

PŘIHLÁŠKA VYNÁLEZU

zveřejněná podle § 31 zákona č. 527/1990 Sb.

(21) Číslo dokumentu:

2003 - 1090

(19)

ČESKÁ
REPUBLIKA



ÚŘAD
PRŮMYSLOVÉHO
VLASTNICTVÍ

(22) Přihlášeno: **16.10.2001**

(32) Datum podání prioritní přihlášky: **17.10.2000**

(31) Číslo prioritní přihlášky: **2000/00309114**

(33) Země priority: **EP**

(40) Datum zveřejnění přihlášky vynálezu: **14.01.2004**
(Věstník č. 1/2004)

(86) PCT číslo: **PCT/EP01/11955**

(87) PCT číslo zveřejnění: **WO02/033596**

(13) Druh dokumentu: **A3**

(51) Int. Cl. ⁷:

G 06 F 17/50

(71) Přihlašovatel:

APPLIED RESEARCH SYSTEMS ARS HOLDING N.
V., Curacao, AN;

(72) Původce:

Colinges Jacques, Neydens, FR;
Church Dennis, Commugny, CH;

(74) Zástupce:

Korejzová Zdeňka JUDr., Spálená 29, Praha 1, 11000;

(54) Název přihlášky vynálezu:

**Způsob provozování počítačového systému pro
provádění diskrétní podstrukturní analýzy**

(57) Anotace:

Způsob provozování počítačového systému pro provádění diskrétní podstrukturní analýzy, který zahrnuje kroky přístupu (210, 220, 410) k databázi (110, 115) molekulárních struktur, přičemž tato databáze je prozkoumatelná prostřednictvím informací o molekulárních strukturách a biologických a/nebo chemických vlastností; identifikování (220) v uvedené databázi podskupiny molekul majících danou biologickou a/nebo chemickou vlastnost; stanovení (230, 420) fragmentů molekul v uvedené podskupině; pro každý fragment vypočítání (230, 430, 610-650) hodnoty skóre, indikující příspěvek příslušného fragmentu k uvedené dané biologické a/nebo chemické vlastnosti; a provedení (240, 250) opakování procesu analyzováním (250) stanovených fragmentů a vypočítaných hodnot skóre, přičemž nejprve se vybere alespoň jeden fragment, který má hodnotu skóre, indukující vysoký příspěvek k uvedené biologické a/nebo chemické vlastnosti, a potom se opakují kroky přístupu, identifikování, stanovení a vypočítání.



Způsob provozování počítačového systému pro provádění diskrétní podstrukturní analýzy

Oblast techniky

5 Předkládaný vynález se týká počítačového systému a způsobu jejího provozování za účelem provádění diskrétní podstrukturní analýzy. Tato analýza umožňuje provádění počítačově realizované identifikace molekul majících určité vlastnosti, jako je biologická a/nebo chemická aktivita. Počítačově řízená diskrétní podstrukturní analýza může být
10 využita při objevování léků nebo v jiných oblastech, kde jsou středem zájmu biologicky, farmakologicky, toxikologicky, pesticidně, herbicidně, katalyticky nebo jinak aktivní sloučeniny.

Dosavadní stav techniky

Pokrok v oblasti například lékařské chemie závisí na identifikaci biologicky aktivních molekul. V mnoha případech jsou výzkumné programy zaměřeny na syntézu malých organických molekul, které budou interagovat se známým enzymatickým nebo
20 receptorovým terčem, aby se vytvořil požadovaný farmakologický účinek. Takové sloučeniny mohou, alespoň částečně, napodobovat nebo zabraňovat v aktivitě známých, přirozeně se vyskytujících substancí, ale jsou určeny pro zajištění silnějšího a/nebo selektivnějšího působení. Sloučeniny, vznikající z tohoto typu výzkumu, mohou zahrnovat
25 určité strukturní znaky relevantních, přirozeně se vyskytujících substancí.

Výzkumné programy mohou být rovněž založeny na
30 přirozeně se vyskytujících sloučeninách nalézáných jako

výsledky zdrojů screeningu, dostupných v přírodě, jako jsou například vzorky půd nebo rostlinné extrakty.

V nedávných letech se zvýšil tlak na identifikování nových a použitelných biologicky aktivních molekul a v důsledku toho byly vyvinuty nové metody generování sloučenin. V tomto ohledu měly obzvláštní důležitost dva vývoje, jmenovitě kombinatorní chemie a vysokovýkonný screening (HTS).

Kombinatorní chemie využívá robotické nebo manuální techniky pro provádění množství menších chemických reakcí, z nichž každá využívá odlišnou kombinaci reagentů, současně nebo "paralelně", čímž se generuje velké množství různých chemických entit pro screening. Souhrn sloučenin, generovaných touto metodou, je známý jako "knihovna". Knihovny pro generování nových chemických vodičů jsou obvykle tak rozmanité, jak jen je možné. Za určitých okolností ale mohou být knihovny nakloněny nebo zacíleny k určitému farmakologickému terči či cíli, nebo mohou být zaměřeny na určitou chemickou oblast, prostřednictvím volby reagentů určených pro zavedení specifických strukturních znaků do finálních sloučenin.

Vysokovýkonný screening zahrnuje využití biochemické analýzy pro rychlé testování aktivity velkých počtů chemických sloučenin in vitro vzhledem k jednomu nebo více biologickým terčům. Tato metoda je ideální pro screening velkých knihoven sloučenin, vytvořených kombinatorní chemií.

Navzdory nezpochybnitelným výhodám kombinatorní chemie a HTS při generování nových vodících struktur, existují u těchto metod určité nevýhody. Vysoký podíl

5

10

25

30

příslušný příspěvek každého jednotlivého prvku k aktivitě. Navíc je vypočítán také průměrný příspěvek každé skupiny substituentů v určitém bodě k biologické aktivitě. Je rovněž uveden příklad jak takovýto příspěvek vypočítat.

5 H. Gao a kol., J. Chem. Inf. Comput. Sci. (39) 1999, 164 až 168 je článek popisující aplikaci techniky QSAR (kvantitativních vztahů strukturní aktivity) pro problém objevování léků. Poté, co jsou zvoleny biologicky aktivní sloučeniny, je optimalizována jejich biologická aktivita. 10 Protože technika QSAR je založena na hypotetickém vztahu mezi biologickou aktivitou a molekulárními strukturami, zabývá se tato technika identifikováním strukturních vlastností, které činí sloučeniny aktivními, a předpovídáním aktivních a neaktivních analogů.

15 WO 00/41060 popisuje způsob korelování aktivit substancí se strukturními znaky substancí. Termín "znak" se týká atomů a vazeb struktury, která odpovídá předloze. V prvním kroku jsou stanoveny prvky sady substancí, které splňují daná omezení strukturních znaků a vlastností. Potom 20 pro každou kategorii aktivit jsou zkonstruovány substance, které spadají do této kategorie. Po rozdělení sady substancí mezi několik kategorií aktivit se vypočítá očekávaná aktivita pro jakoukoliv podsadu a pro každý strukturní znak se zkonstruuje sada bitových vektorů pro aktivitu, vlastnost a 25 znak, které označují počty substancí, které obsahují uvedený znak a jsou v uvedené kategorii aktivit. Tento dokument se týká biologických aktivit a rovněž se zabývá objevováním léků.

30 US 6,185,506 B1 popisuje způsob volby optimální rozmanité knihovny malých molekul na základě ověřených

molekulárních strukturních deskriptorů. Jsou použity sady údajů z množství literatury, které obsahují nejrozličnější chemické struktury a přidružené aktivity. Aktivitou může být biologická a chemická aktivita. Technika je popsána v kontextu farmakologických léčiv. Navíc je popsán způsob výběru podskupiny molekul produktu pro všechny možné molekuly produktu, které by mohly být vytvořeny v kombinatorní syntéze ze specifikovaných reakčních molekul a společných jádrových molekul. V části, popisující dosavadní stav techniky, je uveden odkaz na biologicky specifické knihovny, které byly zkonstruovány na základě znalostí o geometrických uspořádáních strukturních fragmentů abstrahovaných z molekulárních struktur, o nichž je známo, že jsou aktivní. V tomto dokumentu je uvedeno, že je naprosto nezbytné použít menší racionálně konstruovanou screeningovou knihovnu, která ještě zachová rozmanitost kombinatorně dostupných sloučenin.

WO 00/49539 A1 popisuje způsob screeningu sady molekul pro identifikování sady molekulárních znaků, které pravděpodobně korelují se specifikovanou aktivitou. Termín znak se zde týká chemických podstruktur. Sada molekul je roztržena do skupin podle jejich molekulární struktury, jak je charakterizována sadou deskriptorů. Potom jsou identifikovány skupiny, které reprezentují nejvyšší úroveň aktivity, a jsou nalezeny nejběžnější podstruktury mezi molekulami ve skupinách, které tak mohou být přiměřeně korelovány se zjištěnou úrovní aktivity. Je vytvořena sada údajů, která reprezentuje tyto molekuly, z počáteční sady údajů, která obsahuje společnou podskupinu znaků. Tato technika je zde popsána v podobě systému na bázi počítačů pro provádění automatizované analýzy sady údajů.

US 5,463,564 popisuje postup na bázi počítačů pro automatické generování sloučenin prostřednictvím robotického sestavování (syntézy) a analyzování množství chemických sloučenin. Tento proces se provádí iteračně a zaměřuje se na generování chemických entit s definovanými vlastnostmi aktivit. Je sestavena knihovna chemických sloučenin s řízenou rozmanitostí, která zahrnuje množství chemických sloučenin. Údaje o strukturní aktivitě jsou získány robotickou analýzou sestavených sloučenin. Je rovněž popsáno množství databází, z nichž každá obsahuje pole indikující hodnotící faktor přidělený příslušné sloučenině. Hodnotící faktor je přidělen každé sloučenině na základě toho, jak přesně se aktivita této sloučeniny shoduje s požadovanou aktivitou.

Shora popisované postupy jsou buď "prediktivními" modely nebo stále selhávají při uspokojivém zlepšení generování aktivních vodících sloučenin a zvyšování pravděpodobnosti nalezení aktivních sloučenin v dané sadě sloučenin. Navíc tyto známé, běžné techniky nejsou schopné splnit potřebu zvýšeného množství a kvality aktivních a vodících sloučenin, které vstupují do vývojového řetězce.

Cílem předkládaného vynálezu je tudíž navrhnout způsob provozování počítačového systému, a také navrhnout odpovídající počítačový systém, pro dosažení zvýšené možnosti objevení nových, biologicky a/nebo chemicky aktivních molekul.

Podstata vynálezu

Uvedený cíl je vyřešen podle předkládaného vynálezu řešeními definovanými v nezávislých patentových nárocích.

- Obr.5 znázorňuje graf ilustrující, jak mohou být voleny fragmenty na základě vypočítaných hodnot skóre;
- 5 Obr.6 znázorňuje vývojový diagram ilustrující proces výpočtu hodnoty skóre pro fragment podle výhodného provedení předkládaného vynálezu;
- 10 Obr.7 znázorňuje vývojový diagram ilustrující proces analyzování knihovny fragmentů při provádění opakování (re-iterace);
- Obr.8 znázorňuje vývojový diagram ilustrující proces volby nové sloučeniny prostřednictvím využití generických podstruktur;
- 15 Obr.9 znázorňuje vývojový diagram ilustrující proces generování podstruktur pro použití při virtuálním screeningu;
- 20 Obr.10 znázorňuje vývojový diagram ilustrující proces analyzování knihovny fragmentů při provádění opakování, aplikováním techniky vázání za zvýšené teploty (slučování - annealing) podle výhodného provedení vynálezu;
- 25 Obr.11 znázorňuje příklad mapy relativních příspěvků pro ilustraci slučovací techniky aplikované v procesu podle obr. 10;
- 30 Obr.12 znázorňuje graf ilustrující účinek sloučeniny na receptorem zprostředkované vytvoření inositoltrifosfátu;

Obr.13 znázorňuje graf ilustrující účinek sloučeniny na kináze závislé fosforylaci bílkovin;

Obr.14 znázorňuje graf ilustrující účinek sloučeniny na fosfatáze závislé defosforylaci bílkovin;

5 Obr.15 znázorňuje graf ilustrující informace o relativních příspěvcích vynesení determinantů vzhledem k jejich příslušným hodnotám skóre; a

10 Obr.16A až obr. 16H znázorňují další grafy relativního přispění, demonstrující rovnocennost funkcí skóre.

Příklady provedení vynálezu

15 Předkládaný vynález bude nyní popsán detailněji. Navíc budou diskutována výhodná provedení předkládaného vynálezu ve spojení s odkazy na připojené výkresy. Také bude ještě uvedeno množství příkladů, jak předkládaný vynález může být aplikován v množství oblastí objevování či nalézání sloučenin.

20 Podle předkládaného vynálezu je provozován počítačový systém pro provádění diskrétní podstrukturní analýzy. Je přistupováno k databázi molekulárních struktur. Tato databáze je prozkoumatelná prostřednictvím molekulárních informací a biologických a/nebo chemických vlastností. Informacemi o
25 molekulární struktuře jsou jakékoliv informace vhodné pro určení molekulární struktury molekuly. Biologické a/nebo chemické vlastnosti zahrnují biochemické, farmakologické, toxikologické, pesticidní, herbicidní a katalytické
30 vlastnosti.

S použitím databáze technika podle předkládaného vynálezu identifikuje podskupinu molekul majících danou biologickou a/nebo chemickou vlastnost. V uvedené podskupině jsou potom stanoveny fragmenty molekul. Termín "fragment" se
5 týká jakékoliv strukturní podjednotky molekuly, včetně jednoduchých funkčních skupin, dvourozměrných podstruktur a jejich rodin, jednoduchých atomů nebo vazeb, a jakékoliv sestavy strukturních deskriptorů ve dvourozměrném nebo třírozměrném molekulárním prostoru. Osoby v oboru znalé
10 snadno nahlédnou, že fragmentem může být molekulární podstruktura, která nemá žádný známý význam v běžné chemii.

Poté, co molekulární struktury v podskupině jsou rozděleny do fragmentů, je vypočítána hodnota skóre pro každý fragment, která indikuje příspěvek příslušného fragmentu k
15 dané biologické a/nebo chemické vlastnosti. To znamená, že předkládaný vynález umožňuje přidělení hodnot skóre k fragmentům na základě existujících znalostí s ohledem na biologické a/nebo chemické vlastnosti molekul. V následujícím popisu bude molekula, struktura nebo podstruktura označena za
20 "aktivní", pokud bude mít danou vlastnost. Molekula, struktura nebo podstruktura, která není aktivní bude označena jako "neaktivní". Předkládaný vynález tudíž poskytuje podstrukturní analýzu založenou na diskrétních informacích o biologických a/nebo chemických vlastnostech. Hlavní postup
25 podle předkládaného vynálezu je tudíž níže nazýván diskrétní podstrukturní analýzou (DSA).

Protože podle předkládaného vynálezu jsou fragmenty sdružovány s hodnotami skóre, které indikují jejich příspěvek
30 k dané biologické a/nebo chemické vlastnosti, je možné fragmenty považovat za chemické determinanty odpovědné za

systémem schopným souběžného zpracování více úloh (multitasking).

Základní jednotka 100 je spojena s programovou pamětí 130, která ukládá proveditelný programový kód včetně instrukcí pro provádění DSA procesu podle předkládaného vynálezu. Tyto instrukce zahrnují funkce 135 fragmentace pro rozdělování molekulárních struktur do fragmentů, funkce 140 skóre pro výpočet hodnot skóre, funkce 145 zobecnění (například pro nalezení izomerů) pro lokalizování zobecnitelných položek ve strukturách fragmentů a pro náhradu těchto položek zobecněnými výrazy, čímž se vytvářejí generické podstruktury, funkce 150 virtuálního screeningu pro provádění virtuálního screeningu, a funkce 155 slučování (vázáni za zvýšené teploty - annealing) pro provádění procesu slučování fragmentů podle předkládaného vynálezu. Details o jednotlivých funkcích a procesorech realizovaných prostřednictvím základní jednotky 100 při provádění těchto funkcí budou popsány podrobněji níže.

Základní jednotka 100 je dále spojena s databází 115 aktivity struktur, nebo také se seznamem aktivity sloučenin, pro vyhledávání informací o molekulárních strukturách a informací o biologických a/nebo chemických vlastnostech. Tyto informace mohou být podobně přijímány z jednotky 110 pro vstup dat, která umožňuje přístup k vnějším datovým zdrojům.

Přístupem k jednotce 110 a/nebo k databázi 115 může být získána podskupina molekulárních struktur, například z jakéhokoliv dostupného zdroje, jako jsou soukromé nebo veřejné databáze, které jsou prozkoumatelné prostřednictvím podstrukturních a/nebo biologických vlastností. Veřejné databáze zahrnují, ale nejsou omezeny na, databáze dostupné

pod následujícími jmény: MDDR, Pharmaprojects, Merck Index, SciFinder, Derwent. Podskupina molekul může být rovněž získána sestavováním a testováním sloučenin. Molekuly budou obecně zahrnovat úplné sloučeniny, ale mohou rovněž samy být
5 molekulárními fragmenty. Pro jakoukoliv danou biologickou nebo chemickou vlastnost podskupina obsahuje sloučeniny, které nemají tuto vlastnost, například sloučeniny, které nejsou aktivní (nebo spadají pod danou prahovou hodnotu aktivity), a rovněž sloučeniny, které mají tuto vlastnost,
10 například sloučeniny, které vykazují požadovanou aktivitu (to jest mají aktivitu nad danou prahovou hodnotou). Všechny neaktivní sloučeniny jsou relevantní a jsou tudíž analyzovány.

Po přístupu k vnitřním nebo vnějším datům a provedení
15 DSA procesu s využitím funkcí uložených v programové paměti 130 základní jednotka 100 ukládá knihovnu 120 fragmentů, která obsahuje stanovené fragmenty molekul společně s jejich přidruženými hodnotami skóre.

V jednom výhodném provedení podle předkládaného
20 vynálezu je knihovna 120 fragmentů výsledkem hlavního procesu podle vynálezu. Knihovna 120 fragmentů může potom být použita například chemickými nebo biologickými vědci nebo inženýry jako zdroj hodnotných informací, které jsou využitelné v
25 jakémkoliv následujícím výzkumném procesu.

V jiném výhodném provedení je knihovna 120 fragmentů přechodným výsledkem hlavního procesu podle předkládaného vynálezu a může tudíž být uložena v energeticky závislé (nestálé) paměti a rovněž v energeticky nezávislé (trvalé)
30 paměti. Knihovna 120 fragmentů podle tohoto provedení může být čtena základní jednotkou 100 při provádění dalších funkcí

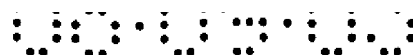
uložených v programové paměti 130 pro generování souboru 125 sloučenin.

Soubor 125 sloučenin je soubor molekul, které byly nalezeny prostřednictvím procesu podle předkládaného vynálezu a mají nebo nemají požadovanou biologickou a/nebo chemickou vlastnost. Molekuly souboru 125 sloučenin mohou být buď již známé nebo mohou být hypotetickými strukturami, které předtím ještě nebyly sestaveny. V každém případě jsou molekuly souboru 125 sloučenin výsledkem vyhodnocení hodnot skóre, přidělených fragmentům podle diskrétní podstrukturní analýzy.

Jak může být patrné z obr. 1, základní jednotka 100 je dále spojena s datovou pamětí 160, která ukládá sady 165 sloučenin, sady 170 fragmentů a hodnoty 175 skóre. Datová paměť 160 je vytvořena pro ukládání dat či údajů používaných pro ukládání vstupních parametrů při vyvolávání funkcí 135 až 155 procesu, nebo pro ukládání vrácených hodnot z těchto funkcí.

Jak je patrné na obr. 2, který ilustruje výhodné provedení hlavního DSA procesu, obsluha počítačového systému ilustrovaného na obr. 1 nejprve v kroku 210 vybírá aktivitu. Jak bylo zmiňováno výše, aktivita znamená jakoukoliv biologickou a/nebo chemickou vlastnost, včetně biochemických, farmakologických, toxikologických, pesticidních, herbicidních a katalytických vlastností. Navíc při použití vynálezu pro identifikování samostatných ligandů může být aktivitou daný účinek na příslušnou bílkovinu (obvykle schopnost vázání - vytváření vazby).

V předkládaném popisu odkaz na určitou vlastnost, jako je biologická aktivita, může být, pokud není uvedeno



jinak, extrapolován na další typy biologických a/nebo chemických vlastností. Dále pro vyloučení jakýchkoliv pochyb termíny "sloučenina", "molekula" a "molekulární struktura" mohou všechny zahrnovat molekulární podstruktury a rovněž celé sloučeniny podle daného kontextu.

Poté, co v kroku 210 již byla zvolena aktivita, je v kroku 220 vybrán soubor 125 sloučenin. Zvolený soubor sloučenin je souborem molekul, které mají být zkoumány pro zjištění, které fragmenty přispívají ke zvolené aktivitě. Jak bude podrobněji popsáno níže, soubor sloučenin, zvolený v kroku 220, obsahuje molekuly, o nichž je známo, že jsou aktivní, a molekuly, o nichž je známo, že jsou neaktivní.

Jakmile již byla zvolena aktivita a soubor sloučenin, proces pokračuje v kroku 230 vytvářením knihovny 120 fragmentů. Proces vytváření knihovny fragmentů může být popsán jako proces vážení účinnosti molekulárních fragmentů, uvnitř podskupiny známých struktur, na chemický a/nebo biologický výsledek. Tento proces může být popsán jako sestávající z kroků:

I. identifikování jedné nebo více podskupin molekul majících dané vlastnosti vzhledem k příslušnému chemickému a/nebo biologickému výsledku;

II. vytvoření předběžné knihovny zahrnující fragmenty molekul v uvedené jedné nebo více podskupinách;

III. aplikování algoritmu pro odhad příspěvku uvedených fragmentů ve vztahu k příslušnému chemickému a/nebo biologickému výsledku; a

IV. získání hodnot skóre pro každý uvedený fragment, na který byl aplikován uvedený algoritmus, přičemž

tyto hodnoty skóre mohou být seřazeny podle velikosti; přičemž ty fragmenty, které s největší pravděpodobností přispívají k příslušnému chemickému a/nebo biologickému výsledku, jsou sdruženy například s vysoce zařazenými hodnotami skóre.

Jak bylo zmiňováno výše, knihovna 120 fragmentů obsahuje fragmenty a rovněž získané hodnoty skóre fragmentů. Jakmile již byla v kroku 230 vytvořena knihovna 120 fragmentů, proces může nebo nemusí provést krok 240 opakování (re-iterace).

Realizováním DSA procesu opakovacím způsobem mohou být výpočetní zdroje využity velmi efektivně. Například může proces výhodně začínat s malými fragmenty. Protože počet možných fragmentů v molekulárních strukturách se zvyšuje přibližně exponenciálně s maximální velikostí fragmentů, které jsou zkoumány, je tato maximální velikost na začátku nastavena na spíše nízkou hodnotu, takže může být zpracován dokonce velmi vysoký počet molekulárních struktur.

Proces v krocích 210 až 230 odhaluje fragmenty s vysokým příspěvkem k požadované aktivitě. Nalezené fragmenty mohou být potom použity v následujícím kole (nebo cyklu) pro nalezení fragmentů větší velikosti, to jest s vyšší molekulární hmotností. Příklad opakovacího procesu je znázorněn na obr. 3. V prvním kole byl nalezen fragment $C=O$ jako mající vysoký příspěvek k požadované aktivitě. Tento fragment je potom použit pro zkoumání fragmentů, které mají větší velikost, než výsledný fragment prvního kola, a které obsahují tento fragment. V příkladu podle obr. 3 druhé kolo ukázalo, že vzhledem k požadované aktivitě je nejlepším fragmentem této velikosti fragment $N-C=O$. Tento opakovací

proces pak pokračuje, čímž se zvětšuje velikost fragmentů, a může vést na sloučeninu, která pravděpodobně má požadovanou biologickou a/nebo chemickou vlastnost a je vhodná pro požadovanou aplikaci.

5 Při opětovném popisu procesu podle obr. 2, pokud je v kroku 240 rozhodnuto o provedení dalšího kola nebo cyklu, je v kroku 250 analyzovaná knihovna 120 fragmentů vytvořená v kroku 230 a proces se vrací do kroku 220. Příklady toho, jak je v kroku 250 analyzována knihovna 120 fragmentů, budou
10 podrobněji popsány níže. Jak lze očekávat, opakovací proces umožňuje aplikování pokročilejších funkcí, jako jsou funkce 145 zobecnění a funkce 155 slučování pro další zlepšení procesu objevování s využitím diskrétní podstrukturní analýzy.

15 Nakonec, když je v kroku 240 rozhodnuto, že již nemá být provedeno žádné další opakování, proces přechází k ukončení, přičemž v kroku 260 se vytváří soubor 125 sloučenin.

20 Pokud se nyní opět týká kroku 230 vytváření knihovny 120 fragmentů, bude níže ve spojení s odkazy na obr. 4 až obr. 6 popsáno výhodné provedení podkroků tohoto vytvářecího procesu. Nejprve poté, co je přistoupeno k vnitřní databázi 115 a/nebo k vnějšímu datovému zdroji a jsou identifikovány
25 podskupiny molekul, jsou v kroku 410 přijata data aktivity struktur, která se týkají identifikovaných molekul. Potom jsou v kroku 420 stanoveny fragmenty molekul v podskupině.

Molekuly mohou být rozdělovány do fragmentů s využitím množství běžných technik. Například může být použit
30 algoritmus pro nalezení jakékoliv permutace atomů, které jsou

než pouhý výběr fragmentů s nejvyšším skóre porovnáním hodnot skóre, protože graf navíc využívá informace o počtu sloučenin, které obsahují příslušné fragmenty.

5 Proces nalézání největší možné hodnoty skóre může být považován za ekvivalentní vytváření fylogenní sítě hierarchicky uspořádaných molekulárních fragmentů, odpovídající dané biologické a/nebo chemické aktivitě. V tomto uspořádání jsou uzly sítě tvořeny samotnými fragmenty a pravděpodobnost, že jakýkoliv jeden fragment je základem
10 biologické aktivity, je dána vzdáleností odpovídajícího uzlu od počátku, to jest základny samotné sítě. Tedy čím větší hodnota skóre je pro jakýkoliv daný fragment, tím dále je odpovídající uzel od počátku mřížky a tím pravděpodobnější je, že tento fragment reprezentuje chemické řešení pro
15 například farmakofor, který je rozpoznáván příslušným daným terčem.

Nyní bude ve spojení s odkazy na obr. 6 podrobněji popsán krok 430 výpočtu skóre fragmentů. Aplikace funkcí 140 skóre odpovídá výše zmiňované soustavě logických pravidel nebo výpočetních kroků. DSA postup podle předkládaného vynálezu zahrnuje ve výhodném provedení krok začlenění proměnných, týkajících se převahy každého fragmentu, do jedné nebo více matematických funkcí, které odhadují hodnotu skóre
20 pro jakýkoliv daný fragment.

25 Uvedený algoritmus je funkcí:

(a) počtu molekul x v podskupině, které splňují danou prahovou hodnotu vzhledem k požadovanému výsledku a které obsahují daný fragment;



(b) počtu molekul y v uvedené podskupině, které obsahují uvedený fragment, ať již splňují či nesplňují uvedenou prahovou hodnotu;

5 (c) počtu molekul z v uvedené podskupině, které splňující uvedenou prahovou hodnotu, ať již obsahují nebo neobsahují uvedený fragment; a

(d) počtu N všech molekul v podskupině.

10 Výsledkem, zmiňovaným v (a), může být jakýkoliv požadovaný parametr týkající se aktivity sloučenin, což zahrnuje, ale není omezeno na, biologickou, biochemickou, farmakologickou a/nebo toxikologickou aktivitu. Každá
15 sloučenina nebo molekula v datovém souboru může potom být analyzována podle toho, zda má požadovaný parametr ve vztahu k dané prahové hodnotě, jako je určitá úroveň aktivity. Prahová hodnota může být nastavena na jakoukoliv požadovanou úroveň. V následujícím popisu je "aktivní" sloučeninou taková
20 sloučenina, která splňuje požadovanou prahovou hodnotu, a "neaktivní sloučeninou" je taková sloučenina, která nesplňuje uvedenou prahovou hodnotu. Tyto termíny nejsou určeny pro vyjádření jakékoliv absolutní vlastnosti příslušných sloučenin.

25 Příspěvek daného fragmentu může být stanoven aplikováním proměnných x , y , z a N , mírou nebo měřítkem asociace nebo funkcí 140 skóre. Jak je dobře známo osobám v oboru znalým, existuje mnoho možných měřítek asociace, která spadají do tří hlavních kategorií:

30 Odčítací měřítko: např.

$Nx-yz$;



Poměrová měřítka: např. $x(N-y-z-x)/(z-x)(y-x)$;

Smíšená měřítka: např. $(x/z) - (z-x)/(N-z)$.

5 Mělo by být zcela zřejmé, že může být zvoleno jakékoliv měřítko asociace a osoby v oboru znalé budou snadno schopné provést vhodnou volbu.

