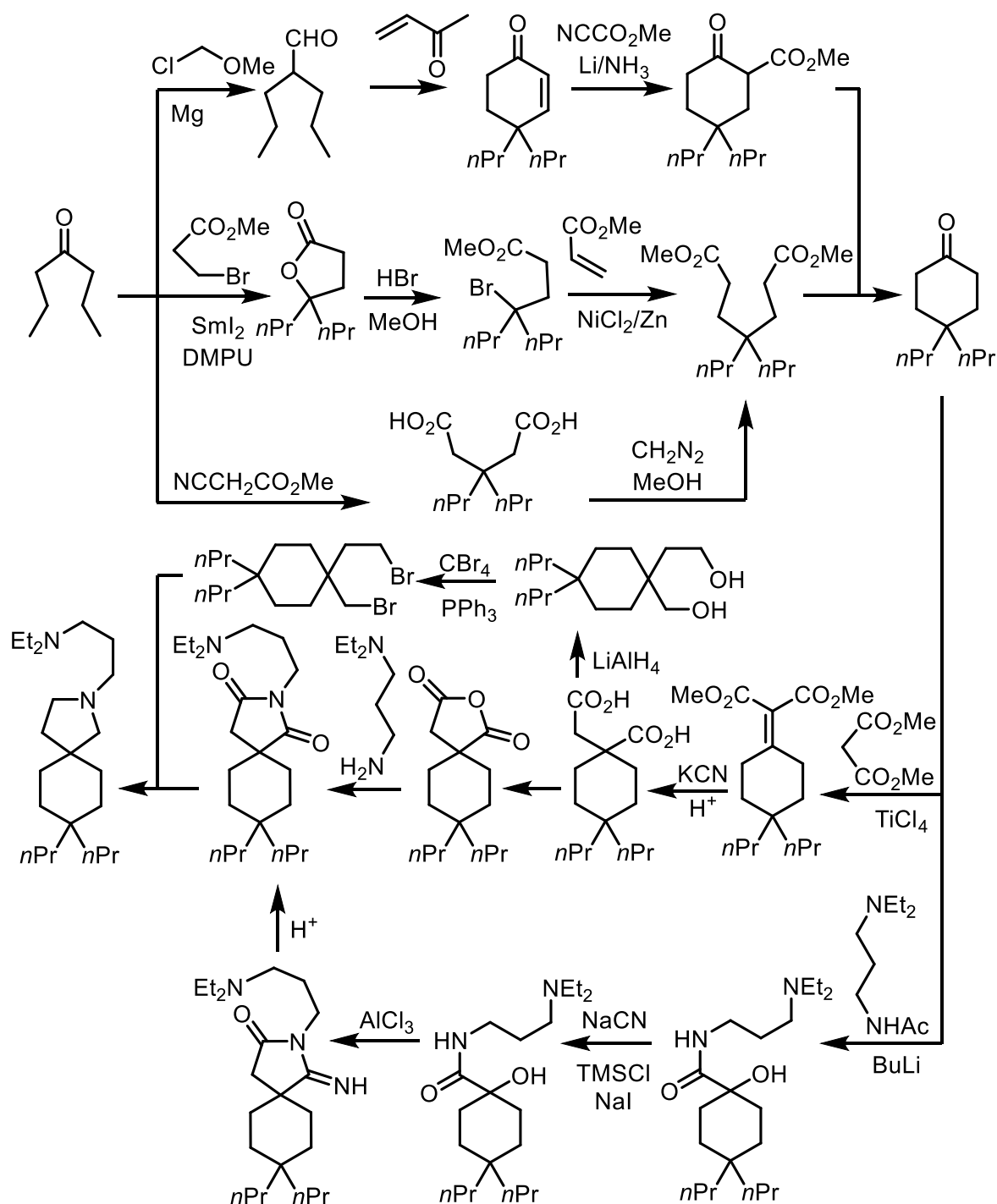
Plan dla węglanu **AII8.14** z Ryc.110b) jest interesujący z powodu zastosowania enancjoselektywnej reakcji **AII8.11** → **AII8.12**. W warunkach dla tego planu narzucono i maksymalną długość 4 etapów i konkretny związek wyjściowy: **AII8.11**. Wprawdzie produkt **AII8.14** nie ma znaczenia w kategoriach komercyjnych, jednak jako ilustracja sekwencji **AII8.11** → **AII8.12** → **AII8.13**, generującej centra chiralności, jest to unikatowy dla swoich czasów (początek lat 90-tych) przykład percepcji przez program komputerowy elementów stereogenicznych w strukturze, oraz wykorzystania transformacji je tworzących.