Algoritmus aplikovaný v kroku 430 tudíž může zahrnovat (viz obr. 6):

10 (i) vyhodnocení počtu sloučenin x v podskupině, které splňují danou prahovou hodnotu vzhledem k příslušnému chemickému nebo biologickému výsledku a které obsahují daný chemický determinant (krok 610);

15 (ii) vyhodnocení počtu sloučenin y v uvedené podskupině sloučenin, které obsahují uvedený chemický determinant, ať již splňují nebo nesplňují uvedenou prahovou hodnotu (krok 620);

20 (iii) vyhodnocení počtu sloučenin z v uvedené podskupině sloučenin, které splňují uvedenou prahovou hodnotu, ať již obsahují nebo neobsahují uvedený chemický determinant (krok 630);

(iv) vyhodnocení celkového počtu N v podskupině sloučenin (krok 640); a

25 (v) aplikování měřítka asociace na dvě nebo více proměnných x , y , z , a N (krok 650), výhodně na tři nebo čtyři proměnné a zvláště výhodně na všechny čtyři proměnné x , y , z , a N .

30 Měřítka asociace může být aplikováno přímo pro stanovení hodnoty skóre, odpovídající příspěvku daného

fragmentu. Výhodně je ale měřítko asociace rozvinuto do funkce skóre, aby se vyhodnotila pravděpodobnost, že podstruktura přispívá k výsledku. To umožňuje jasnější stanovení seřazení hodnot skóre, získaných pro všechny analyzované fragmenty. Měřítko asociace může být rozvinuto do funkce skóre prostřednictvím postupů obecně dobře známých v oboru. Tyto postupy mohou být například výhodně zvoleny ze statistických metod, jako je například metoda kritického poměru (z); Fisherův přesný test, Personova metoda chi-kvadrátu; Mattel-Haenzelova metoda chi-kvadrátu; a metody založené na, ale neomezené na, provádění odhadů strmosti, a podobně. Mohou být ale použity i metody jiné než statistické. Takové metody zahrnují, ale nejsou omezeny na, výpočet a porovnání přesných a přibližných konfidenčních intervalů, korelačních součinitelů, nebo skutečně jakékoliv funkce obsahující měřítko asociace sestávající z kombinace jedné, dvou, tří nebo čtyř z proměnných x , y , z , a N popisovaných výše.

Příklady matematických vzorců reprezentujících měřítko asociace nebo funkce skóre, které mohou být použity v předkládaném vynálezu, zahrnují:

25

(I)

$$x/z$$

(II)

$$x/N$$

(III)

$$Nx-yz$$

30

(IV)

$$(x/z)-(y/N)$$



	(V)	$(x/z) - (z-x)/(N-z)$
5	(VI)	$\frac{x(N-y-z+x)}{(z-x)(y-x)}$
	(VII)	$\frac{Nx-yz}{\sqrt{z(N-z)y(N-y)}}$
10	(VIII)	$e^{[(x/z) - (z-x)/(N-z)]}$
	(IX)	$\frac{(Nx-yz - N/2)^2 N}{z(N-z)y(N-y)}$
15	(X)	$\frac{x(N-y-z+x)}{(z-x)(y-x)} e^{-2\sqrt{1/x + 1/(y-x) + 1/(z-x) + 1/(N-y-z+x)}}$
	(XI)	$\frac{x_1(N-y-z_1+x_1)(z_2-x_2)(y-x_2)}{x_2(N-y-z_2+x_2)(z_1-x_1)(y-x_1)}$
20	(XII)	$\frac{1}{\sqrt{d}} \sum_{i=1}^d \left(\sqrt{\frac{(Nx-yz)^2 N}{z(N-z)y(N-y)}} \right)_i$

25

30 Zkušený pracovník v oboru snadno rozpozná funkci
 (VII) skóre jako výsledek součinitele momentové korelace,

odrážející stupeň sdílené odchylky mezi dvěma dichotomickými proměnnými explicitně nevyjádřenými v uvedeném vzorci.

5 Zkušený pracovník v oboru snadno nahlédne, že funkce (VII) skóre se týká odhadu poměru možných pravděpodobností s využitím strmosti regresní křivky reprezentující stupeň sdílené odchylky, která existuje mezi dvěma dichotomickými proměnnými.

10 Osoba v oboru znalá snadno rozpozná funkci (IX) skóre jako statistiku na bázi chi-kvadrátu, modifikovanou pro různé omezující faktory. Například výraz $N/2$ v čitateli druhého kvocientu logaritmicky váženého výsledku je konzervativním nastavením přirozeného přiblížení k binomickému rozdělení, což je využitelnou modifikací pro práci s relativně malými hodnotami x , y , z , a N . Zkušený pracovník v oboru, snadno
15 nahlédne, že mohou být použita jiná měřítka asociace a/nebo funkce skóre pro stejné účely na místo těch, která byla popsána ve vzorcích (I) a (II), z nichž nejvhodnější z hlediska předkládaného vynálezu obsahují různé kombinace
20 jedné, dvou, tří nebo čtyř z proměnných x , y , z , a N .

25 Zkušený pracovník v oboru rozpozná funkci (X) skóre jako způsob pro odhad hodnoty spodní hranice 95% konfidenčního intervalu měřítka (III) s použitím logaritmické transformace pro lepší porovnatelnost rozdělení poměru s normálním rozdělením, a aproximací Taylorových řad prvního řádu pro odhad odchylky logaritmu stejného uvedeného poměru.

Osoba v oboru znalá nahlédne, že funkce (XI) skóre je způsobem pro porovnání poměrů pravděpodobností, umožňující identifikovat chemické determinanty, které s největší

pravděpodobností budou selektivní pro jeden terč než pro jiný.

5 Zkušený pracovník v oboru rozpozná funkci (XII) skóre jako způsob pro kombinování množství testů asociace, umožňující identifikovat chemické determinanty, které s největší pravděpodobností mají účinky na dvě nebo více daných vlastností současně.

10 Osoba v oboru znalá rovněž snadno nahlédne, že funkce skóre může být modifikována tak, aby zahrnovala přídatné proměnné vztažené k materiálu molekul, biologickým, chemickým a/nebo fyzikálně-chemickým vlastnostem. Například by takovéto modifikace mohly zahrnovat, ale nijak nejsou omezeny pouze na, nastavení pro mocnost sloučeniny, selektivitu, toxicitu, biologickou dostupnost, stabilitu (metabolickou nebo 15 chemickou), realizovatelnost syntézy (sestavení), čistotu, komerční dosažitelnost, dostupnost vhodných reagentů pro syntézu, cenu, molekulární hmotnost, molekulární refraktivitu, molekulární objem, logP (vypočítaný nebo stanovený), počet skupin akceptujících vodíkové vazby, počet 20 skupin poskytujících vodíkovou vazbu, náboje (parciální a formální), protonační konstanty, počet molekul obsahujících přídatné chemické klíče nebo deskriptory, počet otočných vazeb indexy flexibility, indexy tvaru molekul, podobnosti seskupení a/nebo objemy překrytí.

25 Tak například může být funkce (VIII) skóre dále modifikována například pro započtení molekulární hmotnosti každého chemického determinantu při uvažování (MW) následovně:

$$MW \cdot e^{\{(x/z) - (z-x)/(N-z)\}}$$

Podobně může být modifikována funkce (IX) skóre pro
začlenění proměnných MW a [S], které reprezentují molekulární
hmotnost příslušného chemického determinantu (MW) respektive
počet toho, kolikrát se uvedený chemický determinant objevuje
v podskupině aktivních sloučenin x ([S]), následovně:

$$(II) \quad \text{Score} = \text{Log} \left(MW \cdot \frac{x}{[S]} \cdot \frac{(|Nx - yz| - N/2)^2 N}{z(N-z)y(N-y)} \right)$$

aby se tak umožnila identifikace největších možných,
ojedinělých, biologicky aktivních chemických determinantů
během analýzy.

Výsledky algoritmu z kroku 650 poskytují hodnotu
skóre uvažovaného fragmentu. Kroky 610 až 650 algoritmu mohou
být opakovány pro každý ze zvolených fragmentů v datech. Když
již byly vypočítány hodnoty pro všechny zvolené fragmenty,
výsledky poskytují hodnotu skóre odpovídající potenciální
účinnosti každého z fragmentů, které byly analyzovány.
Uvedené hodnoty skóre mohou být seřazeny podle velikosti;
přičemž ty fragmenty, které s největší pravděpodobností
přispívají k uvažovanému chemickému a/nebo biologickému
výsledku, jsou sdruženy například s vysokými hodnotami skóre.
To umožňuje v kroku 440 identifikaci jednoho nebo více
lokálních extrémů hodnot funkce skóre, jejichž odpovídající
chemické determinanty reprezentují úplná nebo částečná
chemická řešení pro požadovaný chemický nebo biologický

výsledek. Nalezení největších hodnot skóre, které může být dosaženo v jakékoliv dané datové sadě, je ekvivalentní identifikování chemickým determinantům obsaženým v podskupině molekul majících požadované vlastnosti, přičemž tyto chemické determinanty mají nejnížší pravděpodobnost výskytu shodou okolností ve stejných podskupinách. Když požadovanou vlastností je daná biologická aktivita fragmenty nebo chemické determinanty s největším skóre reprezentují biologicky aktivní farmakofoř.

Opět s přihlédnutím ke znázornění na obr. 2 budou nyní diskutována výhodná provedení kroku 250 analýzy knihovny 120 fragmentů.

Jeden způsob analýzy knihovny 120 fragmentů je znázorněn na obr. 7. Proces začíná výběrem fragmentu v kroku 710 na základě hodnot skóre, stanovených v předcházejícím kole. Potom jsou v kroku 720 vyjmuty z předcházejícího souboru sloučeniny, které obsahují vybraný fragment. Protože v kroku 710 byl vybrán fragment s vysokým příspěvkem k požadované aktivitě, sloučeniny, které jsou vyjmuty v kroku 720, mohou být považovány za aktivní sloučeniny. Potom je vybrán v kroku 730 soubor neaktivních sloučenin, buď z předcházejícího souboru nebo z databází nebo z jakéhokoliv jiného zdroje. Potom jsou aktivní a neaktivní sloučeniny uvedeny dohromady v kroku 740 pro vytvoření nového souboru sloučenin. Tento nový soubor sloučenin je potom vybrán v kroku 220 jako soubor sloučenin pro následující generace opakování, aby se tak pokračovalo s dalším kolem.

Nyní bude ve spojení s obr. 8 popsáno výhodné provedení pro realizaci kroku 730. Toto provedení využívá

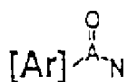
generické podstruktury pro zvolení nového souboru sloučenin pro následující kolo (opakování).

Proces podle obr. 8 začíná v kroku 810 analyzováním struktury fragmentu, který byl vybrán v kroku 710. Při využití obecného aspektu vynálezu může být fragment, vybraný v kroku 710, zvolen prostřednictvím vyhodnocení hodnoty skóre, která byla vypočítána v předchozím kole. Navíc může být výběr fragmentu prováděn závisle na dalších faktorech, které ovlivňují vhodnost fragmentu jako počátečního bodu pro zobecnění. Tato vhodnost by mohla být funkcí počtu atomů nebo vazeb, způsobu jakým jsou atomy vázány, trojrozměrné struktury příslušného fragmentu, a tak dále.

Poté, co v kroku 810 již byla analyzována struktura vybraného fragmentu, je v kroku 820 lokalizována zobecnitelná položka ve struktuře fragmentu. Tato položka je potom v kroku 830 nahrazena zobecněným výrazem pro vytvoření generické podstruktury (například pro nalezení bio-isosterů). Příkladem může být.



Fragment



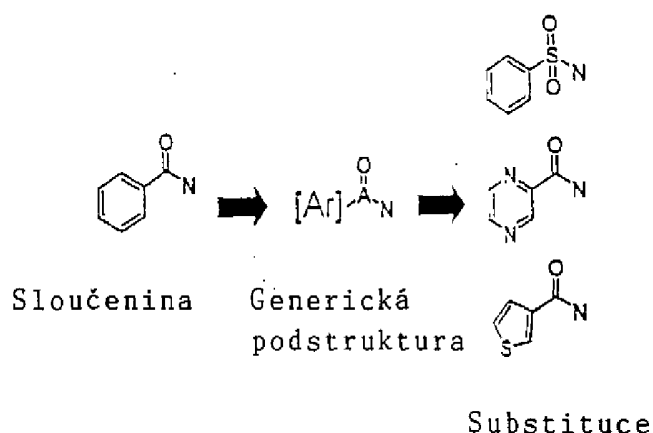
Generická
podstruktura

kde v daném vybraném fragmentu byly lokalizovány dvě zobecnitelné položky a nahrazeny zobecněnými výrazy [Ar] a A,

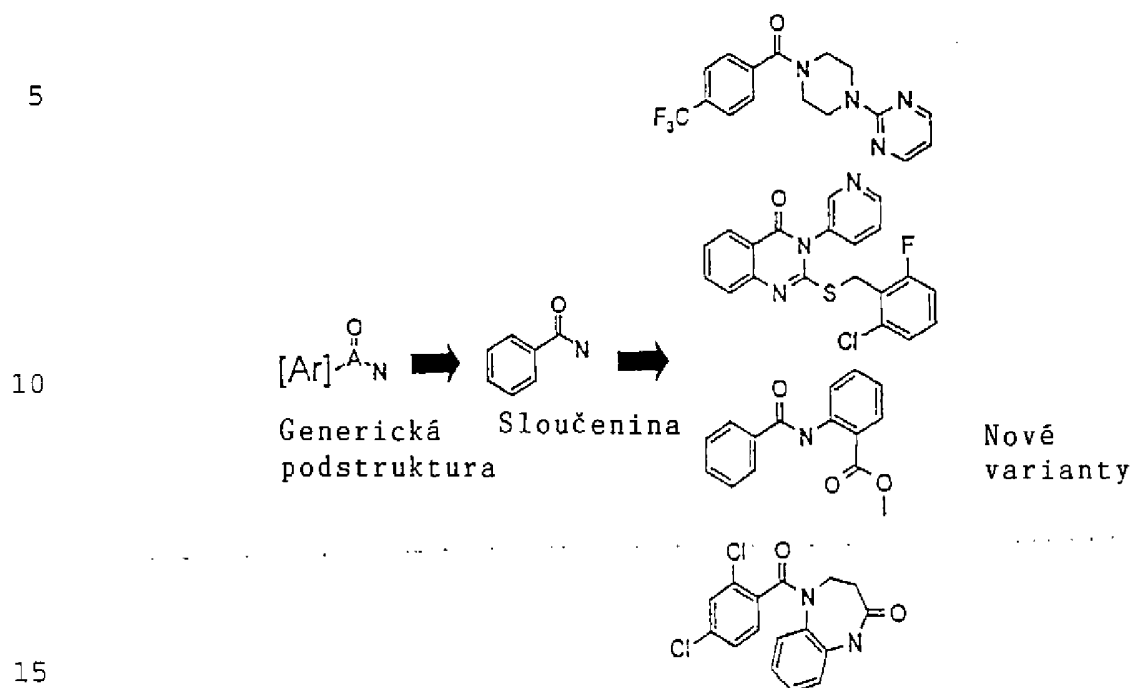
příčemž [Ar] reprezentuje aromatické jádro a A reprezentuje C, nebo S.

Generická podstruktura vytvořená v kroku 830 je potom v kroku 840 použita pro provedení virtuálního screeningu pro nalezení nových sloučenin odpovídajících generické podstruktury. Termín "virtuální screening" označuje jakýkoliv proces třídění, který je prováděn pouze s daty, čímž se vylučuje nutnost sestavovat (synteticky vytvářet) sloučeniny. Nové sloučeniny, které jsou odhaleny prostřednictvím virtuálního screeningu, jsou potom použity pro zkonstruování nového souboru sloučenin v kroku 850, který potom může být použit v následujícím kole opakování.

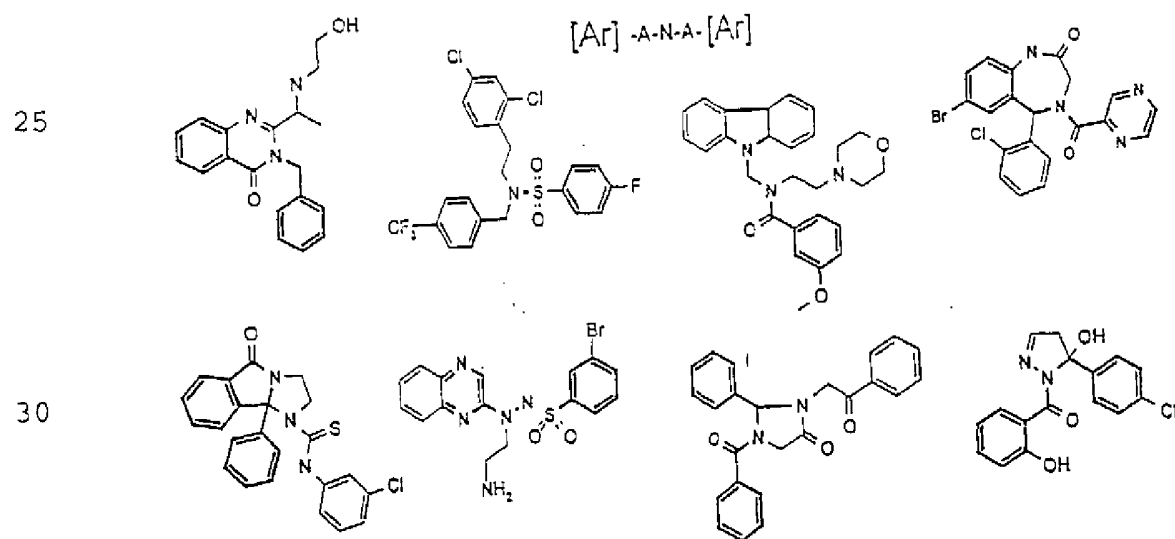
Jak může být patrné z obr. 9, proces virtuálního screeningu může být rozdělen na modifikace fragmentů vnitřní oblasti a vnější oblasti, vytvořených prostřednictvím použití generických podstruktur. Modifikace vnitřní oblasti, prováděné v kroku 910, zahrnují substituce, vřazování, rušení a inverze atomů fragmentu. Počínajíc od výše zmiňovaného přesného fragmentu a zobecnění tohoto fragmentu na generickou podstrukturu jsou získány tři různé substituce v následujícím příkladu:



Modifikace vnější oblasti, prováděné v kroku 920, sestávají ze změn v substituentech fragmentu. Tyto mohou být náhodné, cílené, a tak dále:

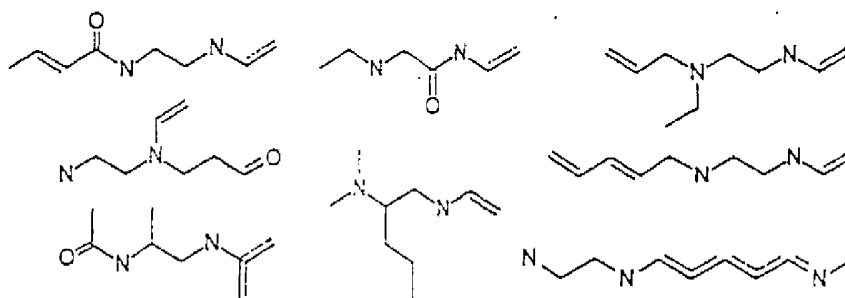


Soubory cílených sloučenin, jsou sbírky molekul, které jsou založeny na modifikacích jedné nebo více generických podstruktur:

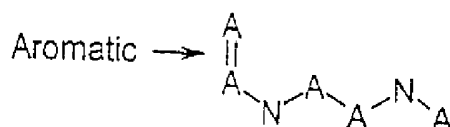


Zatímco na obr. 9 je znázorněno, že kroky provádění modifikací vnitřních oblastí a vnějších oblastí jsou realizovány sériově (po sobě), mělo by být osobám s běžnými znalostmi v oboru zcela zřejmé, že do rozsahu vynálezu spadá rovněž realizace pouze jednoho z těchto různých typů modifikací, nebo realizace obou modifikací v jiném pořadí nebo dokonce paralelně (současně). Mělo by být také zřejmé, že výsledkem virtuálního screeningu je rozmanitá sbírka sloučenin, které mají vysokou pravděpodobnost, že jsou aktivní, protože jsou bohaté na podstruktury sdružené s aktivitou.

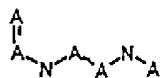
Zatímco v kroku 710 je vybrán fragment, který tvoří základ pro aplikaci funkcí 145 zobecnění pro získání generické podstruktury, je dalším výhodným provedením podle vynálezu volba většího počtu fragmentů s vysokým skóre pro vytváření generických podstruktur. Například bylo zjištěno, že následující fragmenty mají velké příspěvky k požadované aktivitě a mohou tedy být vybrány v kroku 710:



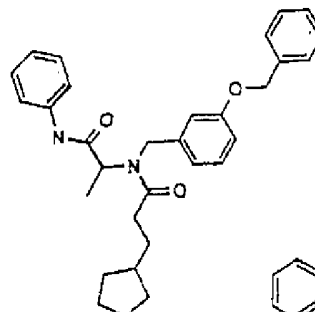
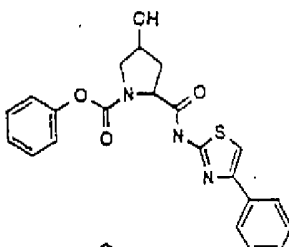
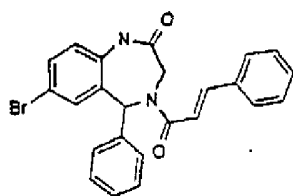
Tyto vybrané fragmenty jsou potom redukovány na generické podstruktury s vysokým skóre, jako:



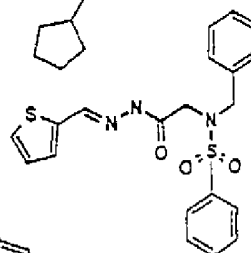
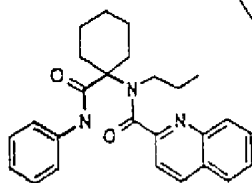
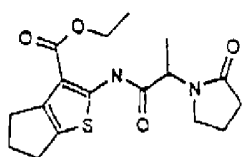
Tyto generické podstruktury jsou potom použity pro virtuální screening komerčních databází



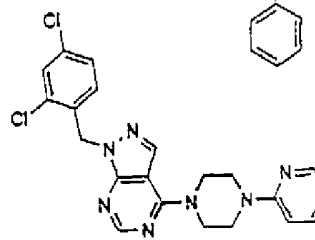
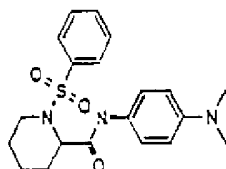
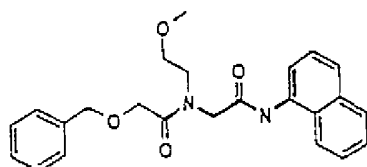
5



10



15



nebo korporativních sbírek sloučenin.

20

Zatímco byl popsán opakovací proces (proces re-iterace) jako výhodný z výpočetních důvodů, protože je výhodné začínat s malými fragmenty a zvětšovat velikost fragmentů mezi po sobě jdoucími koly opakování, a zatímco bylo dále ukázáno, že výkon objevování může být ještě zvýšen prostřednictvím užití generických či obecných aspektů v tomto opakovacím procesu, existuje ještě další přístup podle předkládaného vynálezu pro další zlepšení procesu diskrétní podstrukturní analýzy podle vynálezu. Tento další přístup je založen na technice vázání za zvýšené teploty (slučování - annealing) a bude v následujícím popisu popsán ve spojení s odkazy na obr. 10.

30

Ve výhodném provedení podle obr. 10 začíná krok 250 analýzy knihovny fragmentů, která byla vytvořena v předcházejícím kole, s kroky 1010 a 1020 výběru prvního respektive druhého fragmentu. Oba fragmenty jsou zvoleny na základě vypočítaných hodnot skóre a mohou být považovány za fragmenty s vysokým příspěvkem.

V následujícím kroku 1030 se aplikuje funkce 155 slučování pro spojení prvního a druhého fragmentu. Spojení fragmentů znamená definování molekulární struktury nebo podstruktury obsahující oba fragmenty. Pro tento účel může být použito množství různých funkcí 155 slučování. Tyto funkce slučování se liší v konkrétní realizaci toho, jak jsou určité slučovací parametry vyhodnocovány a použity. Slučovacími parametry jsou, například, (předem stanovená) vzdálenost prvního a druhého fragmentu, třírozměrná orientace prvního a druhého fragmentu, počet atomů, které jsou vloženy mezi fragmenty, počet vazeb, které jsou použity pro spojení fragmentů dohromady, typ vazeb a atomů, a tak dále.

Navíc je slučovací proces dále výhodně kombinován s obecným (generickým) aspektem vynálezu, jak bylo popisováno výše. Pokud například v krocích 1010 a 1020 jsou vybrány fragmenty F1 a F2, o nichž je známo, že mají vysoké hodnoty skóre, mohla by slučovací funkce, která je vybrána v kroku 1030 a spuštěna v kroku 1040 vytvoření nového souboru sloučenin, použít pro spojení fragmentů generický výraz:



Kde obecný výraz $[G]$ je synonymum pro molekulární podstruktury daných vlastností a slučovacích parametrů a závisí na použité slučovací funkci.

Jakmile již byly fragmenty zkombinovány prostřednictvím přesných nebo generických výrazů, je v kroku 1040 vytvořen nový soubor sloučenin, které obsahují oba fragmenty. Příklad molekuly nového souboru sloučenin je znázorněn na obr. 11, který je dvourozměrnou mapou relativních příspěvků, ilustrující relativní příspěví ve vztahu k lokálním souřadnicím. Jak může být patrné z obr. 11, jsou zde dvě lokální maxima, ilustrující přibližné hodnoty skóre 1,2 a 1,7 pro fragmenty F1 a F2.

Slučovací proces (annealing) je výhodný ze dvou důvodů. První výhodou je to, že spojením dvou fragmentů majících vysoký příspěvek k požadované aktivitě mohou být získány větší molekuly, které využívají skutečnosti, že obsahují více než jeden fragment s vysokým skóre. Výsledné struktury mají tudíž velké naděje k tomu, aby měly dokonce větší hodnotu skóre, než je největší hodnota skóre těchto dvou fragmentů.

Například ve struktuře podle obr. 11 výsledná sloučenina obsahuje fragmenty mající hodnoty skóre 1,2 a 1,7, ale může dosáhnout celkové hodnoty skóre pro celou strukturu například 2,1. Slučovací technika tudíž umožňuje objevování či nalézání sloučenin s dokonce vyšší aktivitou.

Druhou výhodou je to, že slučovací technika umožňuje vyloučit zablokování výpočetního procesu. Jak může být patrné z obr. 11, hodnoty relativních příspěvků indikují dvě lokální maxima. Při provádění opakovacího procesu ilustrovaného na



obr. 3, se začátkem s malými fragmenty a se zvyšováním velikosti fragmentů v každém dalším následujícím kole opakování může dojít k zablokování, když je vybraný fragment v jednom z prostředních kroků umístěn na lokálním maximu.

5 Například když je na konci druhého kola vybrán fragment $N-C=O$ a tento fragment je umístěn na lokálním maximu, následující kolo opakování nebude úspěšné. Jak bylo popsáno výše, fragmenty následujícího kola jsou výhodně konstruovány z vybraného fragmentu předcházejícího kola
10 prostřednictvím přírůstkového zvětšování velikosti fragmentu. Tudíž, ať je jakýkoliv atom přidán k vybranému fragmentu, následující kolo posune fragment od lokálního maxima. To znamená, že v tomto případě výsledný fragment má nižší hodnotu skóre, než vybraný fragment z předcházejícího kola.

15 Pro vyloučení tohoto zablokování může být aplikována slučovací technika prostřednictvím výběru dvou dobrých fragmentů z předcházejícího kola, spojením těchto fragmentů, výpočtem hodnoty skóre a pokračováním v procesu. To může být
20 prováděno periodicky v každém kole, nebo kdykoliv je detekováno zablokování.

25 Zatímco předkládaný vynález byl popsán s použitím množství výhodných provedení, mělo by být osobám s běžnými znalostmi v oboru zcela zřejmé, že tento vynález není nijak omezen pouze na tato provedení. Například může být změněno pořadí kroků ve vývojovém diagramu způsobu podle vynálezu, nebo kroky, které jsou ilustrovány jako prováděné postupně po sobě, by mohly být dokonce prováděny paralelně (současně), viz. Například kroky 1010 a 1020 procesu ilustrovaného na
30 obr. 10.

Navíc by osobám s běžnými znalostmi a zkušenostmi v oboru mělo být také zcela zřejmé, že ne všechny z ilustrovaných kroků způsobu jsou požadovány v každé realizaci. Například v procesu výpočtu skóre podle obr. 6 není požadováno, aby byly vypočítávány parametry, které nejsou použity funkcí skóre. Navíc by parametry mohly být také vypočítávány paralelně s využitím operačního systému pro souběžné či paralelní zpracovávání více úloh.

Nyní budou prostřednictvím příkladů popsána další provedení předkládaného vynálezu.

Například knihovna fragmentů, vytvořená v kroku 230, může teoreticky obsahovat všechny možné fragmenty a jejich kombinace. To může být dosaženo v praxi tehdy, když je knihovna generována prostřednictvím počítače. Pokud je ale knihovna generována manuálně, bude pravděpodobně obsahovat pouze výběr všech možných fragmentů. Způsob tudíž může být opakován s použitím kombinací fragmentů, zejména kombinací fragmentů, pro které již byly vysoké hodnoty skóre získány v předchozí analýze.

Tudíž následně po počáteční analýze fragmentů ty fragmenty, které s největší pravděpodobností přispívají k příslušnému chemickému a/nebo biologickému výsledku, mohou být kombinovány a může být aplikován algoritmus podle výše uváděného popisu pro odhad příspěvku uvedeného kombinovaného fragmentu ve vztahu k požadovanému chemickému a/nebo biologickému výsledku. Získaná hodnota skóre může být porovnána s hodnotami skóre jednotlivých fragmentů pro ověření, zda tato kombinace má za následek zlepšení příspěvku k příslušnému chemickému a/nebo biologickému výsledku.

V dalším provedení předkládaného vynálezu je možné vybírat z fragmentů, majících největší příspěvek k příslušnému chemickému a/nebo biologickému výsledku, společnou strukturní část pro identifikování, zda příspěvek
5 uvedeně společné části je stejný nebo vyšší než výchozí fragmenty.

Fragmenty s nejvyššími hodnotami skóre reprezentují chemický determinant nebo molekulární otisk mající největší váhu pro přispění k danému chemickému nebo biologickému
10 výsledku.

Po identifikování uvedeného otisku je potom možné vytvořit knihovnu sloučenin obsahujících uvedené chemické determinanty. Sloučeniny mohou být získány prostřednictvím programu syntézy (sestavování) kolem příslušného strukturního
15 znaku. Alternativně mohou být sloučeniny, obsahující chemický determinant, identifikovány z komerčních katalogů a nakupovány od příslušného zdroje. Sloučeniny nemusí být nezbytně připravovány pro farmaceutické účely a mohou být dostupné z nejrozličnějších zdrojů.

Jakmile již byla sestaveny požadovaná knihovna, může být podrobena screeningu proti určeným požadovaným terčům. Výsledky screeningu mohou identifikovat sloučeniny, které jsou dostatečně aktivní pro další vývoj, nebo mohou poskytnou
25 vodící struktury pro program syntézy. DSA metoda podle předkládaného vynálezu umožňuje vytvoření rozmanitých a přitom ještě vysoce cílených knihoven ve vztahu k určitému biologickému nebo farmakologickému terči. Je tedy podstatně zvýšena pravděpodobnost úspěchu při screeningu pro aktivní
30 sloučeniny a/nebo použitelné vodící struktury.



V dalším provedení předkládaný vynález navrhuje způsob identifikace molekul, majících určité požadované vlastnosti, jako jsou biologicky aktivní molekuly, přičemž tento způsob zahrnuje:

5 • vážení příspěvku molekulárních fragmentů, uvnitř podskupiny molekul, pro daný chemický nebo biologický výsledek, jak bylo popisováno výše,

10 • identifikování jednoho nebo více fragmentů s nejvyšší váhou, a

 • sestavení souboru sloučenin, přičemž tyto sloučenin obsahují jeden nebo více z uvedených fragmentů, a případně

15 • testování uvedených sloučenin na požadované vlastnosti.

20 Mělo by být zcela zřejmé, že způsob může být stejně tak použit pro identifikování fragmentů, které vedou k nežádoucím vlastnostem, jako jsou například škodlivé biologické vedlejší účinky, a tudíž pro vyloučení sloučenin, majících takové fragmenty, z úvah o použití.

25 Způsob podle předkládaného vynálezu tudíž generuje strukturní hypotézy (fragmenty), u nichž se pravděpodobnost toho, že jsou vysvětlením daného biologického, biochemického, farmakologického nebo toxikologického výsledku, odhaduje výpočtem kvantitativní hodnoty skóre. Uvažování hodnoty skóre pro daný fragment umožňuje osobě pracující na vývoji léků provádět informovaná rozhodnutí o přístupech, které s
30 největší pravděpodobností povedou k dosažení požadovaného cíle, jako je identifikace silněji působících sloučenin,

objevení nové řady aktivních sloučenin, identifikace více selektivních nebo více biologicky dostupných sloučenin nebo eliminace toxických účinků.

5 Způsob podle předkládaného vynálezu se zaměřuje na fragmenty přítomné v podskupině příslušných sloučenin, čímž se eliminuje nutnost provádět nadměrná množství zdlouhavých výpočtů pro pravděpodobněji méně relevantní sektory chemického prostoru. To má za následek redukci počtu výpočetních kroků, které jsou potřebné pro vyřešení daného
10 biologického výsledku, při současném zachování základní úrovně molekulárního významu, který je požadovaná za účelem stanovení existence biologicky aktivních chemických determinantů.

15 Jak bylo diskutováno výše, proces podle předkládaného vynálezu zahrnuje hledání lokálních extrémů jedné nebo více funkcí, které mohou být snadno zvoleny tak, aby tyto extrémy odpovídaly pravděpodobnostem daným v běžných statistických tabulkách. To poskytuje elegantní způsob vyhodnocení potenciálního příspěvku daného fragmentu k chemickému nebo
20 biologickému výsledku. Není ale nezbytné založit analýzu na statistické teorii za účelem realizace předkládaného vynálezu.

25 DSA metoda podle předkládaného vynálezu může být využita v širokém aplikací objevování nových léků. Jak bylo popsáno výše, způsob podle vynálezu umožňuje identifikaci farmakoforů, které mají vysokou pravděpodobnost přispívání k dané biologické aktivitě, jako jsou například antagonisty 7-TM receptorů, inhibitory kinázy, inhibitory fosfatázy,
30 blokátory iontových kanálků a inhibitory proteázy a rovněž

aktivní součástí přirozeně se vyskytujících peptidergických ligandů.

5 Způsob rovněž umožňuje identifikaci endogenních modulátorů lékových cílů, což umožňuje identifikaci nových směrů farmakologické intervence a rovněž racionální začlenění nových farmakologických vlastností do molekul předtím nemajících tyto uvedené vlastnosti.

10 Způsob podle vynálezu může být rovněž použit pro identifikování chybně pozitivních a chybně negativních výsledků v datových souborech, například souborech, které jsou odvozeny z vysokovýkonného screeningu. DSA metoda je rovněž použitelná pro předvídaní selektivity sloučeniny například prostřednictvím identifikování potenciálně nežádoucích sekundárních účinků.

15 Způsob může být použit stejným způsobem pro předvídaní toxických účinků sloučeniny prostřednictvím identifikování jeho "toxikoforních" chemických determinantů, což ve spojení s výše uvedeným umožňuje konstruovat databáze chemických determinantů s velkým využitím pro selekci chemických řad. V tomto kontextu způsob podle vynálezu dále umožňuje racionální začlenění nových farmakologických vlastností do chemických sloučenin předtím nemajících takovéto aktivity. Nakonec v důsledku svojí schopnosti
20 identifikovat nejvhodnější úroveň molekulární rozmanitosti, která musí být testována během procesu screeningu, DSA metoda umožňuje efektivní provádění racionálních, do značné míry paralelně realizovaných, automatizovaných a vysoce výkonných procesů screeningu, což je znatelným zlepšením oproti
25 současným strategiím vysoce výkonného objevování.
30

Mělo by být zřejmé, že ve výše popisovaném způsobu podle předkládaného vynálezu se alespoň jeden krok provádí prostřednictvím počítačem řízeného systému. Tak například hodnoty x , y , z a N , získané z databází, mohou být zadávány do a zpracovávány prostřednictvím vhodně naprogramovaného počítače. Do rozsahu předkládaného vynálezu tudíž rovněž spadají takovéto počítačem řízené nebo počítačem realizované postupy.

Z výše uvedeného popisu je zřejmé, že předkládaný vynález poskytuje nový způsob pro rychlou identifikaci molekul majících určité požadované vlastnosti, jako jsou biologicky aktivní molekuly. Zejména se tento vynález týká způsobu vážení účinnosti molekulárních struktur za účelem identifikování biologicky aktivních součástí molekulárních struktur a využití těchto součástí při konstruování cílených souborů chemických sloučenin pro rychlejší a cenově efektivnější objevování léků.

Způsob je navržen pro zvýšení podílu biologicky aktivních sloučenin v daném souboru chemických entit, přičemž o těchto entitách ještě není známo, že by měly požadovanou biologickou aktivitu. Uvedený způsob zahrnuje aplikaci různých matematických technik pro kvantitativní stanovení vztahů strukturní aktivity (QSAR). Tento nový postup, který může být označen jako diskrétní podstrukturní analýza (DSA) poskytuje řešení například pro problém rozpoznání farmakologického obrazu, to jest problém identifikování chemických determinantů (CD), které jsou odpovědné s ohledem na danou sloučeninu za jakýkoliv daný chemický nebo biologický výsledek, kterým může být například biologická,

biochemická, farmakologická, chemická a/nebo toxikologická aktivita.

5 Způsob podle předkládaného vynálezu má široké využití a není omezen na oblast farmakologie. Pokud se týká biologicky aktivních sloučenin, způsob podle vynálezu může být například využit ve spojení s pesticidy a herbicidy, kde požadovanou biologickou aktivitou je pesticidní respektive herbicidní aktivita. Způsob může být rovněž použit v aplikacích reaktivního modelování, kde požadované vlastnosti
10 jsou spíše chemické než biologické atributy, například tedy při vytváření katalyzátorů.

Mělo by být zcela zřejmé, že technikou podle vynálezu je kombinování v podskupině nebo mezi různými podskupinami
15 těch fragmentů, které s největší pravděpodobností přispívají k příslušnému chemickému a/nebo biologickému výsledku, a aplikování algoritmu pro odhad příspěvku uvedeného kombinovaného fragmentu vzhledem k příslušnému chemickému a/nebo biologickému výsledku, přičemž získaná hodnota skóre může být porovnávána s hodnotami skóre jednotlivých fragmentů
20 pro ověření, zda kombinace má za následek zlepšení příspěvku k příslušnému chemickému a/nebo biologickému výsledku.

Navíc předkládaný vynález umožňuje výběr, z fragmentů majících největší příspěvek k příslušnému chemickému a/nebo
25 biologickému výsledku, společné strukturní části pro identifikování, zda příspěvek uvedené společné části je stejný nebo větší než příspěvek výchozích fragmentů.

Navíc je použito míry nebo měřítka asociace, které je výhodně zvoleno z odčítacích měřítek, poměrových měřítek nebo
30 smíšených měřítek. Měřítka asociace je výhodně začleněno ve

funkci skóre nebo rozvinuta na funkci skóre. Funkce skóre může být rozvinuta s použitím statistické metody zvolené z metody kritických poměrů, Fisherova testu přesnosti, Personovy metody na bázi chi-kvadrátu, Mantel Haenzelovy metody na bázi chi-kvadrátu, odhadu strmostí a podobně. Dalším výhodným provedením je, že funkce skóre je rozvinuta s použitím metody zvolené z výpočtu a porovnání přesných a přibližných konfidenčních intervalů, koeficientů korelace nebo jakékoliv funkce explicitně obsahující měřítko asociace zahrnující jakoukoliv kombinaci jedné, dvou, tří nebo čtyř z proměnných x , y , z a N .

Výhodně se podle předkládaného vynálezu provádí krok výběru molekul, obsahujících fragmenty s nejvyšším skóre, jako potenciálních ligandů a případně následně jejich testování jako modulátorů lékového terče. Způsob podle vynálezu může být výhodně použit pro identifikování chybně pozitivních a/nebo chybně negativních experimentálních výsledků. Dalšími výhodnými aplikacemi je provádění zkoumání podobnosti, analýzy rozmanitosti a/nebo analýzy shody.

V následujícím popisu jsou uvedeny ilustrace množství aplikací DSA procesu podle vynálezu. Příklady jsou výhodnými provedeními vynálezu a slouží pro ilustraci vynálezu, ale v žádném případě nemají jakoukoliv omezující funkci pro rozsah vynálezu.

Příklad č. 1 - Racionální identifikace nových a selektivních receptorových ligandů

Zkouška konkurenčních vazeb byly rozvinuta pro povrchový receptor buněk s využitím rekombinačního

20

$$Nx = yz$$

30

k binomickému rozdělení, což je využitelnou modifikací pro práci s relativně malými hodnotami x , y , z , a N . Proměnné MW a $[S]$, které reprezentují molekulární hmotnost příslušného chemického determinantu (MW) respektive počet, kolikrát se stejný uvedený determinant objevuje v podskupině aktivních sloučenin x ($[S]$), byly začleněny do funkce skóre za účelem umožnění identifikace největších možných, samostatných, biologicky aktivních chemických determinantů během analýzy. Zkušený pracovník v oboru snadno nahlédne, že mohou být použita jiná měřítko asociace a/nebo funkce skóre pro stejné účely na místo těch, která byla popsána ve vzorcích (I) a (II), z nichž nejvhodnější z hlediska předkládaného vynálezu obsahují různé kombinace jedné, dvou, tří nebo čtyř z proměnných x , y , z , a N .

$$(II) \quad \text{Score} = \text{Log} \left(MW \cdot \frac{x}{[S]} \cdot \frac{(|Nx - yz| - N/2)^2 N}{z(N-z)y(N-y)} \right)$$

Osoba v oboru znalá rovněž snadno nahlédne, že funkce skóre (II) by mohla být modifikována tak, aby zahrnovala přídatné proměnné vztažené k materiálu molekul, biologickým, chemickým a/nebo fyzikálně-chemickým vlastnostem. Například by takovéto modifikace mohly zahrnovat, ale nijak nejsou omezeny pouze na, nastavení pro mocnost sloučeniny, selektivitu, toxicitu, biologickou dostupnost, stabilitu (metabolickou nebo chemickou), realizovatelnost syntézy (sestavení), čistotu, komerční dosažitelnost, dostupnost vhodných reagentů pro syntézu, cenu, molekulární hmotnost, molekulární refraktivitu, molekulární objem, $\log P$ (vypočítaný

5

10

15

20

25

30

IP₃ s IC₅₀ o hodnotě 22 nM, což je hodnota konzistentní s afinitou sloučeniny pro receptor.

Jak je znázorněno na obr. 12, sloučenina A podstatně snížila receptorem zprostředkovanou tvorbu inositoltrisfosfátu při funkčním testu na bázi buněk (IC₅₀ = 22 nM), což je zjištění konzistentní jak s afinitou sloučeniny pro receptor tak i s použitím receptorových antagonistů ve výše popisovaných výpočtech. Nakonec byla sloučenina A určena jako vysoce selektivní pro příslušný receptor, neboť selhala při demonstrování značné inhibiční (blokační) aktivity při testování při koncentraci 10 μM ve více než 20 jiných testech vázání receptorových radioligandů.

Pátý krok sestával v použití výše popisované reprezentativní kostry pro přímou koncepční konstrukci a syntézu (sestavení) nových chemických sloučenin, ve smyslu sestavování látky, a z hlediska identifikování nových molekul s aktivitami vázání receptoru. Pro tento účel byl sestaven seznam chemických reaktantů a produktů reakcí, ve kterém byla výše popisovaná biologicky aktivní reprezentativní kostra, nebo její fragmenty, obsažena buď v chemických strukturách reaktantů nebo ve výsledných produktech reakcí. Bylo zvoleno více než 2000 kombinací reaktantů a byly sestaveny odpovídající reakční produkty pro testování. Testování těchto sloučenin v testu vázání receptorů vedlo k identifikaci nové třídy chemické sloučeniny ve smyslu sestavování látky, z které množství reprezentantů vykázalo IC₅₀ v rozsahu od 50 do 500 nM.

Příklad č. 2 - Racionální identifikace nových a selektivních inhibitorů kinázy

30

15

25

30

μm . Mezi těmito vykázalo šest molekul hodnotu IC_{50} v rozsahu 0,2 až 2 μm a jedna sloučenina, označená jako sloučenina B, vykázala hodnotu IC_{50} 164 nm (viz obr. 13).

Obr. 13 ilustruje účinek sloučeniny B na proteinovou fosforylaci závislou na kináze. Příslušná kináza byla inkubována radiově označeným ATP a peptidovým substrátem za přítomnosti zvyšujících se koncentrací sloučeniny B. Proteinová fosforylace byla měřena s použitím standardních radiometrických technik. Sloučenina B významně inhibovala fosforylaci proteinového substrátu, závislou na kináze, přičemž vykazovala hodnotu IC_{50} 164 nM.

Mezi 1280 náhodně vybranými sloučeninami, testovanými pro kontrolní účely, pouze tři vykázaly inhibiční aktivitu při testovacím screeningu, z nichž nejsilnější vykazovala hodnotu IC_{50} pouze 7,8 μM . Jako takový byl soubor sloučenin, sestavený na základě reprezentativních otisků, 13,2 krát účinnější při dodání aktivních molekul, než byl soubor náhodně vybraných sloučenin ($p < 0,0001$). Navíc bylo shledáno, že sloučenina B reprezentuje nový, doposud nepopsaný inhibitor ATP-konkureční kinázy, vykazující více než 250 násobnou selektivitu pro příslušnou kinázu při testování v selektivních testech s použitím jak strukturně tak i funkčně alternativních kináz.

Patý krok sestával v použití jedné nebo více z výše popisovaných reprezentativních koster pro přímou koncepční konstrukci a syntézu nových chemických sloučenin, ve smyslu sestavování látky, a z hlediska identifikování nových molekul s aktivitami inhibitoru kinázy. Pro tento účel byl sestaven seznam chemických reaktantů a produktů reakcí, ve kterém byly výše popisované biologicky aktivní reprezentativní kostry,

5

10

15

30

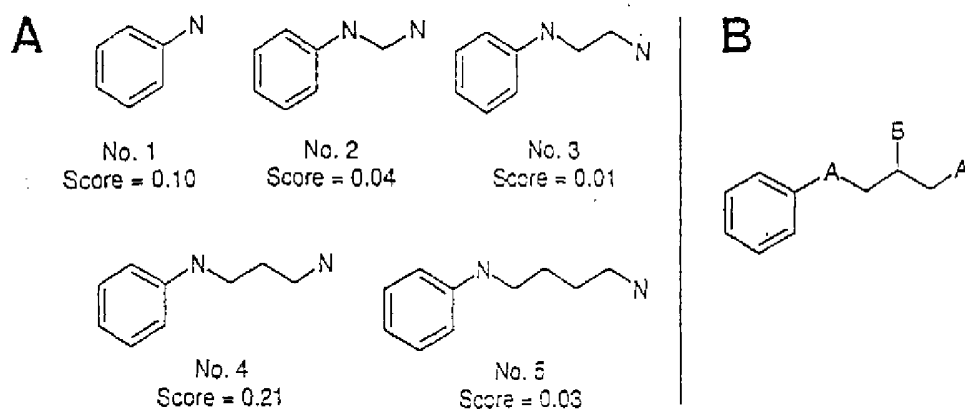
měřítka (I) asociace, ve kterém x reprezentovalo počet aktivních chemických struktur obsahujících příslušný chemický determinant, y reprezentovalo celkový počet chemických struktur obsahujících stejný uvedený chemický determinant, z reprezentovalo celkový počet aktivních chemických struktur v souboru N molekul (to jest $z = 36$), a N reprezentovalo celkový počet chemických struktur podrobených analýze (to jest $N = 3680$). Měřítka (I) asociace potom bylo rozvinuto do funkce (V) skóre, kterou zkušený pracovník v oboru snadno rozpozná jako výsledek součinitele momentové korelace, odrážející stupeň sdílené odchylky mezi dvěma dichotomickými proměnnými explicitně nevyjádřenými v uvedeném vzorci (V).

$$(V) \quad \text{Score} = \frac{Nx - yz}{\sqrt{z(N-z)y(N-y)}}$$

V tomto případě ve funkci skóre nebyly použity žádné další proměnné jiné než x , y , z a N , ačkoliv osoba v oboru znala snadno nahlédne, že funkce (V) skóre by mohla být také modifikována pro zahrnutí dalších proměnných se vztahem k materiálu molekul, biologickým, chemickým a/nebo fyzikálně-chemickým vlastnostem, jak bylo zmiňováno v příkladu 1, ale bez omezení na citace uvedené v příkladu 1. Osoba s dobrými znalostmi v oboru rovněž snadno nahlédne, že pro stejný účel by mohla být použita jiná měřítka asociace a/nebo funkce skóre namísto těch, které byly popsány ve vzorcích (I) a (V), zejména proto, že funkce (V) skóre není neměnná při různých změnách pro studijní konstrukce a/nebo rozdělení y , $(N-y)$, z a $(N-z)$. Nejvhodnější z těchto

alternativních postupů z hlediska předkládaného vynálezu obsahují různé kombinace dvou, tří nebo čtyř z proměnných x, y, z a N.

Následující panely ukazují příklady chemických determinantů použitých pro analýzu a vybraných pro další zkoumání. Celkem 3680 struktur katalogizovaných pro aktivitu inhibice kanálku bylo tetsováno na přítomnost biologicky aktivních podstruktur s použitím sady chemických determinantů zahrnujících pět determinantů ilustrovaných na panelu A. Mezi těmito pěti strukturami vykazoval nejvyšší hodnotu skóre determinant č. 4, což indikuje, že měl nejvyšší pravděpodobnost, že je základem pro aktivitu inhibice kanálku. Výpočty tedy byly opakovány pro struktury obsahující determinant č. 4 a byla identifikována chemická struktura, znázorněná na panelu B, jako jedna z největších, statisticky významných determinantů obsažených v souboru 36 inhibitorů, přičemž tato struktura byla zvolena pro další zkoumání. Symboly: A reprezentuje C, N, O nebo S; B reprezentuje H nebo



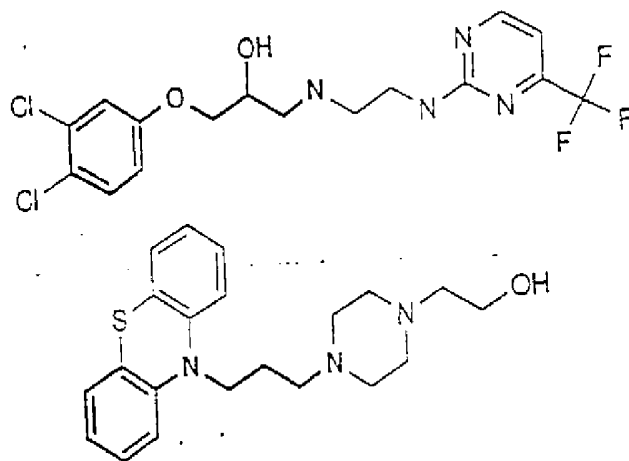
Analýza 3680 katalogizovaných struktur byla porvedena vytvořením skóre pro série chemických determinantů se vzorcem (V) a podržením struktur poskytujících největší, nenulové

15

20

3

vykazovalo hodnoty IC_{50} v podmikromolárním rozsahu a jedna sloučenina, označená jako sloučenina C, vykazovala hodnotu IC_{50} 400 nM. Dva příklady těchto sloučenin inhibujících (iontový) kanálek jsou znázorněny níže, přičemž oba obsahují právě farmakologicky aktivní "otisk" znázorněný na panelu B:



Tyto dvě kanálek inhibující sloučeniny byly vybrány pro testování s použitím způsobu podle předkládaného vynálezu. Obě molekuly významně inhibovaly příslušný kanálek. Jak je znázorněno silněji vyznačenými podstrukturami, chemické struktury těchto dvou sloučenin obsahují farmakologicky aktivní chemický determinant identifikovaný s použitím způsobu podle předkládaného vynálezu a znázorněný na panelu B výše.

Mezi 1280 náhodně vybranými sloučeninami, testovanými pro kontrolní účely, celkem 33 molekul vykazalo inhibitorní aktivitu o hodnotě alespoň 40% při testu screeningu. Jako takový soubor sloučenin, sestavený na základě reprezentativního otisku ilustrovaného na panelu B, byl 1,8 krát účinnější při poskytnutí aktivních molekul, než byl soubor náhodně vybraných sloučenin ($p < 0,005$). Soubor sloučenin, sestavený na základě reprezentativního otisku

ilustrovaného na panelu B, byl také 4,9 krát účinnější při poskytnutí aktivních molekul, než bylo prvních 3680 sloučenin firemní sbírky sloučenin ($p < 0,0001$).

Pátý krok sestával z využití reprezentativní kostry znázorněné na panelu B pro přímou koncepční konstrukci a syntézu nových chemických sloučenin ve smyslu sestavování látky a z hlediska identifikování nových molekul s vlastnostmi inhibice pro (příslušný) kanálek. Pro tento účel jeden ze 120 farmakologicky aktivních inhibitorů, popisovaných výše, byl vybrán pro další zkoumání a byl chemicky modifikován s použitím předtím sestavených kladných a záporných výsledků screeningu jako zdroje informací o aktivitě struktury. To vedlo k syntéze a následné identifikaci nové, doposud nepopsané třídy blokátoru iontového kanálu, ve smyslu sestavování látky, přičemž množství reprezentantů z této třídy vykazovalo hodnoty IC_{50} v rozsahu 100 až 500 nM. Testování selektivity indikovalo, že sloučenina byla selektivní pro příslušný kanálek oproti 30 dalším lékovým terčům a navíc inhibovala úmrtí buněk v modelu nervové apoptózy indukované odebráním růstového faktoru.

Příklad č. 4 - Racionální identifikace nových a selektivních inhibitorů proteázy

Byl rozvinut enzymatický test pro proteázu, o které se předpokládá, že má klíčovou roli při ischemickém poškození a poranění. Příslušná proteáza byla členem rodiny blízce příbuzných enzymů, která sama byla jediným terčem zájmu pro terapeutickou intervenci. Byl sestaven soubor sloučenin pro testování v testu, které byly testovány, a byly identifikovány nové inhibitory enzymu způsobem podle

předkládaného vynálezu. První krok sestával z vytvoření potřebných strukturních dat pro identifikování chemických determinantů inhibitorů enzymu. To bylo provedeno testováním souboru 1680 sloučenin při koncentraci 3 μ M v testu screeningu a popsáním (katalogizací) každé struktury na inhibiční aktivitu. S použitím hranice 40% inhibice jako prahové hodnoty pro třídění bylo 17 struktur identifikováno jako aktivních a zbývajících 1663 molekul bylo kvalifikováno jako neaktivních.

Druhý krok sestával z identifikování biologicky aktivních chemických determinantů obsažených ve strukturách 17 inhibitorů. Pro tento účel bylo 1680 katalogizovaných struktur analyzováno prostřednictvím výběru níže ilustrovaného smíšeného měřítka (VI) asociace, ve kterém x reprezentovalo počet aktivních chemických struktur obsahujících příslušný chemický determinant, y reprezentovalo celkový počet chemických struktur obsahujících stejný uvedený chemický determinant, z reprezentovalo celkový počet aktivních chemických struktur v souboru N molekul (to jest $z = 17$), a N reprezentovalo celkový počet chemických struktur podrobených analýze (to jest $N = 1680$). V tomto případě bylo měřítko (VI) asociace přímo použito jako funkce skóre pro identifikování biologicky aktivních chemických determinantů obsažených v příslušných 17 inhibitorech.

$$(VI) = \frac{x}{z} - \frac{y}{N}$$

V tomto případě ve funkci skóre nebyly použity žádné další proměnné jiné než x , y , z a N , ačkoliv osoba v oboru znalá snadno nahlédne, že vzorec (VI) by mohl být rovněž modifikován pro zahrnutí dalších proměnných se vztahem k materiálu molekul, biologickým, chemickým a/nebo fyzikálně-chemickým vlastnostem, jak bylo zmiňováno v příkladu 1, ale bez omezení na citace uvedené v příkladu 1.

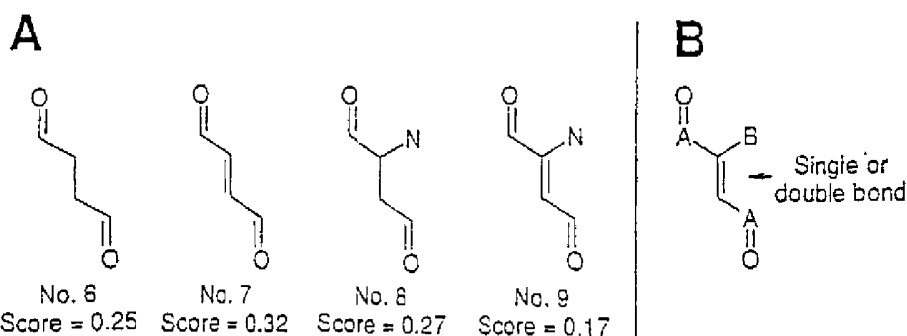
Osoba s dobrými znalostmi v oboru rovněž snadno nahlédne, že pro stejný účel by mohla být použita jiná měřítko asociace a/nebo funkce skóre namísto těch, které byly popsány ve vzorci (VI), zejména proto, že toto měřítko asociace pouze umožňuje relativní odhad pravděpodobnosti, se kterou je daný chemický determinant základem pro biologickou aktivitu. Nejvhodnější z těchto alternativních postupů z hlediska předkládaného vynálezu obsahují různé kombinace dvou, tří nebo čtyř z proměnných x , y , z a N .