Ryc.111. Ilustracja planu syntetycznego dla chiralnej pochodnej naftalenu **AII9.7** wygenerowanego przez program SYNSUP i ogłoszonego w 1992 roku¹⁴².

W planie (Ryc.110c) dla trójpierścieniowego związku **AIH8.19**, na uwagę zasługuje przejście **AIH8.17** → **AIH8.18**, które ilustruje nieoczywistą metodę cyklizacji. Również dwukrotne zastosowanie w kierunku retrosyntetycznym redukcji wiązania podwójnego. Przy implementowaniu do programu tego typu przejść, należy zachować ostrożność, gdyż każdy fragment alkiłowy może zostać uzyskany w ten sposób. Skutkiem czego, niereaktywne a budujące szkielet cząsteczki fragmenty alkiłowe generują znaczną ilość potencjalnie prowadzących doń przekształceń. Nietrudno sobie wyobrazić o ile zwiększa się wówczas w każdej generacji ilość możliwych wyników. Jak zatem poradzić sobie z tym balastem, nie wykluczając tym samym rozsądnych przekształceń, takich jak **AIH8.16** → **AIH8.17** i **AIH8.18** → **AIH8.19**? Twórcy SYNSUP ograniczyli ten problem poprzez arbitralne zróżnicowanie wiązań podwójnych na te warte analizy i pozostałe. Jako warte analizy i tym samym generowania, wytypowano olefiny w pozycjach α,β do grup karbonylowych, olefiny egzocykliczne oraz te zlokalizowane w odległości jednego wiązania od pierścieni aromatycznych. W czasach sprzed upowszechnienia się metatezy olefin, taki wybór skutecznie spełniał rolę limitującą.