Analýza 1680 katalogizovaných struktur byla provedena prostřednictvím zjištění skóre u sérií chemických determinantů s užitím vzorce (VI) a podržením struktur poskytujících největší kladné hodnoty. Příklady některých chemických determinantů, použitých v tomto procesu, jsou znázorněny níže na panelu A společně s jejich vypočítanými hodnotami skóre. Mezi nimi vykazaly determinanty č. 7 a č. 8 nejvyšší skóre a byly akceptovány jako reprezentativní pro jednu nebo více aktivních částí obsažených v podstatné části ze 17 inhibitorů. Výpočty s použitím vzorce (VI) potom byly opakovány, aby se zjistilo, zda by mohly být identifikovány dokonce ještě větší chemické determinanty, což ale nebyl ten případ při využití dostupného souboru 17 struktur a determinanty č. 7 a č. 8 byly sloučeny dohromady pro

vytvoření reprezentativní kostry nebo farmakologicky aktivního "otisku", jak je znázorněno níže v panelu B. Tato reprezentativní kostra byla následně použita pro výběr a syntézu sloučenin.

5

10



15

20

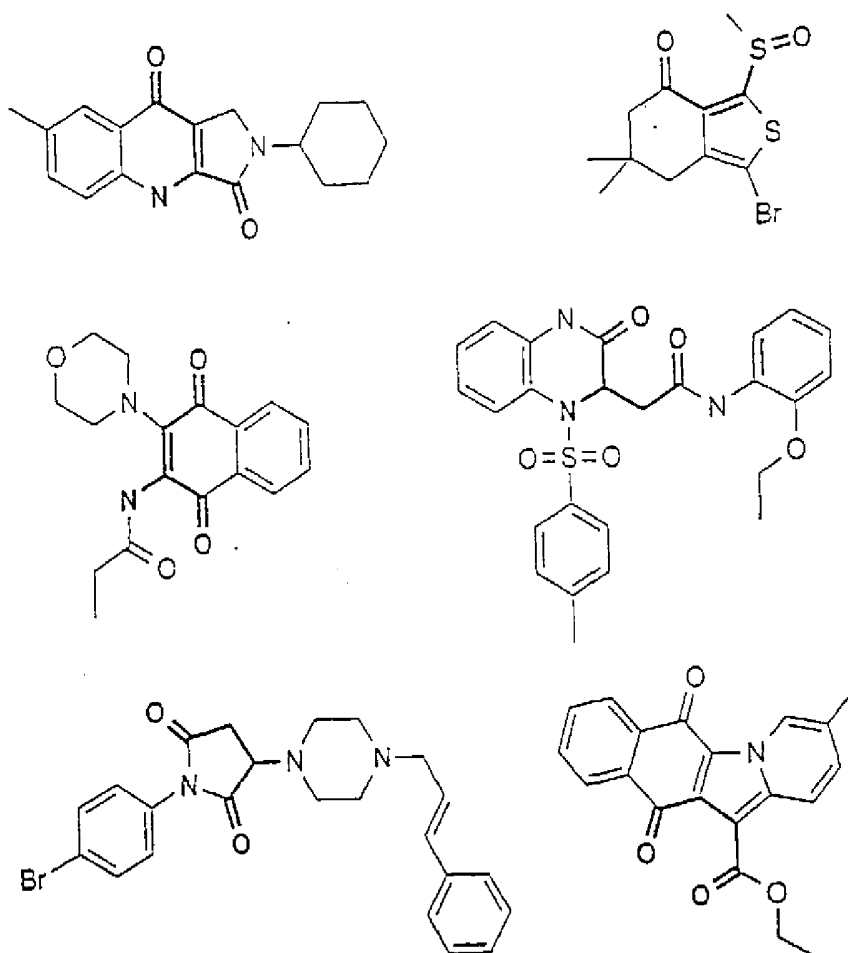
25

30

Na panelech jsou znázorněny příklady chemických determinantů, použitých pro analýzu a vybraných pro další zkoumání. Celkem 1680 struktur, katalogizovaných pro aktivitu inhibice proteázy, bylo testováno na přítomnost biologicky aktivních podstruktur s použitím souboru chemických determinantů zahrnujících čtyři determinanty ilustrované na panelu A. Mezi těmito čtyřmi strukturami determinanty č. 7 a č. 8 vykazovaly nejvyšší hodnoty skóre, což indikovalo, že mají největší pravděpodobnost toho, že tvoří základ aktivity inhibice proteázy. Determinant, sestávající z jednoduchého benzenového kruhu, měl při porovnání skóre 0,02. Protože při opakovaných výpočtech s determinanty č. 7 a č. 8 nebyly identifikovány žádné struktury s větším skóre, byly tyto dvě struktury sloučeny do chemického motivu znázorněného na panelu B, který byl následně použit jako farmakologicky aktivní "otisk" pro virtuální screening a výběr sloučenin. Symboly: A reprezentuje C nebo S; B reprezentuje H, C, N, O nebo atom jakéhokoliv halogenu.

Třetí krok zahrnoval použití reprezentativní kostry, popsané v panelu B, jako šablony pro virtuální screening a výběr sloučenin. Pro tento účel byly provedeny podstrukturní průzkumy v databázi přes 150000 komerčně dostupných sloučenin s použitím jak vypočítaného otisku tak i jeho fragmentů. Na základě těchto průzkumů bylo získáno celkem 589 sloučenin.

Čtvrtý a poslední krok procesu zahrnoval testování získaných sloučenin v enzymatickém testu. Z 589 sloučenin, vybraných na základě reprezentativní kostry, 52 molekul vykazovalo inhibiční aktivity s hodnotou alespoň 40% při testování v testu s koncentrací 3 μ M. Mezi nimi 12 sloučenin vykazovalo hodnoty IC_{50} v podmikromolárním rozsahu a jedna sloučenina, označená jako sloučenina D, vykazovala hodnotu IC_{50} 65 nM. Šest příkladů těchto molekul inhibujících proteázu je znázorněno níže, přičemž všechny tyto molekuly obsahují alespoň jeden výskyt farmakologicky aktivního "otisku" znázorněného na panelu B:

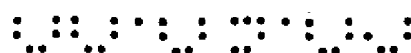


Těchto šest sloučenin inhibujících proteázu bylo vybráno pro testování s použitím způsobu podle předkládaného vynálezu. Každá molekula významně inhibovala příslušný protein, přičemž vykazovaly hodnoty IC_{50} v rozsahu 0,15 až 15 μ m. Jak je znázorněno prostřednictvím silněji vyznačených podstruktur, struktury každé z těchto šesti sloučenin obsahují farmakologicky aktivní chemický determinant identifikovaný s využitím vynálezu a znázorněný na panelu B výše. Některé z těchto sloučenin vlastně obsahují více než jednu variantu tohoto otisku, jako například tetracyklická struktura znázorněná výše v pravém spodním rohu.

Jako takový byl soubor sloučenin, sestavený na základě reprezentativního otisku znázorněného na panelu B, 8,7 krát účinnější při poskytnutí aktivních molekul, než byl původně testovaný soubor 1680 sloučenin ($p < 0,0001$). Navíc bylo zjištěno, že 52 racionálně identifikovaných sloučenin je selektivních pro příslušnou proteázu, neboť většina ($> 90 \%$) selhalo při prokazování inhibiční aktivity při testování při koncentraci 5 μ M na příbuzné proteáze příslušející ke stejné rodině enzymů a rovněž při testování za stejných podmínek pro 12 dalších lékových terčů.

Příklad č. 5 - Racionální identifikace nových a selektivních inhibitorů fosfatázy

Byl rozvinut enzymatický test pro fosfatázu, o které se předpokládá, že má důležitou úlohu při zcitlivění (senzibilaci) a regulaci receptoru. Byl sestaven soubor sloučenin, který byl testován a byly identifikovány nové inhibitory způsobem podle předkládaného vynálezu. První krok sestával z vytvoření potřebných strukturních dat pro



identifikování chemických determinantů inhibitorů příslušného enzymu. To bylo provedeno testováním prvních 12160 sloučenin z firemní sbírky předkladatele přihlášky při koncentraci 3 μM v testu screeningu a popisem (katalogizací) každé struktury v seznamu pro jejich inhibiční aktivitu. S použitím hranice 50% inhibice jako prahové hodnoty pro třídění sloučenin bylo celkem 15 chemických struktur identifikováno jako aktivních a zbývajících 12145 molekul bylo kvalifikováno jako neaktivních.

Druhý krok sestával z identifikování biologicky aktivních chemických determinantů obsažených ve strukturách 15 inhibitorů. Pro tento účel bylo 12160 katalogizovaných struktur analyzováno prostřednictvím výběru níže ilustrovaného smíšeného měřítka (VII) asociace, ve kterém x reprezentovalo počet aktivních chemických struktur obsahujících příslušný chemický determinant, y reprezentovalo celkový počet chemických struktur obsahujících stejný uvedený chemický determinant, z reprezentovalo celkový počet aktivních chemických struktur v souboru N molekul (to jest $z = 15$), a N reprezentovalo celkový počet chemických struktur podrobených analýze (to jest $N = 12145$).

$$(VII) \quad \dots \dots \dots (x/z) - (z-x)/(N-z)$$

Měřítka (VII) asociace potom bylo rozvinuto do funkce (VIII) skóre, přičemž osoba v oboru znalá snadno identifikuje, že tato funkce skóre se týká odhadu poměru možných pravděpodobností s využitím strmosti regresní křivky reprezentující stupeň sdílené odchylky, která existuje mezi

dvěma dichotomickými proměnnými, která byla dále modifikována pro započtení molekulární hmotnosti každého chemického determinantu při uvažování (MW).

5

$$(VIII) \quad \text{Score} = MW \cdot e^{[(x/z) - (z-x)/(N-z)]}$$

10

15

20

V tomto případě ve funkci skóre nebyly použity žádné další proměnné jiné než x , y , z , N nebo MW , ačkoliv je zjevné, že osoba v oboru znalá snadno nahlédne, že vzorec (VIII) by mohl být také modifikován pro zahrnutí dalších proměnných se vztahem k materiálu molekul, biologickým, chemickým a/nebo fyzikálně-chemickým vlastnostem, jak bylo zmiňováno v příkladu 1, ale bez omezení na citace uvedené v příkladu 1. Osoba s dobrými znalostmi v oboru rovněž snadno nahlédne, že pro stejný účel by mohla být použita jiná měřítka asociace a/nebo funkce skóre namísto těch, které byly popsány ve vzorci (VIII), zejména protože porovnání strmostí nemusí v některých případech umožňovat postačující diskriminaci mezi dvěma blízkce příbuznými chemickými determinanty. Nejvhodnější z těchto funkcí skóre z hlediska předkládaného vynálezu obsahují různé kombinace dvou, tří nebo čtyř z proměnných x , y , z a N .

25

30

Analýza 12160 katalogizovaných struktur byla provedena skórováním sérií chemických determinantů s využitím vzorce (VIII) a podržením struktur poskytujících největší kladné hodnoty. To vedlo k identifikaci tří rozdílných chemických determinantů, s molekulární hmotností v rozsahu od 120 do 220, které měly pravděpodobnost menší než 1 z 10, že budou obsaženy v podskupině aktivních chemických struktur

10

20

30

Obr. 14 ilustruje účinek sloučeniny E na defosforylaci proteinu závislou na fosfatázi. Příslušná fosfatáze byla inkubována s fosforylovaným peptidovým substrátem za přítomnosti zvyšujících se koncentrací sloučeniny E. Byla zkoumána defosforylace substrátu prostřednictvím měření uvolňování volného fosfátu do reakčního média s malachitovou zelení. Sloučenina E významně inhibovala defosforylaci závislou na fosfatázi, přičemž vykazovala IC_{50} o hodnotě 87 nM.

Mezi 1280 náhodně vybranými sloučeninami, testovanými pro kontrolní účely, pouze dvě vykázaly inhibiční aktivitu při testovacím screeningu, z nichž nejsilnější vykazovala hodnotu IC_{50} pouze 1,8 μ M. Jako takový byl soubor sloučenin, sestavený na základě reprezentativních otisků, 17,5 krát účinnější při poskytování aktivních molekul, než byl soubor náhodně vybraných sloučenin ($p < 0,0005$), a 22,3 krát účinnější než prvních 12160 sloučenin firemní sbírky sloučenin ($p < 0,00001$).

Navíc nakonec bylo shledáno, že sloučenina E reprezentuje novou, doposud nepopsanou třídu inhibitoru fosfatázy, vykazující více než 20 násobnou selektivitu pro příslušný terč při testování v selektivních testech s použitím jak strukturně tak i funkčně alternativních fosfatáz.

Příklad č. 6 - Zvýšení potence chemických sérií

Předkládaný vynález může být rovněž použit pro zvýšení potence chemických sérií. Pro příkladné znázornění této možnosti byla testována sbírka 1251 sloučenin s

koncentrací 3 μM v testu pro proteázu, což poskytlo 25 sloučenin vykazujících inhibiční aktivitu alespoň 40 %. Byla provedena analýza struktur tak, jak bylo popsáno v příkladu č. 1, což vedlo na identifikaci množství chemických determinantů, z nichž jeden měl pravděpodobnost menší než 1 z 10000, že se bude vyskytovat mezi 7 z 25 inhibitorů proteázy na základě pouze náhodného výběru ($p < 0,0001$). Naneštěstí sedm sloučenin, obsahujících tento determinant, vykazovalo pouze nízké inhibiční aktivity (střední hodnota $\text{IC}_{50} = 3,4 \mu\text{M} \pm 1,34 \mu\text{M}$, $n = 7$), což je činilo neatraktivní pro další chemické zkoumání. Následně byl příslušný determinant akceptován jako reprezentativní pro biologicky aktivní část příslušných inhibitorů a byl přímo použit jako reprezentativní kostra nebo farmakologicky aktivní "otisk" pro další výběr sloučenin.

Pro tento účel byl proveden screening v databázi přes 100000 komerčně dostupných molekul na příslušný determinant, přičemž bylo vybráno 142 molekul pro další testování. Mezi těmito 142 sloučeninami 11 vykazovalo inhibiční aktivity v podmikromolárním rozsahu, přičemž měly střední hodnotu $\text{IC}_{50} 0,48 \mu\text{M} \pm 0,09 \mu\text{M}$ ($n = 11$, střední hodnota IC_{50} podstatně menší než předchozí hodnota při $p < 0,05$). Jako takový tedy způsob podle předkládaného vynálezu umožňuje podstatně zvýšit farmakologickou potenci chemických sérií.

Příklad č. 7 - Zvýšení selektivity chemických sérií

Předkládaný vynález může být rovněž použit pro zvýšení selektivity chemických sérií. Pro příkladně znázornění této možnosti byla testována sbírka 3360 sloučenin s koncentrací 3 μM v testu pro kinázu, označeném test pro

15

30



reprezentuje celkový počet chemických struktur obsahujících uvedený chemický determinant, z_1 reprezentuje celkový počet chemických struktur aktivních na kináze č. 1 v souboru N molekul (to jest $z_1 = 22$), z_2 reprezentuje celkový počet chemických struktur aktivních na kináze č. 2 v souboru N molekul (to jest $z_2 = 92$), a N reprezentuje celkový počet chemických struktur podrobených analýze (to jest $N = 3360$).

$$(IX) \quad \text{Score} = \frac{x_1(N-y-z_1+x_1)(z_2-x_2)(y-x_2)}{x_2(N-y-z_2+x_2)(z_1-x_1)(y-x_1)}$$

Osoba v oboru znalá snadno rozpozná funkci (IX) skóre jako způsob pro porovnání relativních pravděpodobností, umožňující identifikovat chemické determinanty, které s největší pravděpodobností budou selektivní pro jednu kinázu oproti druhé. V této souvislosti je rovněž osobě v oboru znalé zřejmé, že vzorec (IX) by mohl být modifikován pro zahrnutí dalších proměnných se vztahem k materiálu molekul, biologickém, chemickém a/nebo fyzikálně-chemickém vlastnostem, jako bylo zmiňováno v příkladu č. 1, ale bez omezení na citace uváděné v příkladu č. 1. Zkušený pracovník v oboru rovněž snadno nahlédne, že mohou být použita jiná měřítko asociace a/nebo funkce skóre pro stejné účely na místo těch, která byla popsána ve vzorcích (III) a (IX). Například by tak mohlo být ve funkci (II) skóre použito měřítko (I) asociace a výsledné hodnoty skóre pro aktivitu na kinázu č. 2 by mohl být odečteny od hodnot získaných pro aktivitu na kinázu č. 1, nebo obráceně by hodnoty získané pro aktivitu na kinázu č. 1 mohl být děleny hodnotami získanými

5

20

30

Byl rozvinut funkční test pro ligandem uzavíraný iontový kanálek, o kterém se předpokládá, že má úlohu při

10

30

chemický determinant; y_1 reprezentuje celkový počet struktur obsahujících chemický determinant v seznamu struktur katalogizovaných pro účinky blokování sodíkových kanálků, y_2 reprezentuje celkový počet struktur obsahujících chemický determinant v seznamu struktur katalogizovaných pro inhibiční účinky vázání proteinu s nukleotidem purinu, z_1 reprezentuje celkový počet struktur inhibujících sodíkové kanálky v souboru N_1 molekul (to jest $z_1 = 79$), z_2 reprezentuje celkový počet chemických struktur aktivních při vázání proteinů s nukleotidem purinu v souboru N_2 molekul (to jest $z_2 = 2367$), a N_1 a N_2 reprezentují celkový počet chemických struktur podrobených analýze v příslušných seznamech katalogizovaných struktur.

$$(X) \quad \text{Score} = \frac{1}{\sqrt{2}} \left(\sqrt{\frac{(N_1 x_1 - y_1 z_1)^2 N_1}{z_1 (N_1 - z_1) y_1 (N_1 - y_1)}} + \sqrt{\frac{(N_2 x_2 - y_2 z_2)^2 N_2}{z_2 (N_2 - z_2) y_2 (N_2 - y_2)}} \right)$$

Zkušený pracovník v oboru snadno nahlédne, že funkce (X) skóre je způsobem pro zkombinování dvou různých testů asociace, který umožňuje identifikovat chemické determinanty, které s největší pravděpodobností mají účinky současně jak na sodíkové kanálky tak i na proteiny vázající nukleotid purinu. V tomto kontextu je osobě v oboru znalé rovněž zcela zřejmé, že vzorec (X) by mohl být modifikován pro zahrnutí dalších proměnných se vztahem k materiálu molekul, biologickému, chemickému a/nebo fyzikálně-chemickému vlastnostem, jak bylo zmiňováno v příkladu č. 1, ale bez omezení na citace uváděné v příkladu č. 1. Je rovněž jasně patrné, že pro stejný účel mohou být použita jiná měřítko asociace a/nebo funkce skóre

namísto těch, které byly popsány ve vzorcích (I) a (X), zejména prožoe funkce (X) skóre nebere do úvahy směr rozdílů, existujících mezi podíly dvou datových souborů, přičemž po celou dobu vyžaduje, aby tyto podíly byly srovnatelné a navíc aby hodnota N_1 byla srovnatelná s N_2 a aby obě tyto hodnoty byly větší než 20. Například je možné požadovat vážení výsledků pro datové soubory, ve kterých jsou velikosti vzorků podstatně odlišné, prostřednictvím použití funkce skóre, založené na váženém průměru rozdílu mezi podíly (viz příklad 21 níže). Alternativně je možné požadovat začlenění třetí, nebo čtvrté, nebo páté farmakologické vlastnosti do výpočtu, přičemž v tomto případě je zřejmé, že vzorec (X) může být rozšířen na jeho obecnější podobu (XI), ve které d reprezentuje počet seznamů sloučenin procházejících analýzou a kde výsledné hodnoty skóre mohou být přímo vztahovány k tabulkám standardního normálního rozdělení, aby se stanovila pravděpodobnost nalezení jednoho nebo více chemických determinantů, které jsou základem všech uvažovaných farmakologických vlastností. Je rovněž možných množství další přístupů, z nichž nejvhodnější z hlediska předkládaného vynálezu využívají funkce skóre, zahrnující různé kombinace dvou, tří nebo čtyř z proměnných x, y, z a N.

(XI)

$$\text{Score} = \frac{1}{\sqrt{d}} \sum_{i=1}^d \left(\sqrt{\frac{(N_x - yz)^2 N}{z(N-z)y(N-y)}} \right)$$

Analýza dvou seznamů katalogizovaných struktur byla provedena prostřednictvím skórování sérií chemických determinantů s užitím vzorce (X) a podržením struktur

poskytujících největší hodnoty větší než 2. To vedlo na identifikaci chemického determinantu majícího pravděpodobnost menší než 1 z 20 výskytu v obou podskupinách biologicky aktivních struktur pouze na základě náhodného výběru ($p < 0,05$). Tento chemický determinat, označený "determinant č. 12" byl tudíž akceptován jako reprezentativní pro jednu nebo více biologicky aktivních částí, jak pro inhibitory sodíkového kanálku tak i pro inhibitory vázání proteinu s nukleotidem purinu, a byl přímo použit jako reprezentativní kostra nebo farmakologicky aktivní "otisk" pro následný výběr sloučenin.

Třetí krok zahrnoval použití výše popisované reprezentativní kostry jako šablony pro virtuální screening. Pro tento účel byly provedeny podstrukturní průzkumy v databázi přes 250000 komerčně dostupných sloučenin s použitím determinantu č. 12 i jeho fragmentů. Na základě těchto průzkumů bylo získáno celkem 800 sloučenin pro testování a stejný soubor 1280 náhodně vybraných sloučenin, jako bylo popsáno v příkladu č. 1, byl použit pro kontrolní účely.

Čtvrtý a poslední krok procesu zahrnoval testování získaných sloučenin v testu iontového kanálku. Z 800 molekul, vybraných na základě determinantu č. 12, vykazovalo dvacet tři sloučenin inhibiční aktivitu alespoň 40 % při testování s koncentrací 3 μM . Mezi nimi tři sloučeniny vykazovaly hodnoty IC_{50} v podmikromolárním rozsahu a jedna sloučenina, označená jako sloučenina F, vykazovala hodnotu IC_{50} $145 \text{ nM} \pm 56 \text{ nM}$ ($n = 4$). Mezi 1280 náhodně vybranými sloučeninami, testovanými pro kontrolní účely, pouze jedna molekula vykazovala významnou inhibiční aktivitu v nízkém mikromolárním rozsahu a její chemická struktura vlastně

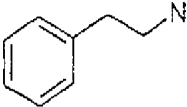
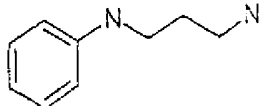
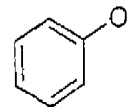


obsahovala podstatnou část determinantu č. 12. Překvapivě, když stejný soubor 800 sloučenin byl testován na kinázu, o které se rovněž předpokládá, že má úlohu při imunitní reakci, osm sloučenin vykazalo inhibitorní aktivity alespoň 40 % při testování s koncentrací 5 μM , sloučenina F vykazala hodnotu IC_{50} 1,2 μM a další sloučenina, označená jako sloučenina G, vykazala IC_{50} 137 nm $\pm$ 48 nM ($n = 4$). Bylo rovněž zjištěno, že sloučeniny F, G a množství blízce příbuzných molekul, rovněž obsahujících determinant č. 12 v jejich strukturách, rovněž inhibovaly (blokovaly) sodíkové kanálky, přičemž obvykle vykazovaly 50 až 100 % inhibicí (blokací) při 1 μM . V souhrnu tedy tyto výsledky demonstrují, že způsob podle předkládaného vynálezu umožňuje vybrat a/nebo zkonstruovat sloučeniny s množstvím farmakologických vlastností, což může být zajímavé při vývoji léků pro použití při léčení chorobných stavů ovlivňovaných více faktory, jako je například bez jakéhokoli omezení zánět. Prostřednictvím analogie může být rovněž zcela jasně zřejmé, že způsob podle předkládaného vynálezu může být použit pro začlenění nových farmakologických vlastností do chemických sérií, které předtím takovéto vlastnosti neměly.

Příklad č. 9 - Sestavování seznamů biologicky aktivních chemických determinantů

Ve výhodném provedení předkládaného vynálezu může být způsob podle vynálezu rovněž použit pro sestavování seznamů biologicky aktivních chemických determinantů, které dále mohou být použity jako referenční databáze pro použití při provádění racionální konstrukce léků, jako například v programech, provádějících počítačově řízená rozhodování, pro použití v lékařské chemii. Pro znázornění příkladem byla

prozkoumána vědecká literatura a bylo sestaveno 25 seznamů farmakologicky aktivních molekul, přičemž každý seznam zahrnoval chemické struktury sloučenin vykazujících danou farmakologickou vlastnost, jako například vázání sigma receptoru, agonismus receptoru dopaminu D_2 a antagonismus receptoru estrogenu. Každý seznam byl následně analyzován podle předkládaného vynálezu prostřednictvím výběru měřítka (III) asociace, jak bylo popsáno v příkladu č. 2, a jeho rozvinutím do funkce (IV) skóre, která byla použita pro skórování různých chemických determinantů obsažených v jednom nebo více seznamech procházejících analýzou. Tyto výpočty vedly k identifikaci velkého počtu farmakologicky aktivních chemických determinantů, z nichž tři jsou znázorněny v části výsledné matice v následující tabulce:

Determinant	Sigma Ligand	D_2 Aconista	Antagonista estrogenu
 No. 13	1.85	8.12	0.05
 No. 14	2.40	0.00	0.00
 No. 15	0.91	2.93	28.17

Tato tabulka poskytuje referenční seznam farmakologicky aktivních chemických determinantů. Bylo sestaveno dvacet pět seznamů struktur obsahujících molekuly, o kterých bylo popsáno, že mají jednu z dvaceti pěti různých farmakologických vlastností, a tyto seznamy struktur byly

analyzovány způsobem podle předkládaného vynálezu s použitím měřítka (III) asociace a funkce (IV) skóre. Uvedených dvacet pět vlastností zahrnovalo například schopnost vázat se na sigma receptory (sigma ligand), agonismus pro receptor dopaminu D_2 (D_2 agonist) a antagonismus pro receptor estrogenu (estrogen antagonist). Malá část výsledné matice s 26 sloupci je znázorněna v tabulce výše. Hodnoty větší než 1 indikují, že daný chemický determinant má pravděpodobnost menší než 1 z 20, že se bude vyskytovat při náhodném výběru v souboru molekul sdílejících stejnou farmakologickou vlastnost, což indikuje, že determinant s největší pravděpodobností bude molekulárním základem pro stejnou uvedenou vlastnost. Tabulky, jako je znázorněná tabulka, tvoří archivy či zdroje biologicky aktivních determinantů nebo "otisků", které mohou být použity jako referenční seznamy pro provádění informovaných rozhodnutí při objevování a vývoji léků.

Interpretace výsledné tabulky se provádí následovně. Sloučeniny, jejichž chemické struktury obsahují determinant č. 13 s větší pravděpodobností budou vykazovat vlastnosti agonistu receptoru dopaminu D_2 spíše než vlastnosti buď vázání sigma receptoru nebo antagonistu receptoru estrogenu, protože $8,12 > 1,85 > 0,05$. Obráceně tedy determinant č. 13 je výhodný determinant pro konstruování souborů potenciálních agonistů receptoru dopaminu D_2 , protože $8,12 > 2,93 > 0,00$. Stejným způsobem potom sloučeniny, jejichž chemické struktury obsahují determinant č. 14, budou s větší pravděpodobností ligandy sigma receptoru než buď agonisty receptoru dopaminu nebo antagonisty receptoru estrogenu, protože $2,4 > 0,00 = 0,00$. Opět je potom determinant č. 14 výhodným determinantem pro sestavování souborů ligandů sigma receptorů, protože $2,40$

> 1,85 > 0,91. Nakonec sloučeniny, jejichž chemické struktury obsahují determinant č. 15, budou s největší pravděpodobností vykazovat vlastnosti inhibice (blokace) receptoru estrogenu, protože 28,17 > 2,93 > 0,91, přičemž alternativně je determinant č. 15 výhodným otiskem pro sestavování souborů potenciálních antagonistů receptoru estrogenu, protože 28,17 > 0,05 > 0,00.

Osobě v oboru znalé je zjevné, že pro konstruování takovýchto tabulek by mohla být použita jiná měřítka asociace a/nebo funkce skóre namísto těch, které jsou popsány ve vzorcích (III) a (IV). Rovněž je zcela zřejmé, že použitá funkce skóre by mohla zahrnovat další proměnné se vztahem k materiálu struktur, biologickým, chemickým a/nebo fyzikálně-chemickým vlastnostem, jak bylo zmiňováno v příkladu č. 1, ale bez omezení na citace uváděné v příkladu č. 1. Dále je rovněž zcela zjevné, že funkce skóre nebo proces skórování by rovněž mohly být modifikovány pro zahrnutí kroků vážení nebo normalizace za účelem snazšího porovnávání jednotlivých hodnot skóre vzájemně mezi sebou, což je jistě případ výše uvedené tabulky, kde při její konstrukci byly použity podobně veliké vzorky, ale nemusí to být případ jiných datových souborů. Nakonec je rovněž zcela zřejmé, že stejný postup by mohl být použit pro sestavování referenčních seznamů struktur skórovaných pro jiné uvažované vlastnosti v procesu objevování, jako je například bez jakéhokoli omezení obecné terapeutické použití, toxicita, absorpce, distribuce, metabolismus, a/nebo vyměšování.

Příklad č. 10 - Předvídání sekundárních farmakologických působení molekuly

Předkládaný vynález může být dále použit pro předvídání sekundárních působení molekuly. Pro ilustraci této možnosti byla identifikována nová třída blokátorů iontového kanálku, jak je znázorněno v příkladu č. 3. Jak bylo v předcházejícím popsáno pro jiné inhibitory tohoto stejného kanálku, základní chemická struktura této nové chemické série inhibitorů obsahovala chemický determinant znázorněný v panelu B příkladu č. 3, zjevně ve formě determinantu č. 5 znázorněného v panelu A příkladu č. 3. Porovnáním determinantu č. 5 s determinanty obsaženými ve výše uváděné tabulce bylo odhaleno, že příslušné inhibitory měly velmi vysokou možnost vázání na sigma receptory, zejména proto, že chemická struktura determinantu č. 5 je shodná s chemickou strukturou determinantu č. 14. Následně byly blokátory kanálku, obsahující determinant č. 5, tetovány v testech na vázání sigma σ_1 a σ_2 receptory, a bylo zjištěno, že vykazují podmikromolární afinity pro obě místa. Jako takové tyto výsledky demonstrují, že hodnoty skóre, získané s použitím způsobu podle předkládaného vynálezu, umožňují předvídat sekundární působení chemických sérií, což je extrémně užitečné pro sériový postup v lékařské chemii.

Příklad č. 11 - Identifikace a předvídání toxického působení molekuly

Z předcházejících příkladů je zřejmé, že způsob podle předkládaného vynálezu může být rovněž použit pro identifikování toxikoforních chemických determinantů obsažených v pesticidech, herbicidech, insekticidech a podobně, a to jednoduše prostřednictvím analyzování seznamů struktur, které jsou popsány či katalogizovány pro



toxikologické vlastnosti namísto farmakologických. V této souvislosti může být předkládaný vynález přímo aplikován pro identifikaci mocnějších selektivnějších nebo širěji působících toxických chemických sérií pro použití, například, v programech zemědělské chemie pro ochranu úrody.

Alternativně může být předkládaný vynález použit pro sestavování referenčních seznamů nebo databází toxických chemických determinantů způsobem shodným s postupem popisovaným v příkladu č. 9. Takové seznamy mohou být potom použity pro odhadování pravděpodobnosti, že chemické řady budou vykazovat daný toxický účinek, což lze využít, například, při screeningu potravinářských příměsí a chemikálií ovlivňujících životní prostředí.

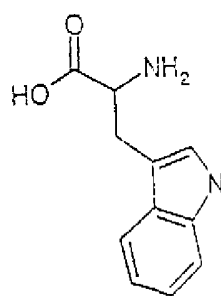
Pro ilustraci možnosti předvídání toxických účinků při nasazení pro farmaceutický výzkum bylo testováno 4480 sloučenin na zainteresovanou buněčnou fosfatázu pro léčení zánětu. Celkem 25 sloučenin vykazovalo inhibiční aktivity o hodnotě alespoň 40 % při testování v testu s koncentrací 10 μM , přičemž všechny vykazovaly hodnoty IC_{50} v nízkém mikromolárním rozsahu. Následně byla způsobem podle předkládaného vynálezu provedena analýza, což vedlo na identifikaci dvou molekulárně rozdílných chemických determinantů, které s největší pravděpodobností budou základem farmakologické aktivity, přičemž tyto determinanty byly označeny jako determinanty č. 16 a č. 18. Protože tyto dva determinanty byly přítomny v rovnocenných molekulách a oba byly vnímány jako schopné poskytovat chemické série, které by byly rovnocenně přístupné pro další chemické zkoumání, bylo rozhodnuto zvolit mezi těmito dvěma

determinanty na základě předvídaných toxických vedlejších účincích.

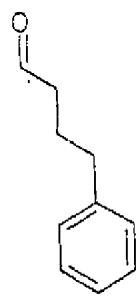
Pro tento účel byly struktury determinantů č. 16 a č. 17 porovnávány se strukturami obsaženými v toxikologické databázi a bylo zjištěno, že molekuly, obsahující determinant č. 16 v jejich strukturách, měly podstatně větší pravděpodobnost, že budou cytotoxické, než sloučeniny obsahující pouze determinant č. 17. To indikovalo, že inhibitory fosfatázy, nesoucí determinant č. 16, by byly méně zajímavé pro další postup v důsledku vlastní cytotoxicity farmakologického otisku. Tato hypotéza byla ověřena experimentálně vystavením kultivovaných buněk koncentracím 1 μ M obou tříd inhibitoru a změřením životaschopnosti buněk s použitím standardního testu MTT, přičemž bylo zjištěno, že všechny sloučeniny, obsahující determinant č. 16, indukovaly smrt buněk během 24 hodin po aplikaci, což nebyl případ u většiny sloučenin nesoucích determinant č. 17. Jako takové tyto výsledky zjevně demonstrují, že způsob podle předkládaného vynálezu umožňuje identifikovat a/nebo předvídat chemické série, které s největší pravděpodobností vykazují toxické vlastnosti v daném nasazení. V tomto kontextu je zcela zjevné, že shodné výpočty mohou být prováděny s použitím, například údajů o mutagenicitě (schopnost působit změny dědičných vlastností - mutace) (testy Ames), údajů o inhibování isozymu P4502, nebo údajů odvozených z jakýchkoliv jiných dalších relevantních testů toxicity.

Příklad č. 12 - Identifikace biologicky aktivních částí receptorových ligandů

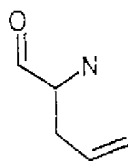
Receptor buněčného povrchu byl vybrán jako příslušný terč pro řízení jistých endokrinních chorob. Tento receptor byl popsán jako endogenně aktivovaný prostřednictvím hormonu nonapeptidu, vytvářenému podvěskem mozkovým. Byl sestaven seznam chemických struktur, popsanych jako ligandy stejného uvedeného receptoru, prostřednictvím prozkoumání vědecké literatury. Tento seznam byl následně analyzován způsobem podle předkládaného vynálezu s využitím měřítka (III) asociace, funkce (IV) skóre, přičemž seznam chemických determinantů zahrnoval fragmenty dvaceti běžných aminokyselin (glycin, alanin, valin, leucin, isoleucin prolin, serin, theorin, tyrosin, fenylalanin, tryptofan, lysin, arginin, histidin, aspartát, glutamát, asparagin, cysteín a methionin), doplněné fragmenty struktury hlavního peptidového řetězce $(\text{NH}-\text{CH}-\text{CO})_n$. Příklady těchto determinantů jsou ilustrovány níže:



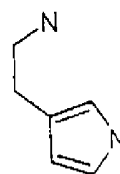
Tryptophan



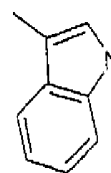
No. 18



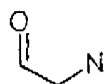
No. 19



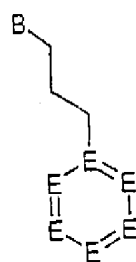
No. 20



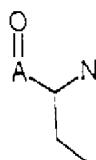
No. 21



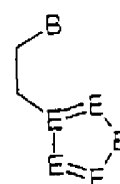
No. 22



No. 23



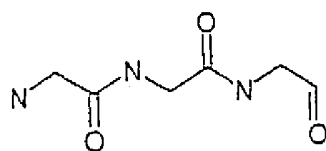
No. 24



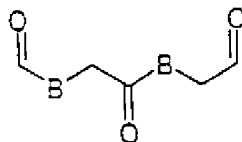
No. 25



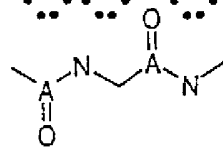
No. 26



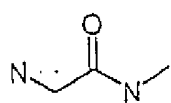
Kostra peptidu



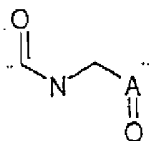
No. 27



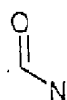
No. 28



No. 29



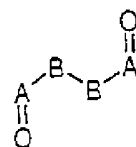
No. 30



No. 31



No. 32

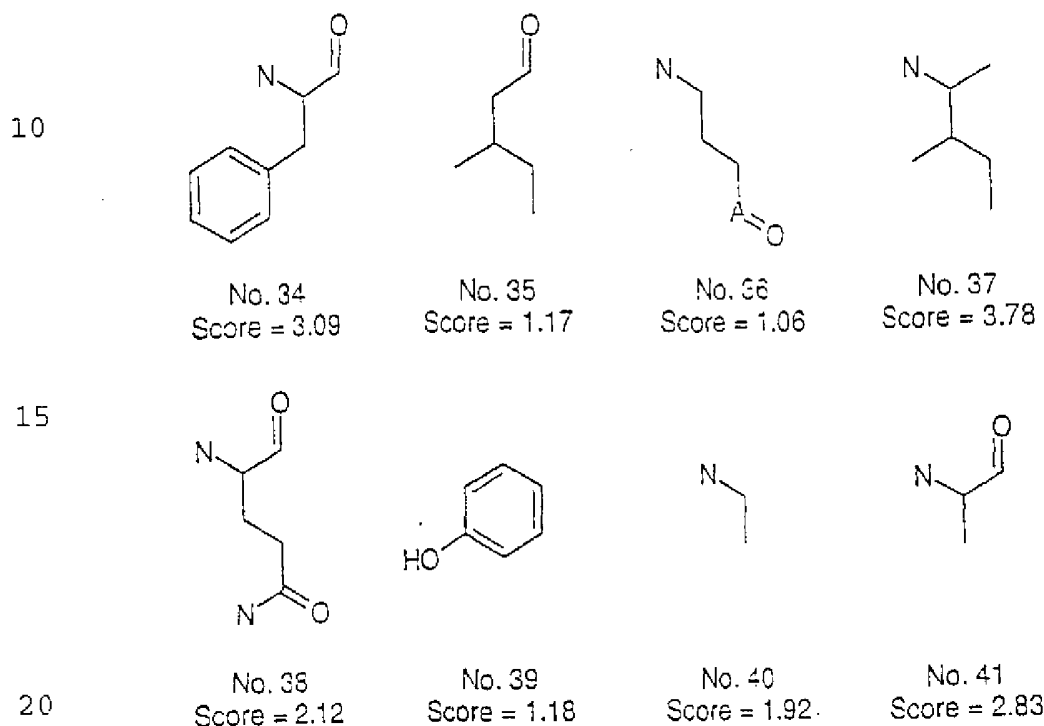


No. 33

Toto jsou příklady chemických determinantů odvozených z aminokyselin a hlavního peptidového řetězce a použitých pro analýzu. Seznam receptorových ligandů byl sestaven prostřednictvím prozkoumání vědecké literatury a analyzován způsobem podle předkládaného vynálezu s využitím měřítka (III) asociace, funkce (IV) skóre a seznamu chemických determinantů zahrnujících různé fragmenty dvaceti běžných aminokyselin, doplněné fragmenty struktury hlavního peptidového řetězce (NH-CH-CO-)₃. Příklady některých z determinantů odvozených z tryptofanu jsou znázorněny v prvních dvou řádcích. Tyto determinanty byly buď přesnými fragmenty (například determinanty č. 18, č. 19, č. 20, č. 21 a č. 26), sestavami přesných fragmentů (například determinant č. 22), nepřesnými fragmenty (například determinanty č. 23, č. 24 a č. 25), nebo sestavami přesných a nepřesných fragmentů (nejsou ilustrovány). Spodní dvě řady ilustrují příklady determinantů, odvozených ze struktury hlavního peptidového řetězce (NH-CH-CO-)₃ a reprezentujících přesné fragmenty (determinanty č. 29, č. 31 a č. 32) a nepřesné fragmenty (determinanty č. 27, č. 28, č. 30 a č. 33). Symboly: A reprezentuje C nebo S; B reprezentuje C nebo N; E reprezentuje C, N, O nebo S.

Skórování fragmentů prostřednictvím vzorce (IV) vedlo na identifikaci množství chemických determinantů majících

hodnoty skóre větší než 1, což indikuje, že odpovídající struktury měly pravděpodobnost menší než 1 z 20, že budou obsaženy v podskupině farmakologicky aktivních sloučenin pouze na základě náhodného výběru ($p < 0,05$). Příklady takovýchto determinantů jsou znázorněny níže společně s jejich příslušnými hodnotami skóre:



Toto jsou příklady vysoce skórujících chemických determinantů identifikovaných v prvním kole analýzy. Soubor receptorových ligandů byl analyzován podle předkládaného vynálezu prostřednictvím skórování výše ilustrovaných chemických determinantů a rovněž množství dalších s využitím funkce (IV) skóre. Hodnoty větší než jedna indikují, že determinant má pravděpodobnost menší než 1 z 20, že se bude vyskytovat v podskupině receptorových ligandů pouze na základě náhodného výběru. Obrázek výše ilustruje některé z

více skórujících chemických determinantů, které byly identifikovány tímto procesem.

Tyto determinanty byly tudíž akceptovány jako reprezentativní pro jednu nebo více aminokyselin, obsažených v primární sekvenci peptidového hormonu, a byly sestaveny do druhého seznamu. Potom byly opakovány výpočty s použitím vzorce (IV) za účelem identifikování nejvíce skórujících kombinací těchto nových determinantů, z nichž množství získalo hodnoty skóre větší než 10. Struktura chemického determinantu s nejvyšším skóre, označená jako determinant č. 42, potom byla následně porovnávána se strukturami 800 dipeptidů zahrnutých v různých kombinacích 20 aminokyselin, přičemž bylo zjištěno, že pouze jedna dipeptidová sekvence, označená A_1-A_2 , obsahovala celý determinant č. 42. Tento výsledek byl uvažován pro indikaci, že příslušný hormon s největší pravděpodobností zahrnuje sekvenci A_1-A_2 někde ve své primární struktuře a navíc, že alespoň jedna z těchto dvou aminokyselin má důležitou úlohu při vázání endogenního ligandu na jeho receptor. Ověření sekvence hormonu odhalilo, že skutečně zahrnuje předvídanou sekvenci A_1-A_2 , což je skutečnost, která bylo vypočítána, že má pravděpodobnost pouze 0,019 výskytu na základě náhodného výběru. Překvapivě další práce prokázala, že peptidy, obsahující mutaci v poloze A_2 sekvence A_1-A_2 (například A_1-A_3 nebo A_1-A_4 namísto A_1-A_2 , kde A_1 , A_2 , A_3 a A_4 jsou různé aminokyseliny) vykazovaly znatelně nižší afinitu pro receptor, což ilustruje, že alespoň jeden ze dvou předvídaných zbytků, skutečně tvoří důležitou součást, která tvoří základ pro biologickou funkci příslušného hormonu. Souhrnně tedy tyto výsledky demonstrují, že způsob podle předkládaného vynálezu umožňuje identifikovat

biologicky aktivní části peptidových ligandů, což je užitečné v programech lékařské chemie, které se zaměřují na racionální konstruování, například, inhibitorů peptidomimetických enzymů a/nebo receptorových ligandů.

5

Příklad č. 13 - Předvídání proteinových interakcí

Předkládaný vynález rovněž umožňuje předvídat existenci proteinových interakcí způsobem analogickým k postupu popsanému v předcházejícím příkladu. Pro ilustraci byl realizován screening iontového kanálku tak, jak bylo popsáno v příkladu č. 3, což vedlo na identifikaci více než 10 dvou tuctů molekul vykazujících alespoň 40% inhibici při testování při koncentraci 5 μ M. Chemické struktury těchto inhibitorů byly sestaveny do seznamu, který byl analyzován 15 podle popisu v příkladu č. 12. To vedlo na identifikaci sérií vysoce skórujících, chemických determinantů odvozených z aminokyselin a peptidového hlavního řetězce, přičemž po další analýze bylo zjištěno, že příslušný kanálek s největší pravděpodobností bude interagovat s inhibitory peptidem 20 nebo proteinem specificky obsahujícím určitou dipeptidovou sekvenci, označenou A_5-A_6 . Překvapivě tyto inhibitory proteiny již byly předtím popsány v literatuře, přičemž všechny z nich obsahovaly "kanálek inhibující" oblast 25 aminokyselin, obsahující přesně předpovězenou dipeptidovou sekvenci A_5-A_6 . Protože může být stanoveno, že jakákoliv sekvence z 20 aminokyselin má pravděpodobnost pouze 0,046, že bude obsažena v daném sekvenčním uspořádání dvou daných reziduí na základě náhodného výběru, lze odhadnout, že pravděpodobnost správně předpovězené existence 30 rozdílných dipeptidových sekvencí, existujících ve dvou

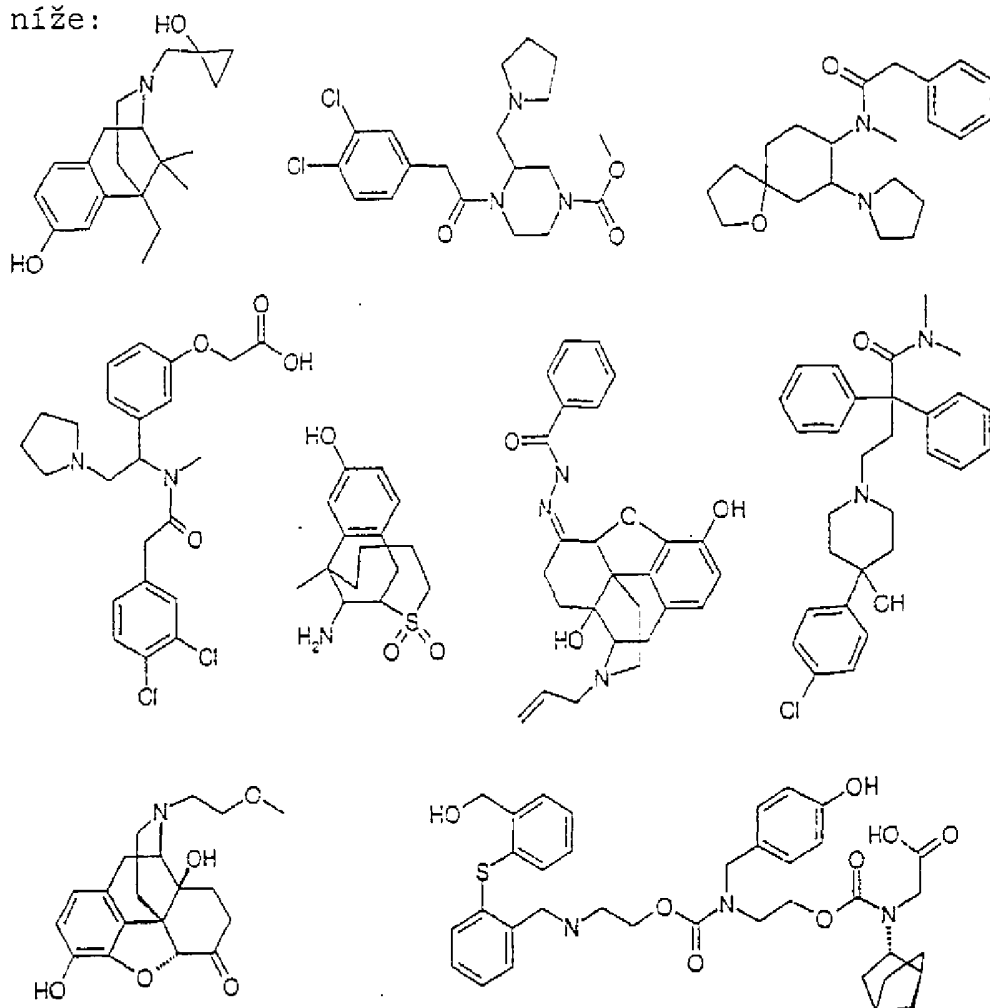
20

25

30

generovány množstvím způsobů, jako je, ale bez omezení na tyto postupy, provedení NMR studií, měření souhlasných změn prostřednictvím kruhového dichroismu, měření interakcí proteinů a ligandů prostřednictvím povrchové plazmové rezonance, nebo v případě samostatného receptoru provedení testu s konstitutivně aktivovanými mutanty příslušného receptoru.

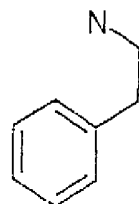
Pro ilustraci tohoto konceptu předpokládejme, že experimenty výše popisovaného typu jsou prováděny na samostatném receptoru, což poskytuje struktury ilustrované níže:



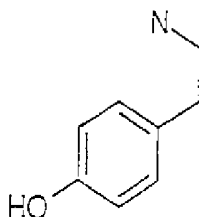
To je hypotetický seznam struktur analyzovaných pro biologicky aktivní chemické determinanty. Devět struktur,

znázorněných výše, bylo analyzováno podle předkládaného vynálezu, jak bylo popsáno v příkladu č. 12, s využitím výše zmiňovaného seznamu chemických determinantů odvozených z aminokyselin a peptidového základního řetězce.

Analýza struktur podle popisu v příkladu č. 12 vedla na indentifikaci množství chemických determinantů odvozených z aminokyselin a peptidového základního řetězce s hodnotami skóre většími než 1. Příklady takovýchto determinantů jsou znázorněny níže společně s odpovídajícími hodnotami jejich skóre.



No. 43
Score = 4.43

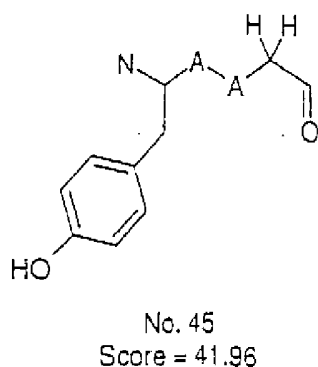


No. 44
Score = 4.90

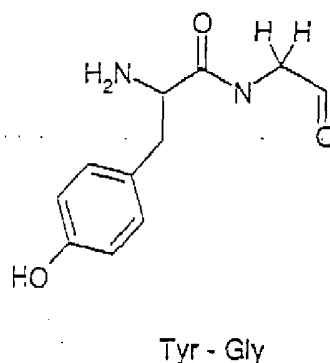
Toto jsou příklady vysoce skórujících chemických determinantů identifikovaných v prvním kole analýzy. Soubor hypotetických receptorových ligandů byl analyzován podle vynálezu prostřednictvím skórování chemických determinantů znázorněných v prvním panelu příkladu č. 12 a rovněž množství dalších prostřednictvím funkce (IV) skóre. Hodnoty větší než jedna indikovaly, že determinant měl pravděpodobnost menší než 1 z 20, že se bude vyskytovat v podskupině ligandů pouze na základě náhodného výběru. Výše jsou znázorněny dva chemické determinanty s nejvyšším skóre, které byly identifikovány uvedeným procesem.

Z těchto příkladů je zřejmé, že determinanty č. 43 a č. 44 mohou být obsaženy pouze v chemických strukturách aminokyselin fenylalaninu a tyrosinu. Z toho lze tedy usuzovat, že peptidy, které interagují se samostatným receptorem, pravděpodobně obsahují reziduum buď tyrosinu nebo fenylalaninu ve svých sekvencích a že tato rezidua pravděpodobně mají důležitou úlohu buď při vázání ligandů a/nebo aktivaci receptorů těmito peptidy. Pokud jsou vysoce skórující determinanty č. 43 a č. 44 následně opětovně analyzovány, aby se zjistilo, zda kombinace s fragmenty dalších aminokyselin neposkytují dokonce ještě vyšší hodnoty skóre, mohou být dále identifikovány fragmenty, jako je determinant č. 45 ilustrovaný na následujícím panelu A.

A



B



Tyto panely znázorňují vysoce skórující chemické determinanty identifikované v druhém kole analýzy. Chemické determinanty, jako jsou výše popsané determinanty, byly opětovně analyzovány podle předkládaného vynálezu pro stanovení, zda kombinace s fragmenty dalších aminokyselin by nevytvořily struktury s ještě vyššími hodnotami skóre. Jedna z nich, označená jako determinant č. 45 (panel A), vykazovala hodnotu skóre větší než 40. Překvapivě celý tento determinant

č. 45 je obsažen ve struktuře dipeptidové sekvence Tyr-Gly (panel B), což znamená, že endogenní ligand příslušného samostatného terče obsahuje dipeptidovou sekvenci Tyr-Gly ve své primární struktuře.

protože je zjevné, že celý determinant č. 45 je obsažen ve struktuře dipeptidu tyrosin-glycin (Tyr-Gly), lze usuzovat, že samostatné ligandy, které hledáme, s největší pravděpodobností obsahují sekvenci Tyr-Gly někde v jejich primárních strukturách. Na základě této informace mohou být podrobeny screeningu databáze sekvencí aminokyselin, aby se tak identifikovaly známé a/nebo samostatné ligandy obsahující předvídanou Tyr-Gly sekvenci, které po selekci a vyjádření mohou být testovány v původním testu biochemického screeningu. Alternativně může být chemický determinant č. 45 přímo použit pro sestavení souborů sloučenin potenciálních mimetik Tyr-Gly.

Nakonec je vhodné uvést, že chemické struktury, použité v tomto příkladu, jsou vlastně agonisty opioidních receptorů, nalezené v literatuře, a že přirozeně se vyskytující agonisty opioidních receptorů, to jest dynorfin A, β -endorfin, leu-enkefalin a met-enkefalin, všechny obsahují předvídanou sekvenci Tyr-Gly ve svých primárních strukturách. Protože reziduum tyrosinu bylo ilustrováno jako bezvýhradně požadované pro aktivitu opioidního agonistu, tento příklad dále ilustruje schopnost předkládaného vynálezu identifikovat biologicky aktivní části receptorových ligandů. Je rovněž patrné, že výše popisované odhady mohou být zlepšeny prostřednictvím použití alternativních algoritmů využívajících proměnné x, y, z a N, jako je například Fischerův přesný test. Ovšem bylo analyzováno pouze devět

struktur s použitím způsobu, pro který nebyla provedena adekvátní korekce pro vzorky malých velikostí, což může vést k tomu, že hodnota skóre 41,96 pro determinant č. 45 může být poněkud nadhodnocena.

5

Příklad č. 15 - Identifikace endogenních modulátorů lékových terčů

Osobě v oboru znalé je zřejmé, že předkládaný vynález může být rovněž aplikován na identifikaci endogenních modulátorů lékových terčů. Pro uvedení příkladu byl rozvinut funkční test pro příslušný iontový kanálek při léčení neurodegenerace. Soubor sloučenin byl podroben screeningu a výsledný seznam inhibitorů byl analyzován na přítomnost biologicky aktivních chemických determinantů, jak bylo popsáno v příkladu č. 2. To vedlo na identifikaci vysoce skórujícího chemického determinantu, o kterém bylo zjištěno, že je obsažen v podskupině molekul endogenně vytvářených v eukaryotických buňkách. Odpovídající sloučeniny byly následně opatřeny (nakoupeny) a testovány v testu, přičemž bylo zjištěno, že příslušný kanálek je selektivně inhibován (blokován) podmikromolárními koncentracemi určité podtřídy buněčného fosfolipidu, který velmi překvapivě byl již předtím sdružen s neuronální apoptózou prostřednictvím neznámých mechanismů jiných skupin. V souhrnu tedy tyto výsledky demonstrují, že předkládaný vynález umožňuje identifikaci endogenních modulátorů lékových terčů.

30

Příklad č. 16 - Identifikace klamně kladných experimentálních výsledků

Byl rozvinut enzymatický test pro proteinovou kinázu, o které se předpokládá, že má důležitou úlohu při imunitní reakci. Podle předkládaného vynálezu byl sestaven soubor sloučenin pro screening na daný terč, přesněji podle popisu v příkladu č. 2. Sloučeniny souboru potom byly následně testovány v testu s koncentrací 5 μM , což vedlo na identifikaci 35 molekul vykazujících inhibiční aktivitu alespoň 40%. Struktury těchto sloučenin byly analyzovány s použitím zjednodušené varianty vzorce (II) jako funkce skóre, a odpovídající hodnoty skóre byly přímo porovnávány s hodnotami statistické tabulky, což poskytlo odhady pravděpodobností, že se dané chemické determinanty vyskytují mezi podskupinou 35 farmakologicky aktivních sloučenin pouze na základě náhodného výběru.

S použitím prahové hodnoty pro pravděpodobnost náhodného výskytu $p < 0,05$ bylo stanoveno, že 14 z 35 inhibitorů s největší pravděpodobností reprezentuje klamně kladné výsledky. Následné opětovné testování 14 sloučenin v testu potvrdilo tuto hypotézu, což ilustruje, že předkládaný vynález umožňuje identifikaci klamně kladných experimentálních výsledků.

Příklad č. 17 - Identifikace klamně záporných experimentálních výsledků

Provedením výpočtů analogických k výpočtům podle popisu v příkladu č. 16 předkládaný vynález dále umožňuje identifikaci klamně záporných experimentálních výsledků. Pro

uvedení příkladu byly analyzovány chemické struktury sérií inhibitorů fosfatázy na přítomnost farmakologicky aktivních chemických determinantů, jako bylo popsáno v příkladu č. 16. Výsledné, nejvíce skórující chemické determinanty byly
5 použity jako farmakologicky aktivní "otisky" pro provedení podstrukturních průzkumů v seznamu chemických struktur odpovídajících sloučeninám, které byly původně testovány v testu. To odhalilo množství molekul, které obsahovaly jeden nebo více z výše zmiňovaných chemických determinantů, ale
10 které byly nicméně přesto identifikovány jako negativní v testu screeningu. Odpovídající molekuly byly následně opětovně testovány v testu, přičemž bylo zjištěno, že více než 15 % z nich bylo klamně negativních, přičemž dokonce jedna sloučenina vykazovala podmikromolární inhibitorní
15 aktivitu. Tyto výsledky jasně demonstrují, že způsob podle předkládaného vynálezu umožňuje identifikaci klamně záporných experimentálních výsledků.