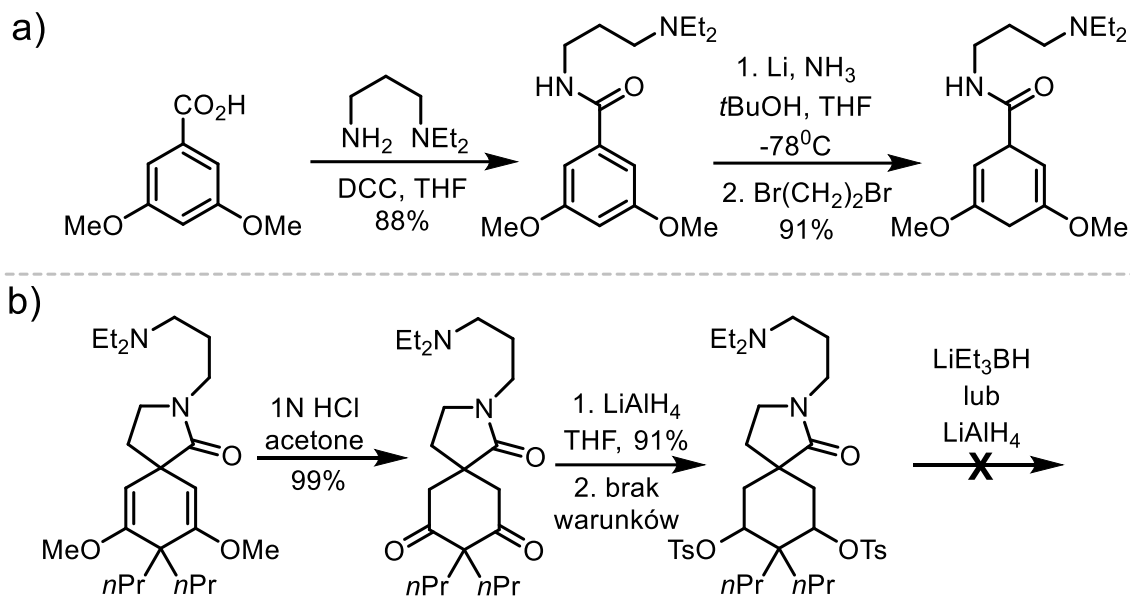
Pochodzący z pracy¹⁴² z 1992 roku przykład pochodnej naftalenu **AIH9.7** (Ryc.111.) pokazuje pewną osobliwość stosowanych reakcji. Z jednej strony proponowane przekształcenia są chemicznie poprawne, z drugiej jednak, proponowane przez SYNSUP warunki trudno zawsze nazwać typowymi. I tak dla acylowania Friedla-Craftsa **AIH9.1** → **AIH9.2**, jako środek acylujący proponuje się fluorek acetylu. Precedens literaturowy podkreśla selektywność tej reakcji³⁷¹, intencja twórców, by w bazie danych posiadać jak najwięcej selektywnych reakcji, rozmija się w tym miejscu z praktyką ze względu na małą dostępność i niską temperaturę wrzenia fluorku acetylu. Jeszcze ciekawsza jest propozycja eliminacji drugorzędowego alkoholu **AIH9.3** → **AIH9.4**, przy użyciu ThO₂. I znów, reakcja możliwa do wykonania, ale podpowiedzenie takich warunków nie jest najtrafniejsze. Eliminację tę łatwo wykonać w warunkach kwasowych lub alkohol przeprowadzić w ester mesyłowy i bez oczyszczania dokonać eliminacji mesyłanu. Również wykorzystanie organoolewiniowego związku jako donora grupy etylowej w **AIH9.6** → **AIH9.7** nie jest pierwszym z narzucających się warunków dla tej reakcji. To ciekawe, że o ile same przekształcenia i ich kolejność nie budzą poważnych zastrzeżeń, to już zaproponowane „najpopularniejsze warunki”, trudno takimi nazwać. Może być to świadectwo małej ilości reakcji w bazie danych, gdzie takie nietypowe acz selektywne warunki występują częściej. Interesujący jest sposób wprowadzenia centrum stereogenicznego, w transformacji **AIH9.4** → **AIH9.5**. Wykorzystano tu asymetryczne formylowanie Stille, co zasługuje na dwa komentarze. Po pierwsze, jest to pierwszy przykład katalizowanej kompleksem palladu reakcji w komputerowo generowanych analizach retrosyntetycznych. W żadnej z analiz LHASA czy SYNCHEM, takie reakcje nie występują. Niewykluczone, że znajdowały się w ich bazach danych, ale nie znalazło to odzwierciedlenia w publikowanych wynikach. Komentarz drugi ponownie nawiązuje do wielkości bazy danych, gdzie z konieczności ilość reakcji generujących centra stereogeniczne jest ograniczona. Widać to wyraźnie na tym przykładzie, redukcja ketonu metylo-arylowego **AIH9.2** → **AIH9.3** jest tu tylko etapem na drodze do uzyskania grupy winylowej. Tymczasem przegapiona zostaje dogodna okazja na asymetryczną redukcję takiego ketonu, dobrze opisanego dla redukcji Corey-Bakshi-Shibata³⁷²⁻³⁷⁴. Wydłużenie fragmentu węglowego wymagałoby na przykład mesyłowania, stereospecyficznej reakcji S_N2 za pomocą cyjanków i redukcji nitrylu za pomocą DIBALH. Tego typu sytuacje są typowe dla jeszcze małych eksperckich baz danych (ok. 2500 reakcji w tamtym momencie), gdzie znajdują się już specyficzne reakcje typu formylowania Stille, ale brakuje jeszcze popularnych reakcji typu redukcji CBS. Łatwo można to poznać przez próbę wygenerowania przez program nietypowego związku przejściowego, który jednak stanowi dlań jedyną możliwość na stworzenie zadanego motywu strukturalnego.



Ryc.112. Ilustracja różnych wariantów planów syntetycznych dla: azaspiranu **L16.1**; wygenerowanych przez program SYNSUP i ogłoszonych w pracy z 2006 roku¹⁴⁰.

Dla opisanej w głównym tekście syntezy azaspiranu **L16.1** przedstawiono (Ryc.112.) szereg alternatywnych dróg syntezy. Pozwala to ocenić stopień wszechstronności SYNSUP pod koniec jego rozwoju.

Podczas prowadzenia walidacji syntetycznej dla azaspiranu **L16.1**, próbowano alternatywnych metod syntezy wykorzystujących m.in. redukcję Bircha na późniejszym etapie (Ryc.113a). Skutkuje to jednak niewbudowaniem się do szkieletu fragmentu 1,2-dibromoetanu.



Ryc.113. Ilustracja nieudanych wariantów syntezy azaspiranu **L16.1** testowanych podczas walidacji planu inspirowanego analizą uzyskaną za pomocą programu SYNSUP i ogłoszonego w pracy z 2006 roku¹⁴⁰.