20 **Příklad č. 18 - Provedení kvantitativních konfiguračních a konformačních analýz**

V dále zlepšeném provedení předkládaného vynálezu je možné rovněž využít algoritmy zahrnující různé kombinace proměnných x, y, z a N pro kvantitativní konformační a/nebo
25 konfigurační analýzu. Pro ilustrování této možnosti je zřejmé z výsledků, uvedených v příkladu č. 4, že struktura farmakologicky aktivního, proteázu inhibujícího "otisku", znázorněná v panelu B příkladu č. 4, není ani konfiguračně ani konformačně definována. Skutečně je nemožné z
30 reprezentace této struktury sdělit, zda je vzhledem ke dvěma karbonylovým nebo sulfonylovým skupinám trans-oid nebo

15

20



5

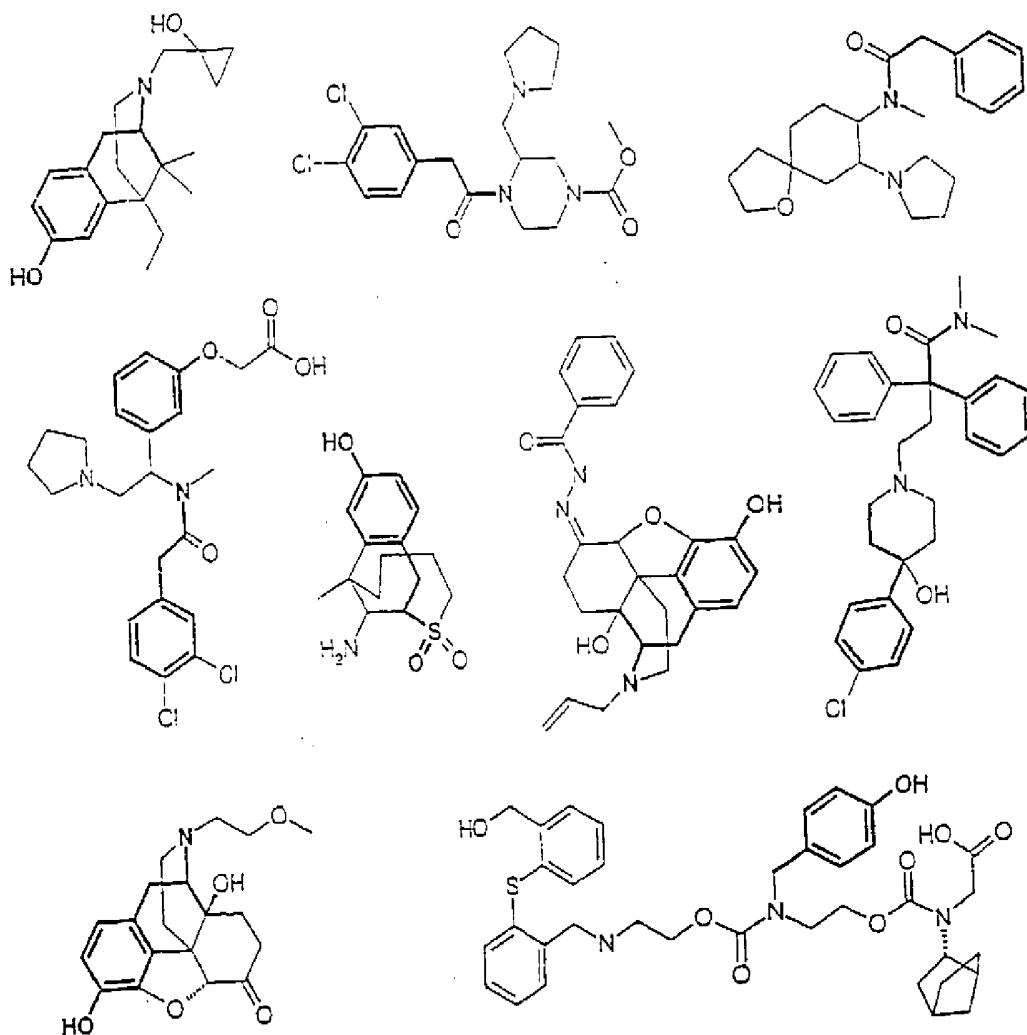
10

20

25

Příklad č. 19 - Provádění průzkumů podobnosti

Z předcházejících příkladů je zřejmé, že koncept molekulární podobnosti, jak je nahlížen způsobem podle předkládaného vynálezu, je nápadně rozdílný od toho, co je obecně akceptováno jako význam tohoto termínu. Například sloučeniny v hypotetickém seznamu podle příkladu č. 14 jsou velmi nepodobné, dokud zde není zjevný způsob pro zatřídění těchto devíti molekul do jedné chemické rodiny s použitím klasických shlukovacích technik. Přesto nicméně bylo v příkladu č. 14 ilustrováno, že tyto sloučeniny jsou ve skutečnosti extrémně podobné v té míře, že každá z nich obsahuje alespoň jeden výskyt chemického determinantu, který je reprezentativním fragmentem aminokyseliny tyrosinu; viz následující panel:



5

10

15

20

25



velké podíly podobnosti ve smyslu předkládaného vynálezu s farmakologicky aktivními referenčními sloučeninami.

Příklad č. 20 - Analyzování rozmanitosti souborů sloučenin

Předkládaný vynález může být dále použit pro analyzování rozmanitosti souboru sloučenin způsobem analogickým k tomu, co bylo popsáno v předcházejícím příkladu. V této souvislosti je zjevné osobě v oboru znalé, že koncept chemických determinantů může být snadno využít pro porovnávání daného souboru sloučenin s jakýmkoliv dalším. Například může být vybrán soubor sloučenin pro vysoce výkonný screening prostřednictvím analyzování odpovídajícího seznamu chemických struktur podle předkládaného vynálezu, přičemž jako referenčního souboru "lékových" molekul se použije referenční soubor chemických struktur, jako jsou chemické struktury obsažené v databázích Merck Index, Derwent, MDDR nebo Pharmaprojects. V tomto případě molekuly, jejichž struktury jsou v podstatě zahrnuty v nízce skórujících chemických determinantech, jsou přítomné ve velké části referenčních struktur. Naproti tomu molekuly, které jsou v podstatě zahrnuty ve vysoce skórujících chemických determinantech, jsou posuzovány jako "nelékové", protože stejné determinanty jsou pouze slabě reprezentovány v souboru referenčních sloučenin. Tyto informace jsou velmi užitečné pro konstruování objevovacích experimentů, protože pomáhají výzkumníkovi při identifikování chemických struktur, které by měly být začleněny do nebo vyloučeny ze souboru sloučeniny pro provádění screeningu. V tomto kontextu je zjevné, že pro

tento účel může být použito množství algoritmů zahrnujících různé kombinace proměnných x , y , z a N .

Příklad č. 21 - Speciální algoritmy

Je zcela zřejmé, že předcházející příklady neposkytují vyčerpávající seznam všech algoritmů, využívajících různé kombinace proměnných x , y , z a N , které mohou být použity pro provádění diskrétní podstrukturní analýzy. V této souvislosti je rovněž zcela zřejmé osobě v oboru znalé, že funkce (XII), (XIII) a (XIV) skóre mohou být rovněž použity pro adresování množství otázek prezentovaných v předcházejících příkladech. V některých příkladech je skutečně dokonce mnohem vhodnější ze statistického hlediska využít jeden z těchto vzorců namísto jednoho ze vzorců explicitně uváděných v příkladech. Protože je ale předkládaný vynález primárně konstruován pro identifikování chemických determinantů obsažených v seznamu chemických struktur, které jsou s největší pravděpodobností základem daného biologického účinku, byl tento popis primárně zaměřen na relativní skórování a následné roztrídění chemických determinantů. Přesto jsou nicméně níže uvedeny vzorce (XII), (XIII) a (XIV) pro případ, že a) je požadován přesný odhad pravděpodobnosti výskytu při náhodném výběru pro malé soubory vzorků (viz vzorec (XII), kde s odpovídá nejmenší hodnotě mezi proměnnými x , $(y-x)$, $(z-x)$ a $(N-y-z+x)$); b) předpokládá se, že poměrně vážený odhad současných příspěvků dvou determinantů je vhodnější pro použití v příkladu č. 8 (viz vzorec (XIII), kde d odpovídá počtu samostatných chemických determinantů); nebo c) považuje se za důležité odhadnou sledované účinky při vyhodnocování současných příspěvků dvou vzájemně propojených

chemických determinantů (viz vzorec (XIV)). V tomto kontextu jsou definice proměnných x , y , z , a N přesně takové, jak bylo popsáno v předcházejícím popisu.

$$(XII) \quad \text{Score} = \sum_{i=1}^s \left(\frac{y! (N-y)! z! (N-z)!}{x! (y-x)! (z-x)! (N-y-z+x)! N!} \right)$$

$$(XIII) \quad \text{Score} = \sum_{i=1}^d \left(\frac{Nx - yz}{N} \right) / \sqrt{\sum_{i=1}^d \left(\frac{z(N-z)y(N-y)}{N^3} \right)}$$

$$(XIV) \quad \text{Score} = \frac{(|y+z-N|-1)^2}{(N-y-z+2x)}$$

Nakonec je rovněž zjevné osobě v oboru znalé, že použití určitých proměnných ve funkcích skóre a/nebo algoritmech zkonstruovaných pro identifikování biologicky aktivních chemických determinantů, ale ne explicitně popsaných v předcházejících příkladech, může být matematicky ekvivalentní použití různých kombinací proměnných x , y , z a N . Pro názornou ilustraci funkce skóre, využívající proměnou q definovanou jako reprezentující počet neaktivních molekul, jejichž chemické struktury obsahují daný chemický determinant, je ekvivalentní použití proměnných x a y jako $q=y-x$. Podobně funkce skóre, využívající proměnnou r definovanou jako reprezentující celkový počet aktivních sloučenin, které neobsahují daný chemický determinant, je algebraicky ekvivalentní použití proměnných x a z , protože může být snadno ukázáno, že $r=z-x$. Rovněž funkce skóre, využívající proměnnou s definovanou jako reprezentující celkový počet neaktivních sloučenin, které neobsahují daný



chemický determinant, je ekvivalentní použití proměnných x , y , z a N , protože $s=N-y-z+x$. Nakonec algoritmy využívající proměnné t a u , reprezentující celkový počet molekul, jejichž struktury neobsahují daný determinant (t), respektive celkový počet neaktivních molekul (u), jsou ekvivalentní použití proměnných N , y a/nebo z , protože může být snadno ukázáno, že $t=N-y$ a $u=N-z$.

Příklad č. 22 - Mapování relativních příspěvků

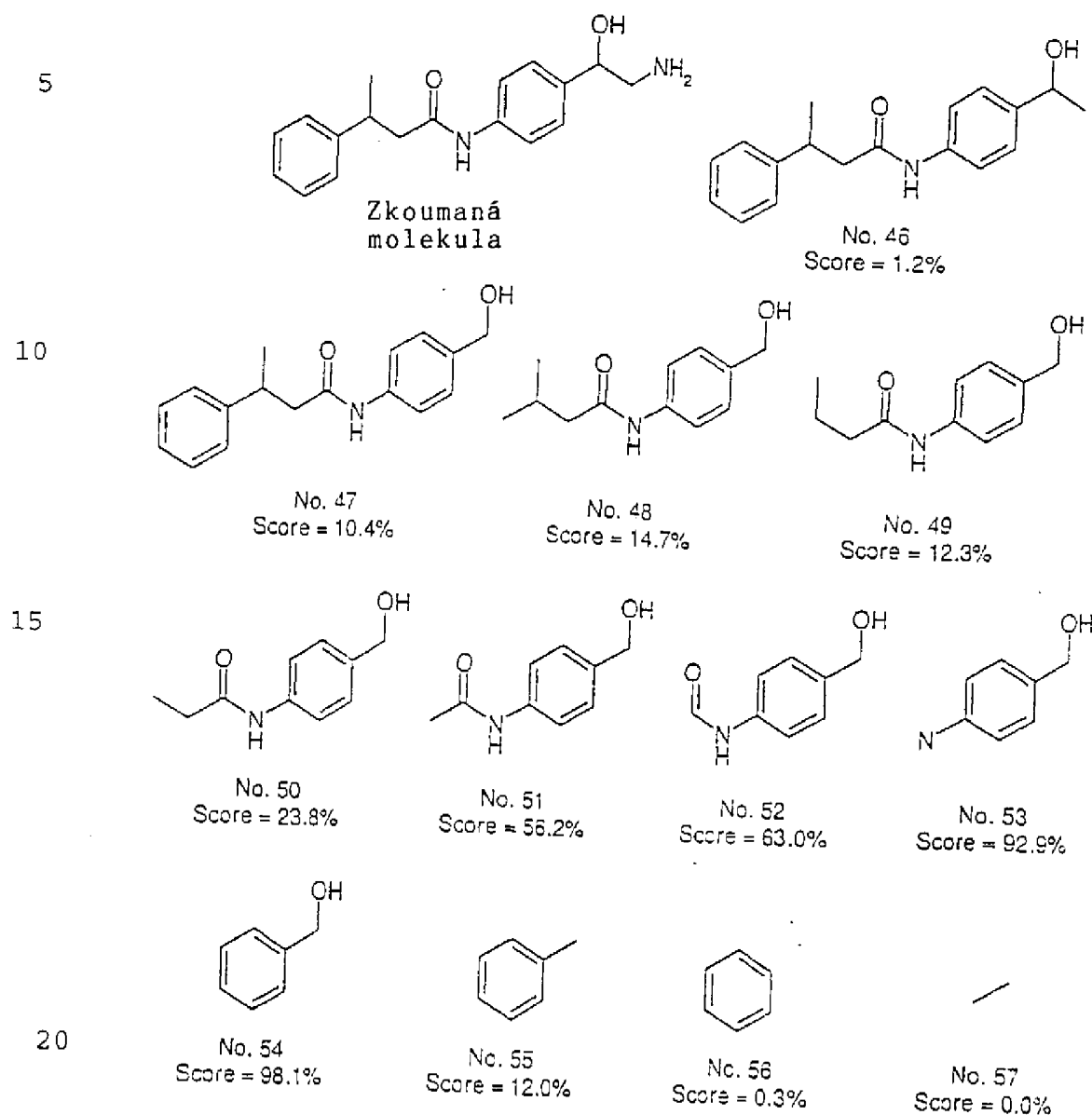
Předkládaný vynález rovněž umožňuje konstruování diagramů relativních příspěvků. To jsou grafické reprezentace chemických struktur, ve kterých jsou indikovány relativní příspěvky různých atomů, vazeb, fragmentů a/nebo podstruktur pro daný biologický výsledek prostřednictvím hodnot skóre, vypočítaných podle popisů v předcházejících příkladech. Ve výhodném provedení způsobu podle vynálezu jsou použity pravděpodobnostní hodnoty skóre, jako jsou ty, které byly vypočítány s použitím vzorce (XII), kde $P(A)$ reprezentuje pravděpodobnost, že daný chemický determinant je obsažen v podskupině biologicky aktivních struktur na základně náhodného výběru, což je vypočítáno s využitím vzorců používajících nejrozumnější kombinace proměnných x , y , z a N , jak bylo popsáno v předcházejícím.

(XII)

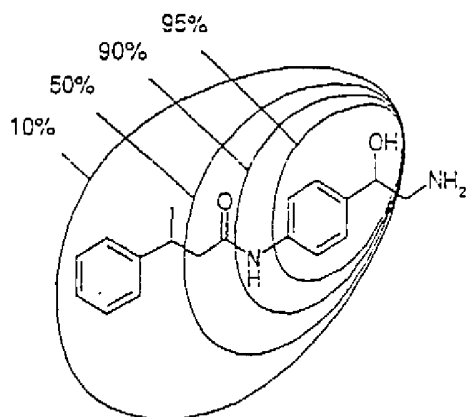
$$\text{Score} = [1 - P(A)] \cdot 100\%$$

V tomto kontextu je zcela zřejmé, že množství měřítek asociace a/nebo funkcí skóre je možné použít pro odhad $P(A)$.

Dva příklady diagramů relativních příspěvků budou nyní diskutovány poněkud podrobněji. Následující panel



ukazuje příslušnou molekulu doplněnou sériemi chemických determinantů sestavených z fragmentů stejné uvedené molekuly, které byly skórovány s použitím vzorce (XII) a modifikací měřítka (I) asociace pro stanovení $P(A)$. Obr. 15 znázorňuje stejné informace v grafické podobě, kde determinanty jsou vyneseny proti jejich příslušným hodnotám skóre. V tomto kontextu je zjevné, že stejné informace mohou být reprezentovány ve formě pravděpodobnostních obrysových map, jak je znázorněno na následujícím panelu:



Celkově jsou takovéto diagramy velmi užitečné pro konstruování souborů sloučenin, protože pomáhají výzkumníkovi při výběru sloučenin na základě matematických odhadů úspěšnosti náhodného výběru v daném testu, což snižuje potřebu spoléhat se na koncept molekulární rozmanitosti pro identifikování nových, biologicky aktivních chemických sérií. Tyto diagramy jsou rovněž zajímavé v lékařské chemii, protože reprezentace jaká je například znázorněná na panelu výše, jasně indikuje, které části molekuly mohou být přiměřeně modifikovány s minimálním rizikem ztráty farmakologické aktivity. Naproti takovéto grafy upozorňují toxikologa na to, které části toxické sloučeniny musí být modifikovány za účelem eliminování nežádoucího účinku.

Pro získání mapování relativních příspěvků, znázorněného výše a na obr. 15, byly podle předkládaného vynálezu skórovány chemické determinanty, odpovídající fragmentům biologicky aktivní molekuly, s užitím funkce skóre využívající proměnné x , y , z a N , což umožňovalo přímý odhad pravděpodobnosti náhodného výskytu v souboru aktivních molekul ($P(A)$). Odpovídající hodnoty $P(A)$ byly transformovány s použitím funkce (XII) skóre, poskytující pravděpodobnostní hodnoty skóre pro každý determinant, což odráží relativní

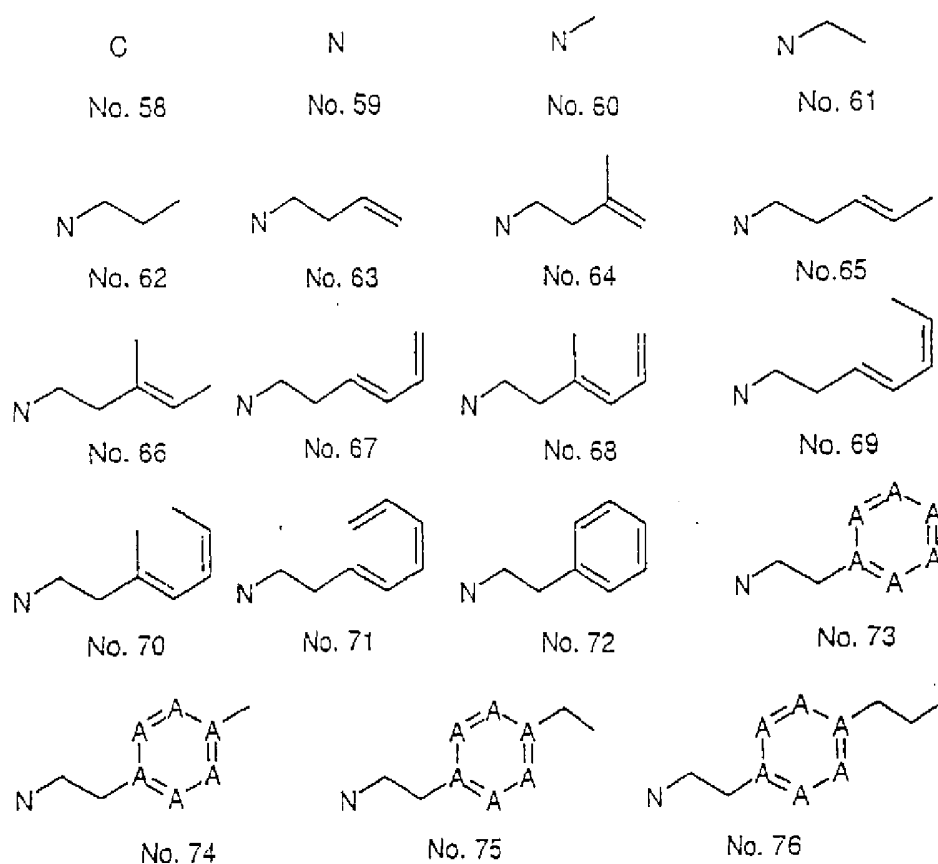
pravděpodobnost, že odpovídající chemická struktura byla základem příslušné biologické aktivity. Hodnoty mohou být ilustrovány jako na obr. 15, který je grafickou reprezentací hodnot skóre pro různé chemické determinanty. Chemický determinant č. 54 odpovídá lokálnímu maximu této série. Nebo mohou být tyto hodnoty ilustrovány jako na výše uvedeném panelu, který je pravděpodobnostní obrysovou mapou indikující, který fragment nebo sektor příslušné chemické struktury s největší pravděpodobností přispívá k biologické aktivitě (determinant č. 54 obsažený v oblasti vymezené 95% obrysovou linií). Další způsob prezentování uvedených hodnot je znázorněn na obr. 11.

Příklad č. 23 - Ekvivalence funkcí skóre

Funkce skóre, použité v předcházejících příkladech jsou všechno způsoby pro identifikování chemických determinantů, které jsou s největší pravděpodobností základem daného biologického, farmakologického a/nebo toxikologického účinku. Zatímco je osobě v oboru znalé zcela zjevné, že určitá měřítko asociace a/nebo funkce skóre jsou nejlépe využita pro adresování pouze určitých typů otázek, při použití ve způsobu podle předkládaného vynálezu podle výše uvedeného popisu každý vzorec umožňuje identifikaci stejného, nejvíce skórujícího chemického determinantu, který s největší pravděpodobností je základem daného biologického účinku. Jako takové vzorce, prezentované v předcházejících příkladech, jsou funkčně ekvivalentní ve smyslu diskrétní podstrukturní analýzy.

Za účelem uvedení názorného příkladu byla provedena analýza chemických struktur 131 agonistů receptoru dopamin D_2

osm krát paralelně s využitím osmi měřítek asociace a funkcí skóre, obsahujících různé kombinace proměnných x , y , z a N , jak je znázorněno níže. Studie byla prováděna jako podle předchozího popisu, zejména přidáním chemických struktur 101207 molekul, které byly popsány jako mající nulový účinek na receptor dopamin D_2 , k prvnímu seznamu 131 molekul a skórováním sérií 19 chemických determinantů, znázorněnými níže, s užitím funkcí (XV) až (XXIII) skóre, přičemž znalý čtenář snadno nahlédne, že tyto funkce reprezentují stejné funkce, které byly použity v množství předcházejících příkladů, a nebo jejich blízce příbuzné varianty.



Toto jsou chemické determinanty skórované s osmi různými funkcemi skóre. 19 chemických determinantů, znázorněných výše, bylo skórováno s použitím funkcí (XV) až (XXII) a seznamu chemických struktur katalogizovaných pro

aktivitu agonistu receptoru dopamin D_2 . Použité funkce jsou následující:

$$(XV) \quad \text{Score} = MW \cdot (x/z)$$

$$5 \quad (XVI) \quad \text{Score} = (x/z) - (y/N)$$

$$(XVII) \quad \text{Score} = Nx - yz$$

$$(XVIII) \quad \text{Score} = \frac{x(N-y-z-x)}{(z-x)(y-x)}$$

$$10 \quad (XIX) \quad \text{Score} = \frac{(|Nx - yz| - N/2)^2 N}{z(N-z)y(N-y)}$$

$$(XX) \quad \text{Score} = \frac{x(N-y-z-x)}{(z-x)(y-x)} e^{-2\sqrt{1/x+1/(y-x)+1/(z-x)+1/(N-y-z+x)}}$$

$$15 \quad (XXI) \quad \text{Score} = \frac{Nx - yz}{\sqrt{z(N-z)y(N-y)}}$$

$$(XXII) \quad \text{Score} = e^{[(x/z) - (z-x)/(N-z)]}$$

20
Obr. 16A až obr. 16H znázorňují odpovídající diagramy relativních příspěvků. Chemické determinanty, znázorněné v panelu výše, byly skórovány podle předcházejícího popisu a vyneseny proti jejich odpovídajícím hodnotám skóre. Obr. 16A znázorňuje skóre získaná s funkcí (XV), obr. 16B znázorňuje skóre získaná s funkcí (XVI), obr. 16C znázorňuje skóre získaná s funkcí (XVII), obr. 16D znázorňuje skóre získaná s funkcí (XVIII), obr. 16E znázorňuje skóre získaná s funkcí (XIX), obr. 16F znázorňuje skóre získaná s funkcí (XX), obr. 16G znázorňuje skóre získaná s funkcí (XXI), a obr. 16H

25
30

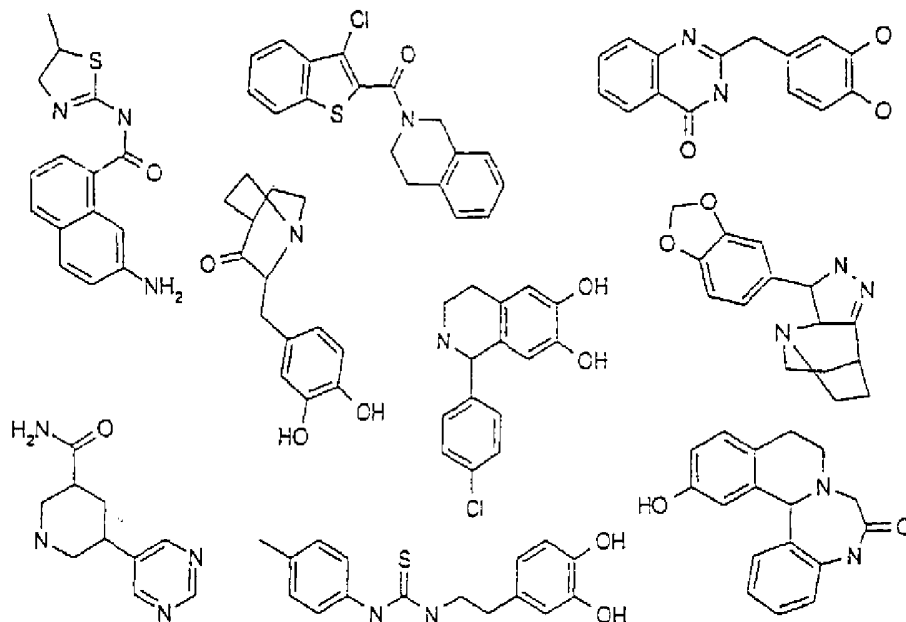
znázorňuje skóre získaná s funkcí (XXII). Každá funkce skóre nezměnitelně zvýrazňuje stejný chemický determinant (č. 73) jako mající největší pravděpodobnost, že je základem biologické aktivity.

5 Jak je znázorněno diagramy relativních příspěvků, prezentovanými na obr. 16A až obr. 16H, každá z osmi funkcí skóre správně identifikovala chemický determinant č. 73 jako odpovídající lokálnímu maximu, což znamená, že to je chemický
10 motiv s největší pravděpodobností zakládající aktivitu agonistu dopaminu D_2 v seznamu 19 testovaných determinantů. Překvapivě se různé funkce skóre lišily, pokud se týká ohodnocení méně skórujících chemických determinantů, takže například determinant č. 62 byl považován za důležitý pro biologickou aktivitu ohodnocením na třetím místě ve výpočtech
15 s použitím funkcí (XV), (XVI) a (XVII) skóre, zatímco s použitím funkce (XXII) skóre byl jako třetí hodnocen determinant č. 63, s použitím funkcí (XIX) a (XXI) skóre byl jako třetí ohodnocen determinant č. 65 a nakonec s použitím funkcí (XVIII) a (XX) skóre byl jako třetí hodnocen
20 determinant č. 66.

celkově tyto drobné rozdíly mají malou důležitost pro úspěšný výsledek metody, protože v každém případě méně skórující determinanty jsou vlastně fragmenty většího,
25 nejvíce skórujícího determinantu č. 73 (viz panel výše). Takto postačuje přímé použití chemického determinantu č. 73 a jeho fragmentů pro zkonstruování souborů sloučenin pro vysoce výkonný screening, protože tyto bude neměnně obsahovat struktury obsahující každý z méně skórujících determinantů. Vzorky typu sloučenin, které by mohly být začleněny do
30 takového souboru jsou znázorněny níže:

5

10



15

Tyto vzorky struktur jsou příklady sloučenin, které by mohly být vybrány pro začlenění do souboru sloučenin, zkonstruovaného pro identifikaci agonistů receptoru dopamin D_2 . Každá ze struktur, znázorněných výše, obsahuje chemický determinant č. 73 nebo jeho podstatnou část.

20

25

Ačkoliv tedy matematické zdůvodnění, které je podkladem konstruování, a použití osmi různých funkcí skóre je v každém případě jiné, všechny tyto postupy identifikují zcela stejný a tentýž chemický determinant, který je s největší pravděpodobností základem biologické aktivity. Jako takové jsou algoritmy, obsahující různé kombinace proměnných x , y , z a N nebo q , r , s , t a u , jak bylo zmiňováno v předcházejícím popisu, funkčně ekvivalentní ve smyslu předkládaného vynálezu.

30

Příklad č. 24 - Nástroje pro objevování léků na bázi informatiky

Z předcházejících příkladů je zcela zřejmé, že předkládaný vynález může být začleněn do jedné nebo více

či nikoliv, pro výběr chemických sloučenin, sérií a/nebo
koster pro vysoce výkonný screening, lékařskou chemii a/nebo
optimalizaci vodících struktur, přičemž uvedené
experimentální výsledky a hodnoty skóre se týkají jakékoliv
5 dané farmakologické, biochemické, toxikologické a/nebo
biologické vlastnosti.