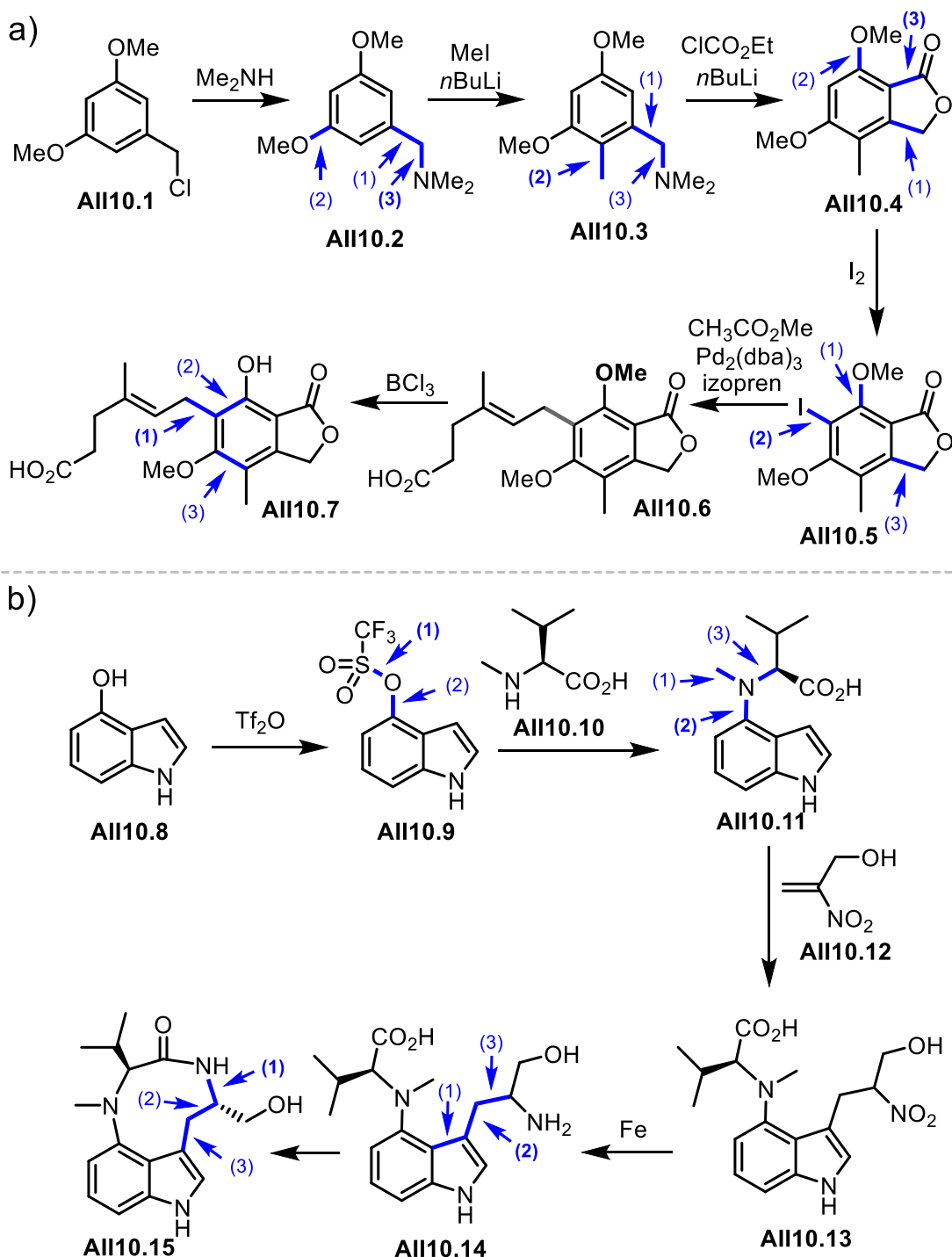
Podobnie problemy napotkano podczas prób redukcji grup ketonowych do metylenowych, wykorzystanie estrów tosylowych jako substratów do redukcji nie powiodło się (Ryc.113b).

W tekście głównym zamieszczono przykład planu syntetycznego dla kamptotecyny (Ryc.18.), pozostałe 2 związki, dla których udało się przy użyciu komputera odtworzyć plany syntez bardzo zbliżone do znanych w literaturze¹⁴³ to kwas mykofenolowy **AII10.7** i indolaktam **V AII10.15**. Są to zarazem ostatnie przykłady planów syntetycznym z SYNSUP, toteż analiza rodzajów transformacji pozwala na ocenę do jakiego stopnia rozwinięto program i jaka jakość analiz kryła się za niespełna 6000 reguł reakcji.