• Proces, ve kterém je předkládaný vynález, jak je
popsán v kterémkoliv z předcházejících příkladů, použit pro
identifikování farmakologických modulátorů lékových terčů,
10 jako například, ale bez omezení na, receptorové ligandy,
inhibitory kinázy, modulátory iontových kanálků, inhibitory
proteázy, inhibitory fosfatázy a ligandy steroidních
receptorů.

• Proces, ve kterém je předkládaný vynález, jak je
15 popsán v kterémkoliv z předcházejících příkladů, přímo
použit, nebo využít ve zkonstruovaném počítačovém programu,
pro analýzu chemických struktur za účelem zvýšení potence
chemických sérií, zvýšení selektivity chemických sérií,
konstruování sloučenin s více farmakologickými účinky,
20 předvídání potenciálních sekundárních farmakologických
působení molekuly, předvídání potenciálních toxikologických
působení molekuly, identifikování biologicky aktivních částí
receptorových ligandů, předvídání potenciálních proteinových
interakcí, identifikování párů samostatných ligandů a
25 receptorů a/nebo identifikování endogenních modulátorů
lékových terčů. Poslední využití se vztahují zejména na
oblasti funkčních genomů a proteomů, ve kterých, například,
mohou být vybrány sekvence nukleotidů a/nebo aminokyselin pro
průzkumy na základě chemických struktur molekul
30 identifikovaných v testu biochemického screeningu a

zpracovaných podle vynálezu, jako například pro identifikaci samostatných ligandů.

5

10

15

20

25

30

P A T E N T O V É N Á R O K Y

1. Způsob provozování počítačového systému pro provádění diskrétní podstrukturní analýzy, **vyznačující se tím, že** zahrnuje kroky:

5 přístupu (210, 220, 410) k databázi (110, 115) molekulárních struktur, přičemž tato databáze je prozkoumatelná prostřednictvím informací o molekulárních strukturách a biologických a/nebo chemických vlastností;

10 identifikování (220) v uvedené databázi podskupiny molekul majících danou biologickou a/nebo chemickou vlastnost;

 stanovení (230, 420) fragmentů molekul v uvedené podskupině;

15 pro každý fragment vypočítání (230, 430, 610-650) hodnoty skóre, indikující příspěvek příslušného fragmentu k uvedené dané biologické a/nebo chemické vlastnosti; a

20 provedení (240, 250) opakování procesu analyzováním (250) stanovených fragmentů a vypočítaných hodnot skóre, přičemž nejprve se vybere alespoň jeden fragment, který má hodnotu skóre, indikující vysoký příspěvek k uvedené biologické a/nebo chemické vlastnosti, a potom se opakují kroky přístupu, identifikování, stanovení a vypočítání.

25 2. Způsob podle nároku 1, **vyznačující se tím, že** krok vypočítání hodnoty skóre obsahuje krok:

 vypočítání (610) počtu (x) molekul v uvedené podskupině molekul, které obsahují daný fragment.

3. Způsob podle nároků 1 nebo 2, **vyznačující se tím, že** dále zahrnuje krok:

30 identifikování v uvedené databázi druhé podskupiny

molekul, které nemají uvedenou biologickou a/nebo chemickou vlastnost;

přičemž krok vypočítání hodnoty skóre zahrnuje krok:

5 vypočítání (620) počtu (y) molekul v uvedené podskupině
a uvedené druhé podskupině molekul, které obsahují daný
fragment.

4. Způsob podle jednoho z nároků 1 až 3, **vyznačující se tím, že** uvedený krok vypočítání hodnoty skóre zahrnuje krok:

10 vypočítání (630) počtu (z) molekul v uvedené podskupině
molekul.

5. Způsob podle jednoho z nároků 1 až 4, **vyznačující se tím, že** dále zahrnuje krok:

15 identifikování v uvedené databázi druhé podskupiny
molekul, které nemají uvedenou danou biologickou a/nebo
chemickou vlastnost;

přičemž uvedený krok vypočítání hodnoty skóre zahrnuje
krok:

20 vypočítání (640) celkového počtu (N) molekul v uvedené
podskupině a uvedené druhé podskupině molekul.

6. Způsob podle jednoho z nároků 1 až 5, **vyznačující se tím, že** proces opakování se provádí výběrem fragmentů
následujícího kola, které mají vyšší molekulární hmotnost,
než fragmenty předcházejícího kola.

25 7. Způsob podle jednoho z nároků 1 až 6, **vyznačující se tím, že** dále zahrnuje kroky:

výběru (710) fragmentu na základě vypočítaných hodnot
skóre;

30 analyzování (810) struktury vybraného fragmentu;

lokalizace (820) zobecněné položky ve struktuře

vytvoření (740) nové podskupiny molekul, obsahující vyjmuté a vybrané sloučeniny.

11. Způsob podle jednoho z nároků 1 až 10, **vyznačující se tím, že** dále zahrnuje krok:

5 vytvoření (230) knihovny (120) fragmentů, obsahující stanovené fragmenty a vypočítané hodnoty skóre.

12. Způsob podle jednoho z nároků 1 až 11, **vyznačující se tím, že** databáze je autorizovaná databáze.

10 13. Způsob podle jednoho z nároků 1 až 12, **vyznačující se tím, že** databáze je veřejná databáze.

14. Způsob podle jednoho z nároků 1 až 13, **vyznačující se tím, že** databáze je databází sekvencí aminokyselin a/nebo nukleových kyselin, a uvedená biologická a/nebo chemická
15 vlastnost je daný účinek na příslušný protein.

15. Způsob podle jednoho z nároků 1 až 14, **vyznačující se tím, že** uvedená biologická a/nebo chemická vlastnost je farmakologická vlastnost a způsob se použije pro objevování
20 léků.

16. Způsob podle jednoho z nároků 1 až 15, **vyznačující se tím, že** dále zahrnuje krok:

25 sestavení (260) souboru sloučenin, které obsahují alespoň jeden ze stanovených fragmentů.

17. Způsob podle nároku 16, **vyznačující se tím, že** dále zahrnuje krok:

testování sloučenin z uvedeného sestaveného souboru na uvedenou danou biologickou a/nebo chemickou vlastnost.

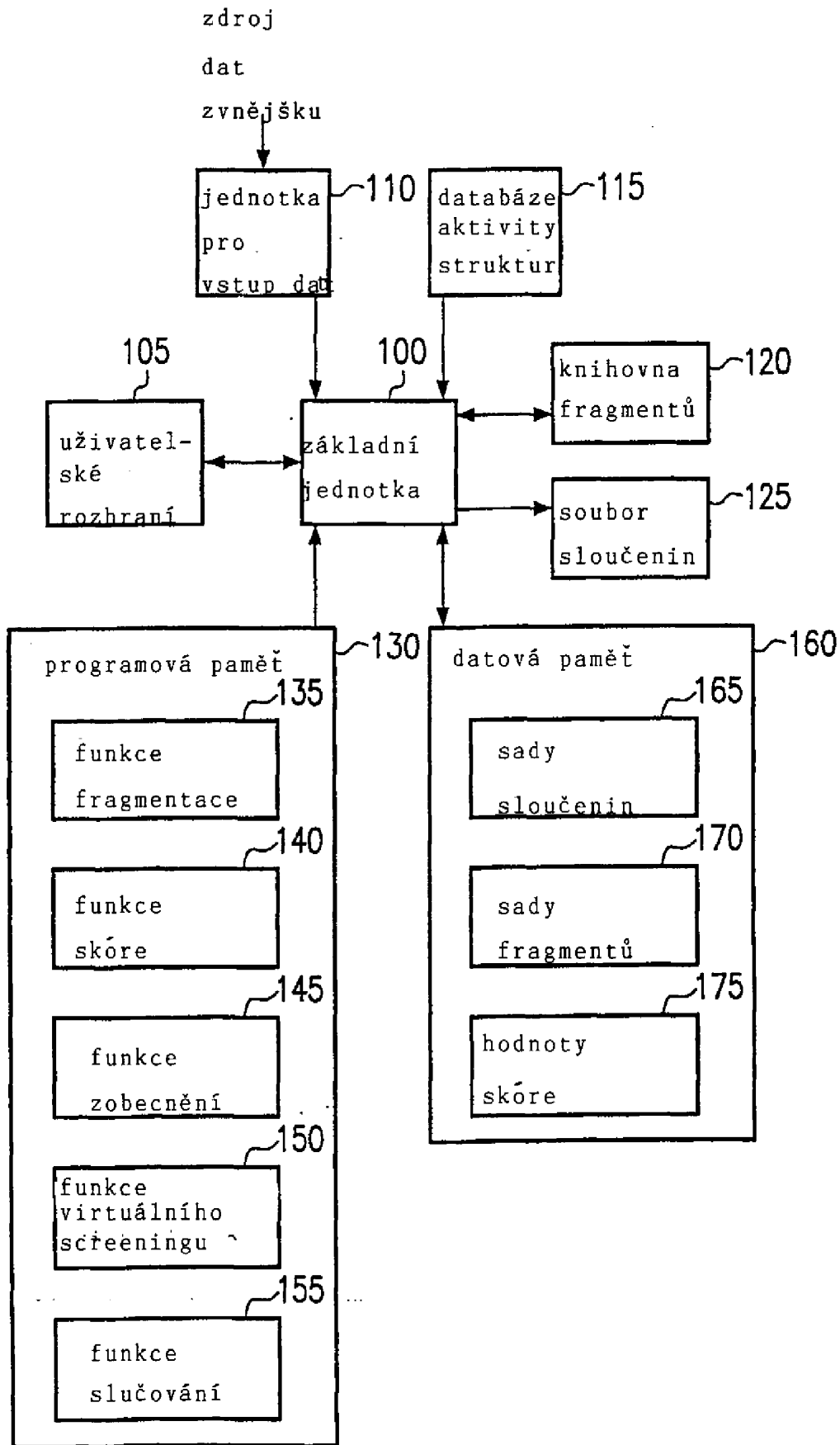


Fig. 1

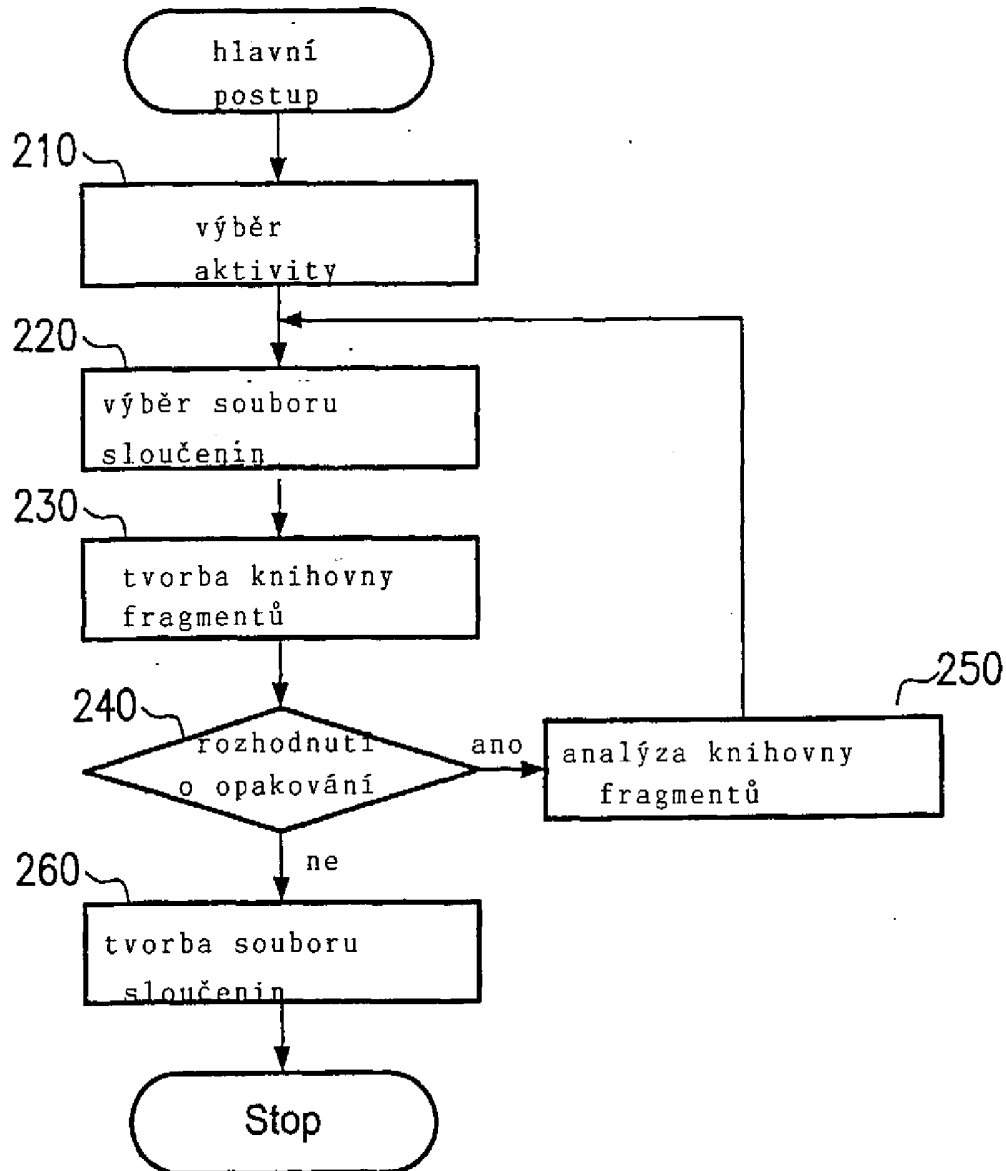


Fig. 2

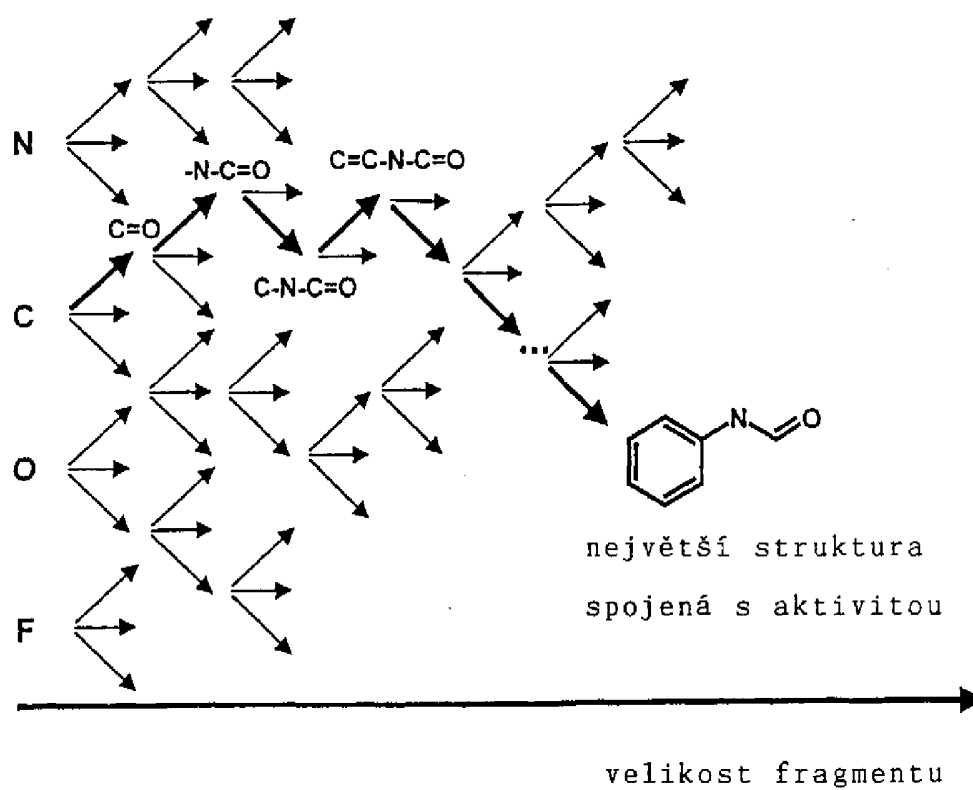
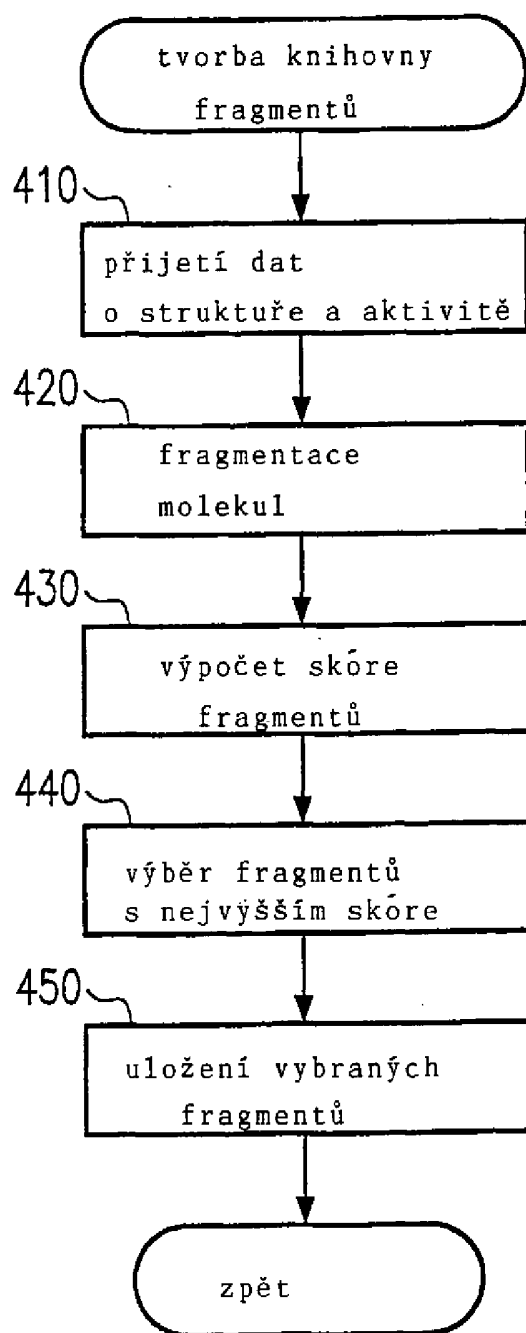
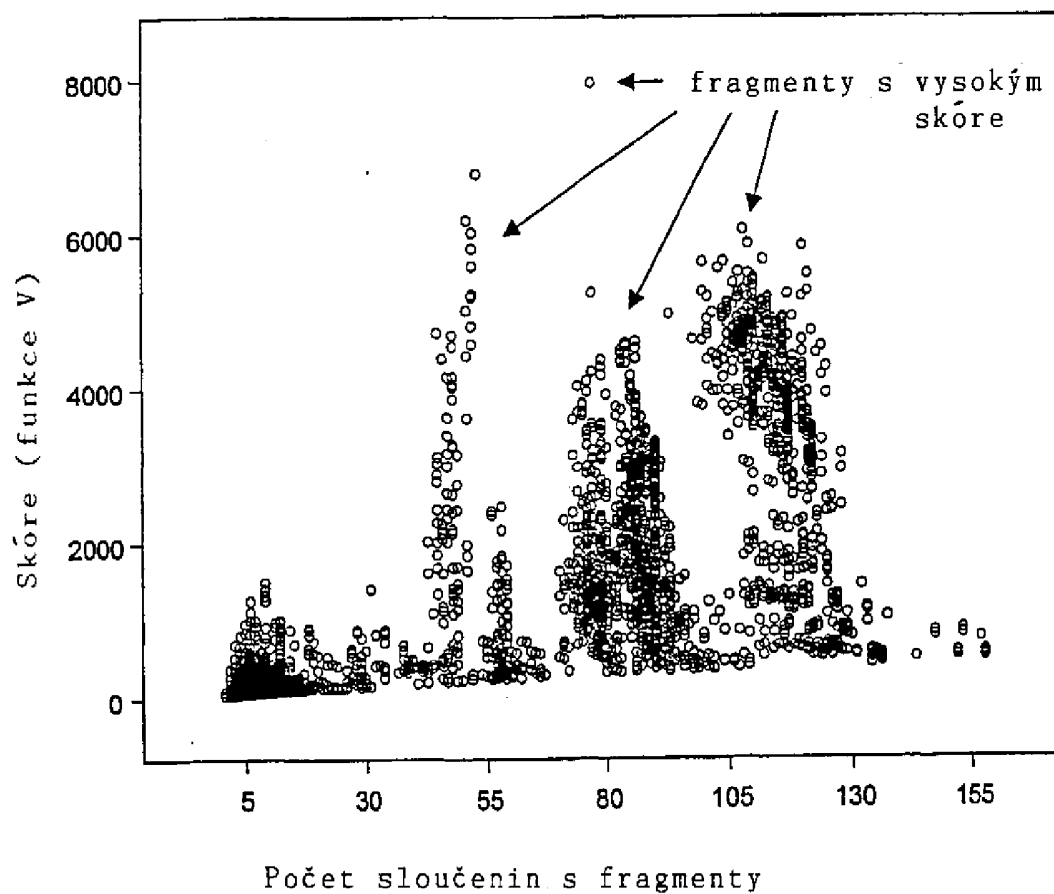
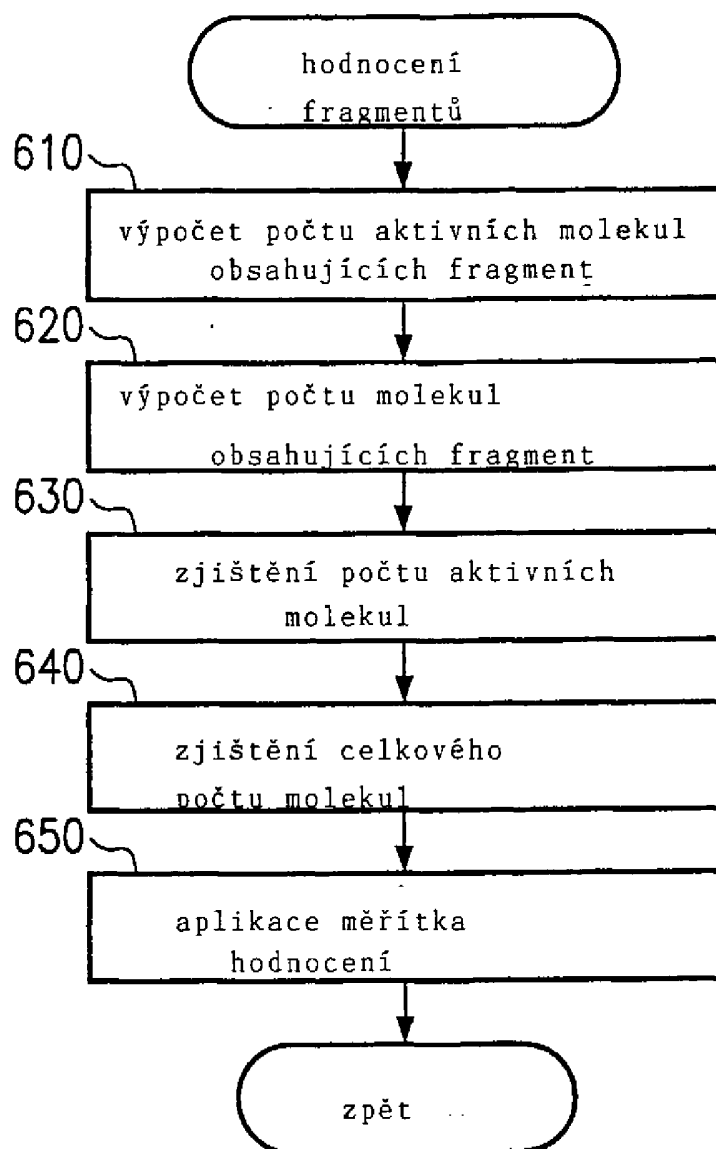
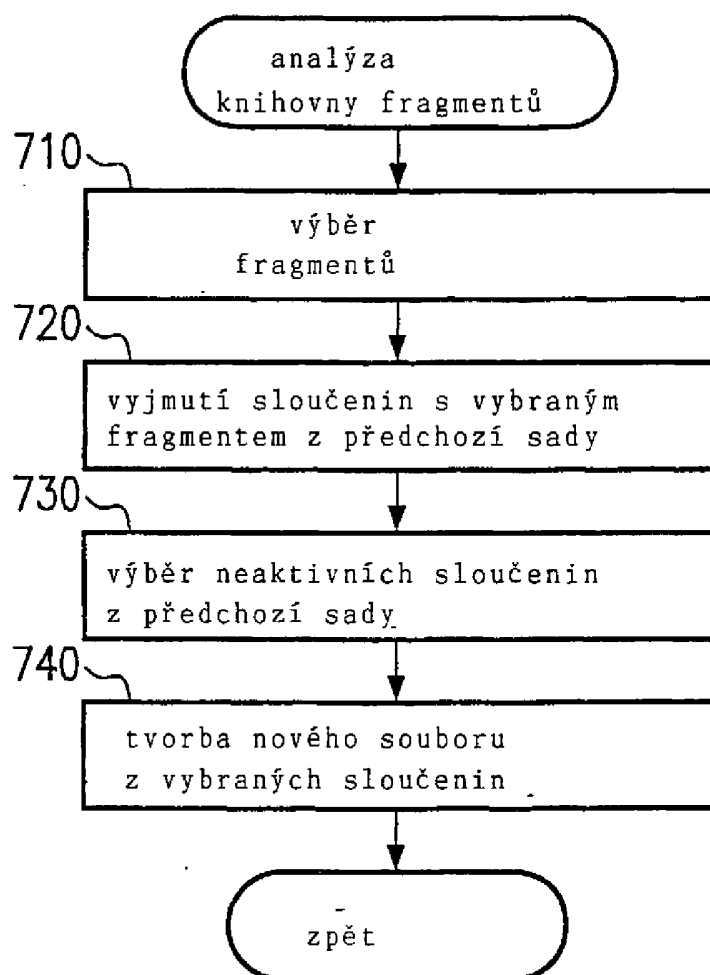


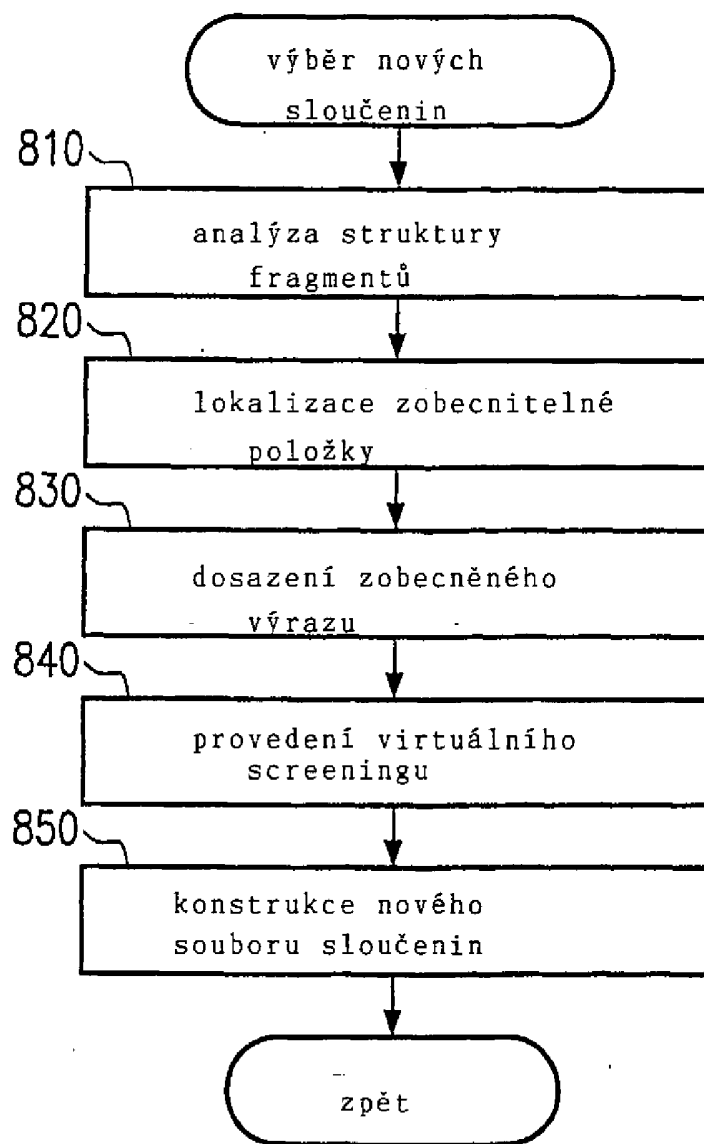
Fig. 3

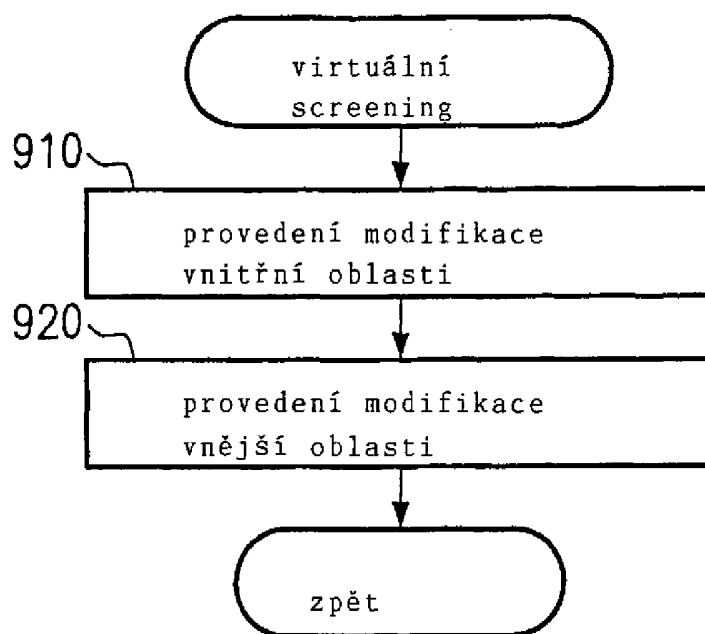
**Fig. 4**

**Fig. 5**

**Fig. 6**

*Fig. 7*

*Fig. 8*

**Fig. 9**

10/19

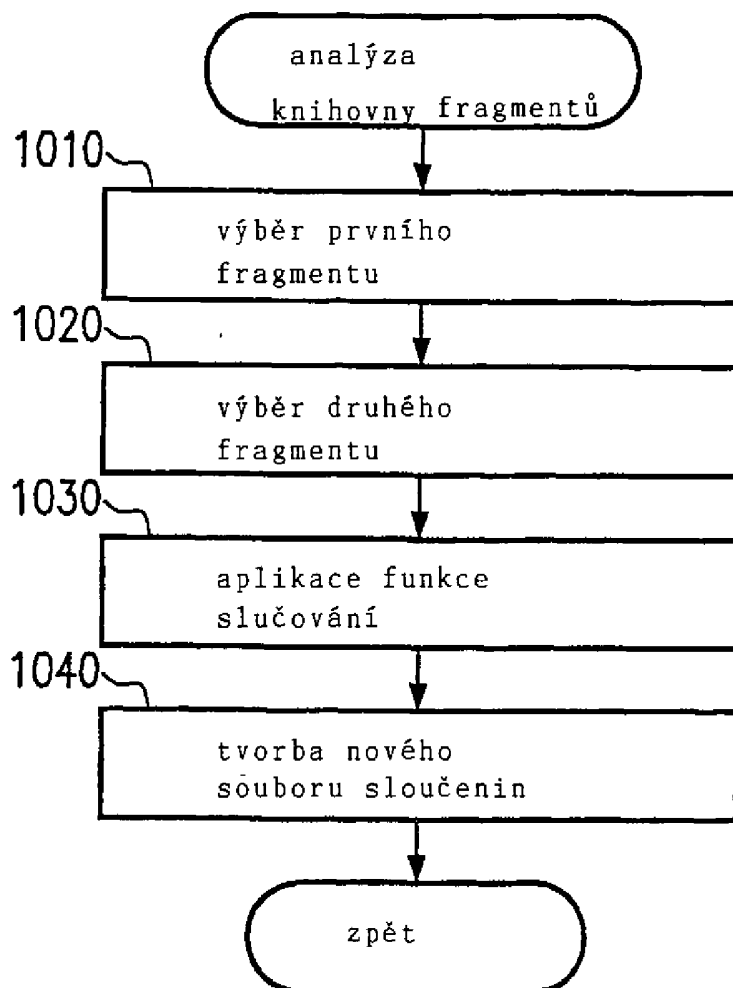
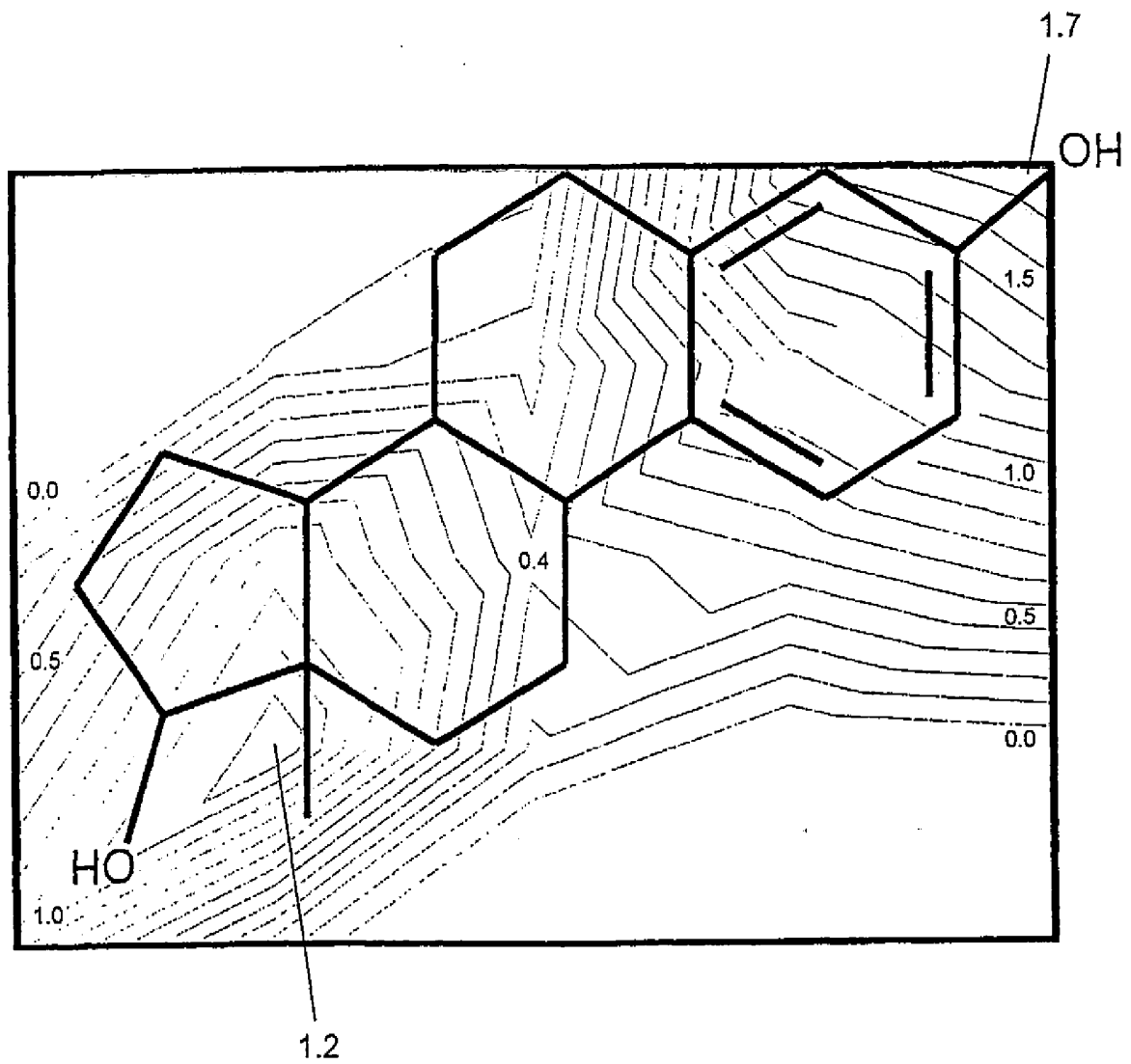
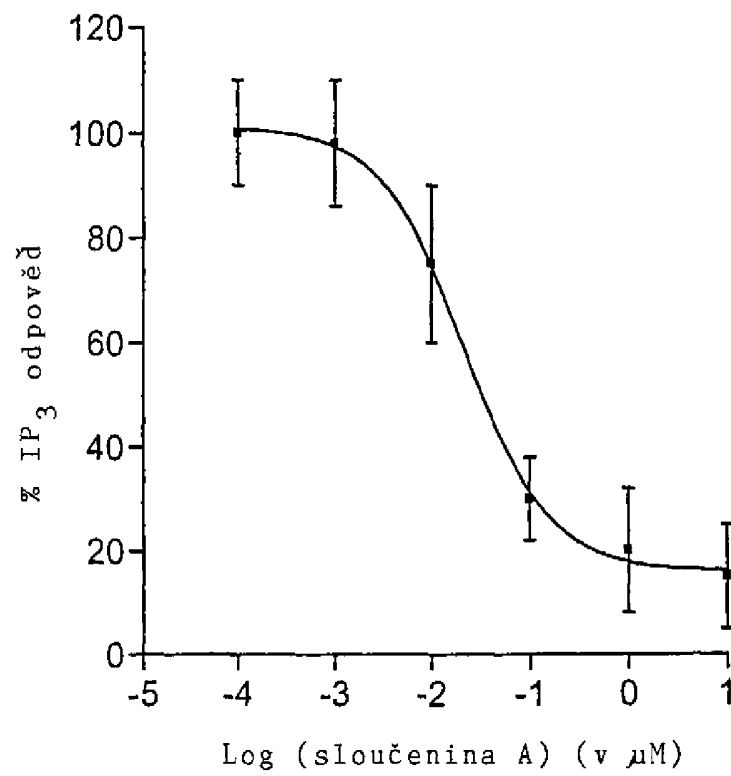
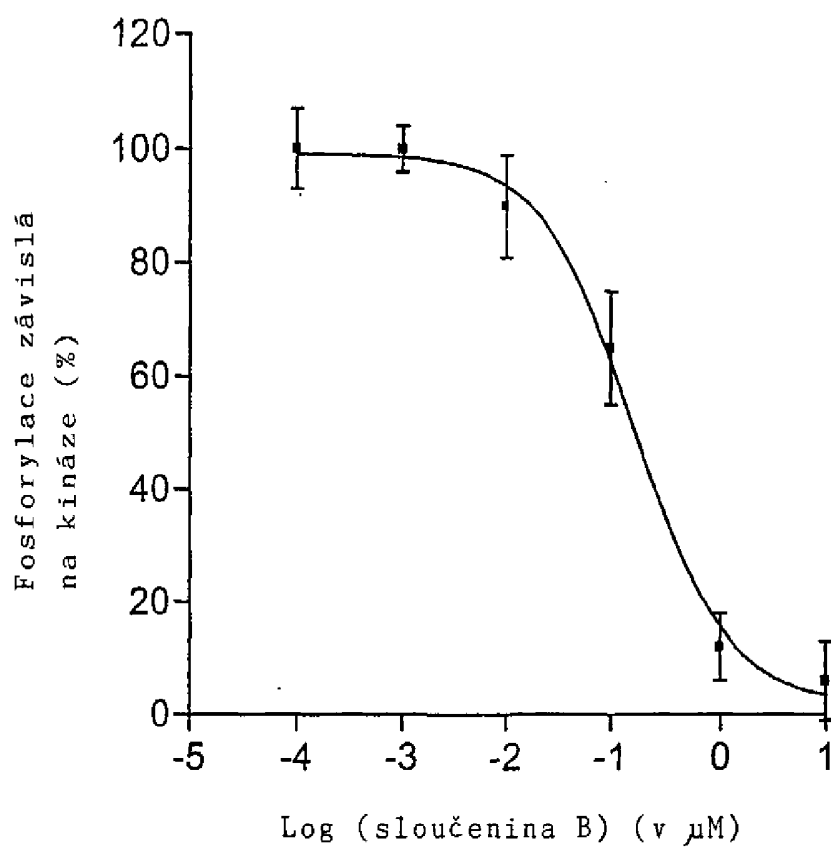


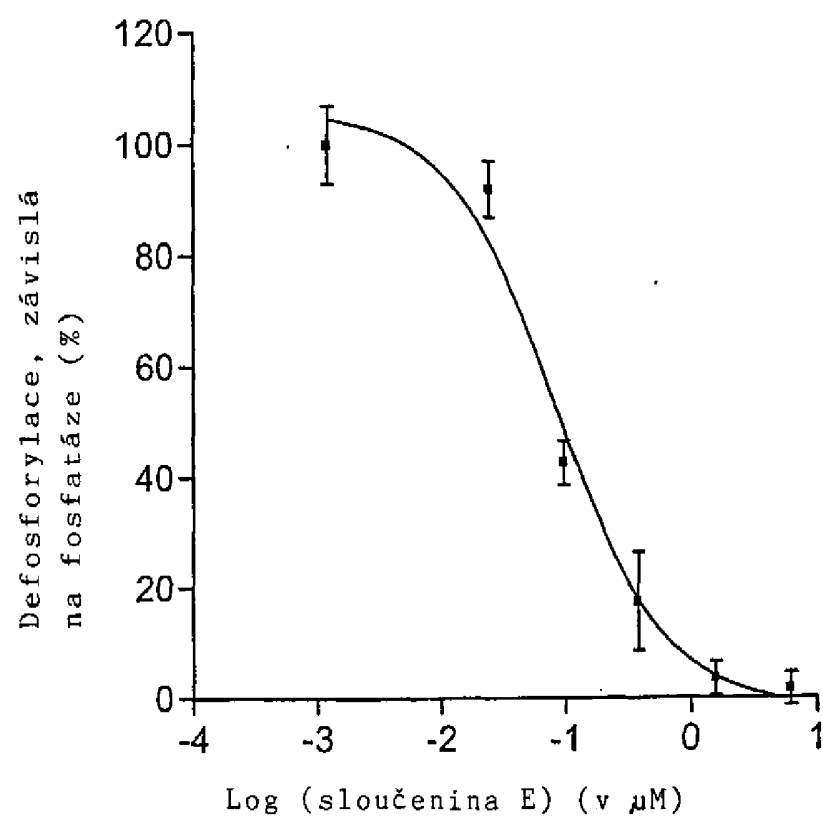
Fig. 10

*Fig. 11*

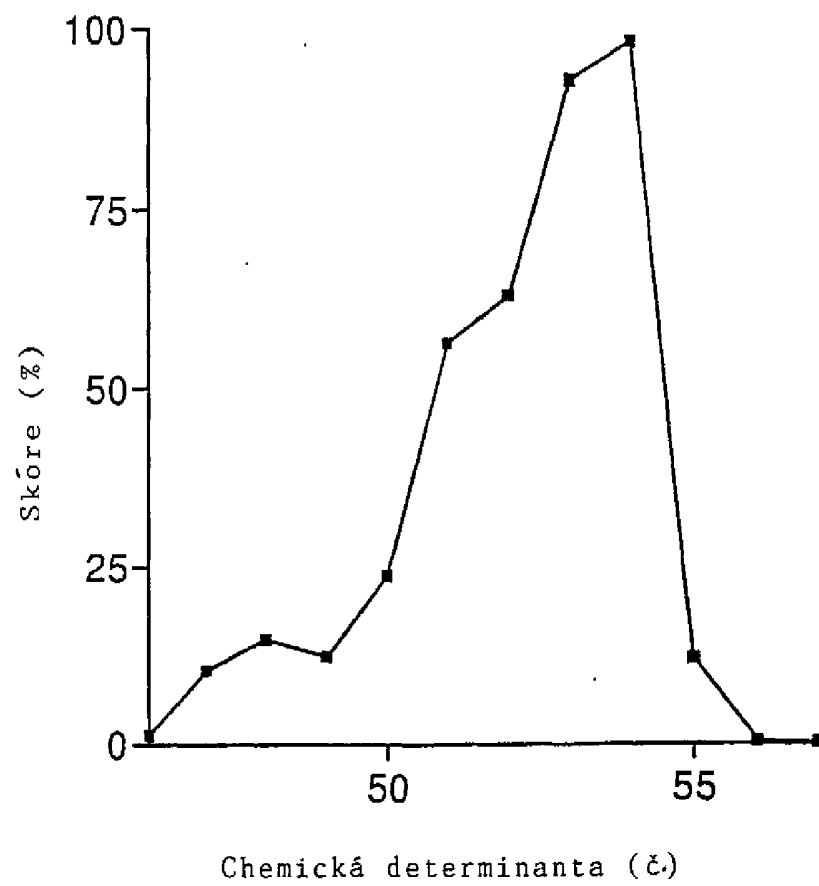
12/19

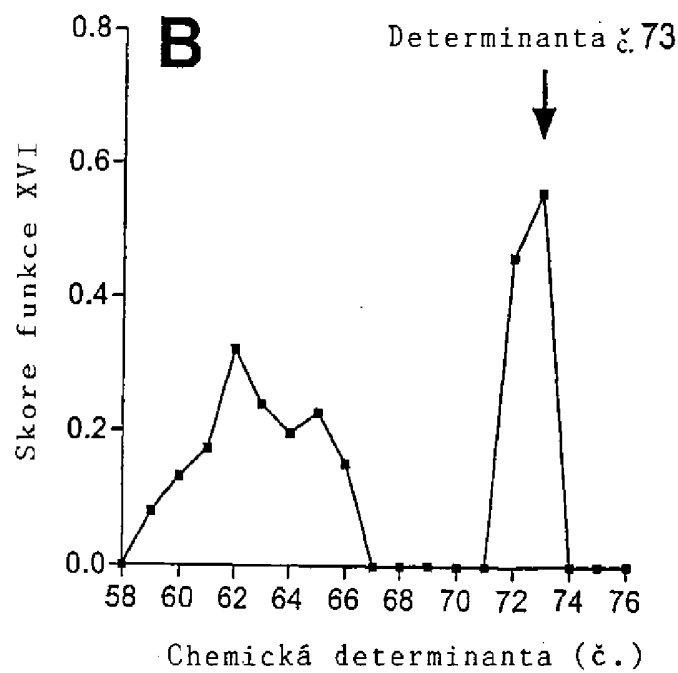
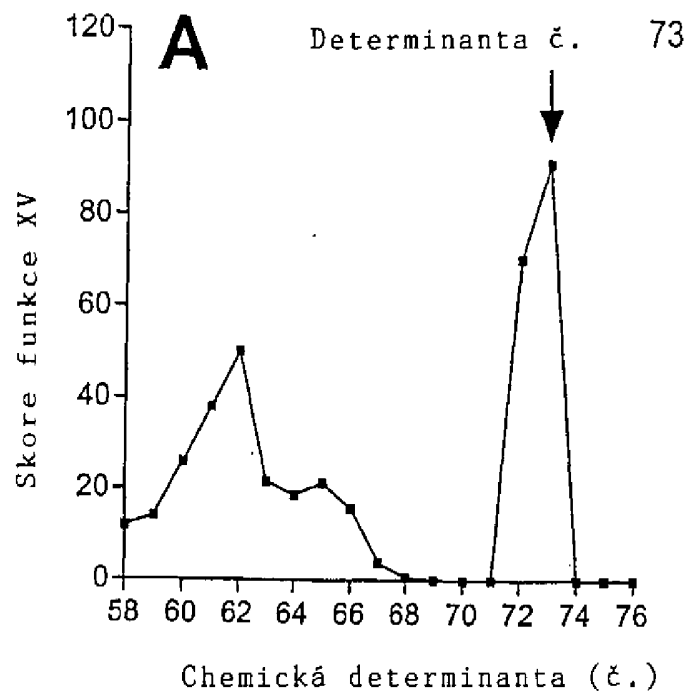
*Fig. 12*

*Fig. 13*

**Fig. 14**

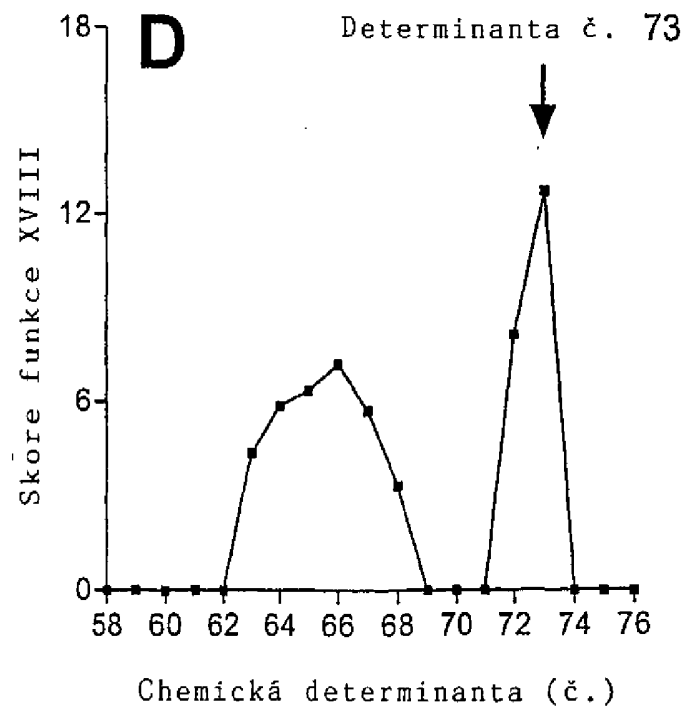
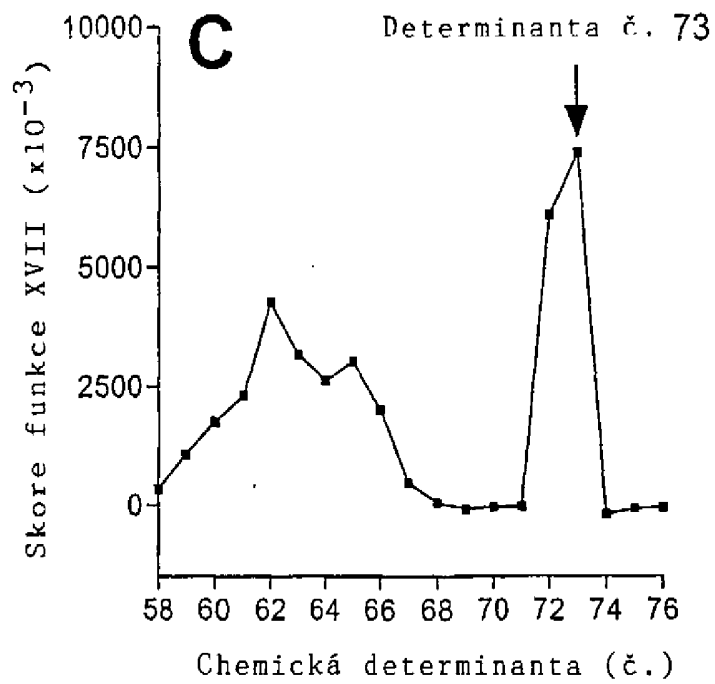
15/19

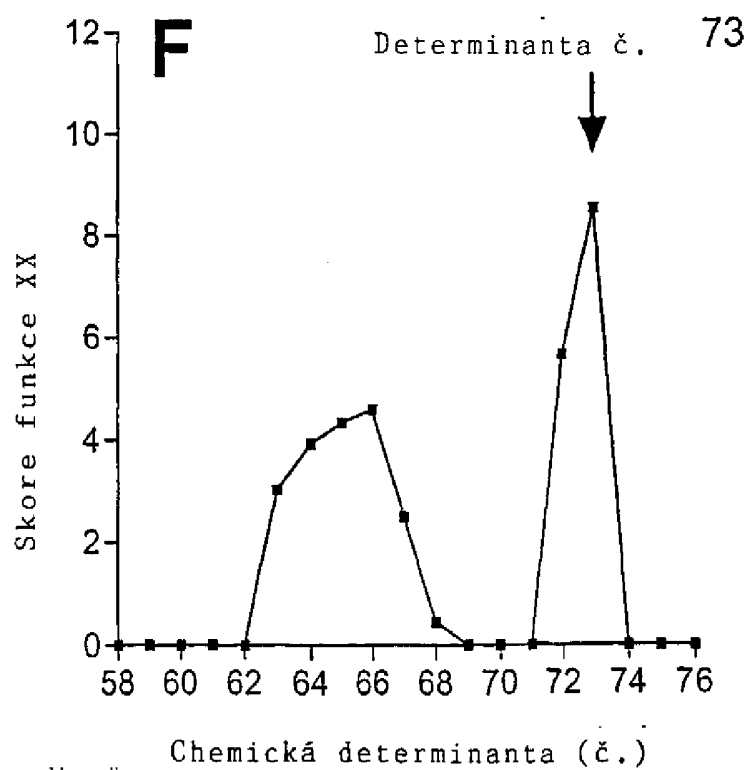
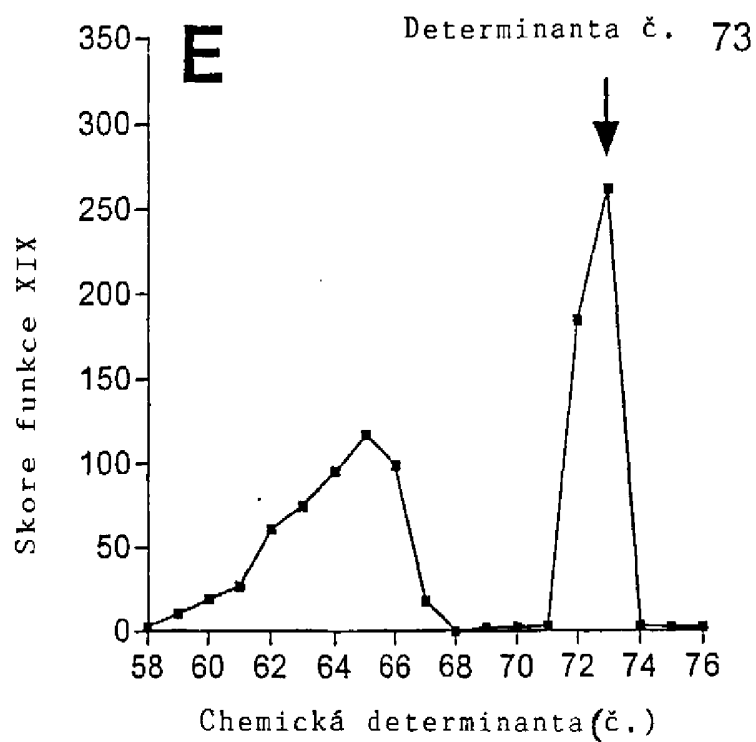
*Fig. 15*



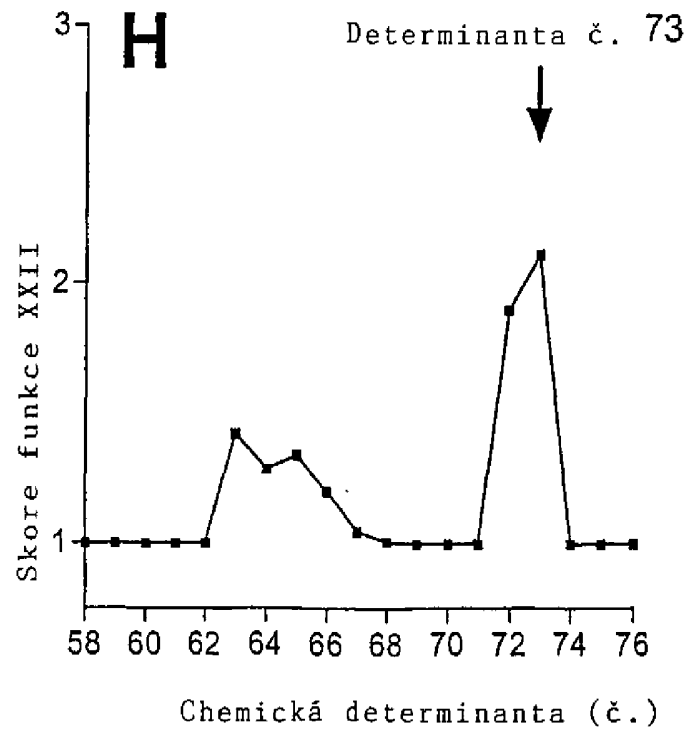
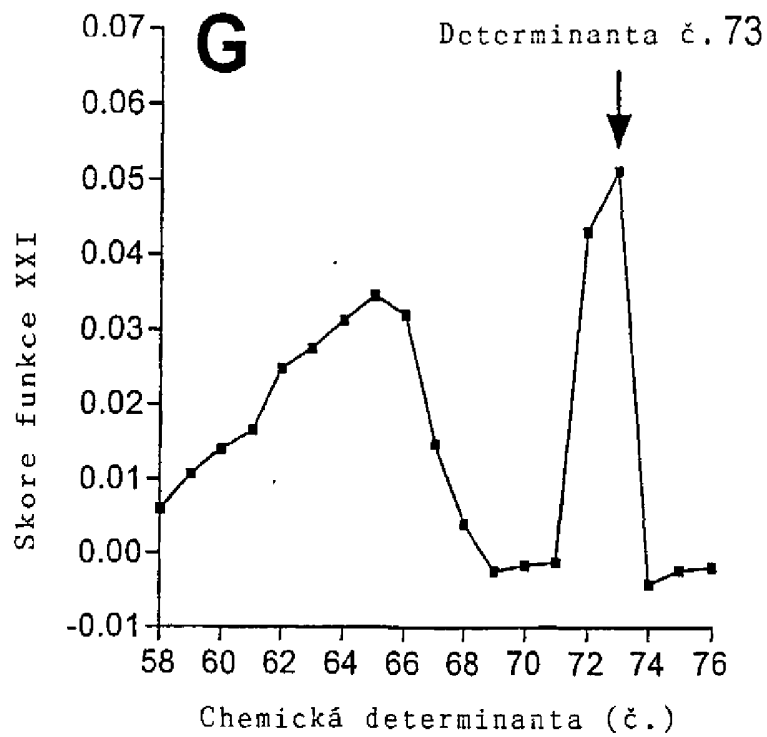
Figs. 16A, 16B

17/19

*Figs. 16C, 16D*

*Figs. 16E, 16F*

19/19

*Figs. 16G, 16H*