Kwas mykofenolowy **AII10.7** – pochodna fenolu o właściwościach przeciwnowotworowych, przeciwwirusowych, przeciwpasożytniczych i immunosupresyjnych, wymaga w syntezie skutecznej kontroli reaktywności pierścienia fenyloвого. Tak jak wcześniejsze programy w zakresie reakcji specyficznych dla układów aromatycznych, wykazywały poważne braki, tak SYNSUP proponuje tu interesujące i warte dyskusji transformacje (Ryc.114a). Dyskusyjną należy nazwać propozycję **AII10.2** → **AII10.3**, dyskusja nad tym metylowaniem sprowadza się do pytania, która z dwóch pozycji na pierścieniu aromatycznym będzie lepiej stabilizowała arylolitowy stan przejściowy. Czy wpływ obu grup metoksylowych, czy też grupy metoksylowej i na drugiej flance dimetyloaminometylenowej – będzie ważniejszy. Odpowiedzi udziela (na korzyść zaproponowanej przez SYNSUP) praca³⁷⁵ dla ortolitowania związku **AII10.2**, przykładów na przeciwną regioselektywność brak w literaturze. Po zidentyfikowaniu możliwości regioselektywnego przebiegu ortolitowania, funkcjonalizowano pozostałą pozycję **AII10.3** → **AII10.4**, w celu uzyskania laktonu. Przyłączenie izoprenowego fragmentu wymaga aktywacji pozostałego protonu arylowego, w tym celu jodowanie należy przeprowadzić w ostrożny, elektrofilowy sposób, gdyż obecność laktonu wyklucza ponowne użycie ortolitowania. Dla reakcji podobnej do proponowanej **AII10.5** → **AII10.6** istnieje również odnotowany przez autorów precedens literaturowy. Na pierwszy rzut oka, można mieć zastrzeżenia co do selektywności demetylowania **AII10.6** → **AII10.7**. Jednak obecność różnicującej grupy metoksylowej tlenu z grupy karbonylowej laktonu, pozwala na selektywne demetylowanie dla kwasu mykofenolowego,

jak wcześniej wykazano dla jego szkieletu³⁷⁶⁻³⁷⁷. Należy w tym miejscu odnotować postęp jakiego w zakresie implementacji reaktywności związków aromatycznych dokonali twórcy programu, w porównaniu z omówionymi wcześniej rezultatami osiąganymi na początku lat 90-tych.

Natomiast skrócenia czasu analizy jest tu jeszcze większe niż w przypadku kamptotecyny. Pierwotnie wygenerowanie 30 planów o maksymalnej długości 7 etapów, zajmowało ponad 35 godzin. Po wprowadzeniu typowania wiązań strategicznych, czas analizy skrócono do 60 sekund, generując 2 plany o maksymalnej długości 6 etapów.



Ryc.114. Ilustracja planów syntetycznych dla: a) kwasu mykofenolowego; b) indolaktamu V; wygenerowane przez program SYNSUP i ogłoszonego w pracy z 2009 roku¹⁴³.

Indolaktam V **AIII10.15** – związek o właściwościach kancerogennych, łączący pierścień heterocykliczny z 9-członowym pierścieniem makrocyklicznym udekorowanym dwoma centrami stereogenicznymi. W tym planie swobodnie wykorzystywana jest reakcja Buchwalda-Hartwiga **AIII10.9** + **AIII10.10** → **AIII10.11**, pozwalająca wprowadzić centrum stereogeniczne za pomocą pochodnej waliny. Przejście **AIII10.11** + **AIII10.12** → **AIII10.13** budzi większe zastrzeżenia, po pierwsze grupa hydroksylowa powinna być zabezpieczona, na co również autorzy zwracają uwagę. Ponadto w tej reakcji generowane jest centrum stereogeniczne przy grupie nitrowej. Zaproponowana reakcja nie pozwala na przeprowadzenie tego procesu w sposób stereokontrolowany. Amidowanie **AIII10.14** → **AIII10.15** jako reakcji zamknięcia pierścienia makrocyklicznego indolaktamu, zostało wykorzystane w jego syntezie totalnej z 2019 roku³⁷⁸.

Przejścia zaproponowane przez SYNSUP dla indolaktamu mogą budzić wątpliwości, jednak tego typu reakcje można skorygować nie skreślając całego planu, jak to było przy pierwszych rezultatach z lat 90-tych. Zastosowanie typowania wiązań poskutkowało skróceniem czasu pracy z ponad 25 godzin dla 47, maksymalnie 5 etapowych planów, do 15 sekund dla wygenerowania 2 planów, maksymalnie 7 etapowych.

Na powyższych przykładach widać postęp, jaki dokonał się w ciągu dekad rozwoju programu SYNSUP, zarówno jeżeli chodzi o czas potrzebny na wygenerowanie odpowiedzi, jak i jakość analiz retrosyntetycznych.



B. Org. 417/20

Biblioteka Instytutu Chemii Organicznej PAN

Org.-B.417/20



80000000341314