

© Borgis

Zastosowanie różnorodnych metod bioinformatycznych w badaniach biologicznych i medycznych

*Paweł Kowalczyk¹, Dorota Dziuban², Katarzyna Hulka³, Maciej Filocha¹

¹Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego
Dyrektor Centrum: prof. dr hab. Marek Niezgódka

²Instytut Biochemii i Biofizyki Polskiej Akademii Nauk, Zakład Biologii Molekularnej
Dyrektor Instytutu: prof. dr hab. Piotr Zielenkiewicz

Kierownik Zakładu: prof. dr hab. Barbara Tudek

³Wydział Biologii Uniwersytetu Warszawskiego

Dziekan Wydziału: prof. dr hab. Joanna Pijanowska

USING A VARIETY OF BIOINFORMATIC METHODS IN BIOLOGICAL AND MEDICAL RESEARCH

Summary

Examination of potential mutagenic and carcinogenic properties of new products entering the market or to the environment that may induce human cancers has been made compulsory for the various sectors of the economy in the various fields as medicine and molecular biology. The number of new chemical carcinogenic compounds present in the human environment with the ability of mutations in critical genes (oncogenes or tumor suppressor genes) requires the formation of ever new methods, techniques and screening tests used in the micro-organisms (bacteria, yeast, viruses, cell culture) or higher organisms (fruit fly). Studies in animals (rodents) are carried out only in the final testing phase of the compound. A very large amount of information required for pre-selection of active compounds) was frequently replaced *in silico* studies. Therefore, with the help of bioinformatics has developed a number of computational programs (information and math), explaining in a simple way to very complex biological systems and medical devices within the entire genome, because they are much more sensitive and they can study the changes in organisms exposed to low doses of mutagens or diversity within the population. The ability to process a gigantic amount of information obtained during the study of genomes and proteomes of living organisms is essential not only in scientific work, but also for practical applications as the design of therapeutic substances, crystallochemistry drugs, magnetic resonance spectroscopy in biological research, quantum chemistry and molecular modeling, instrumental analysis used in organic synthesis of classical and molecular biology.

Key words: PCR, mathematical methods, DNA repair genes

MODELOWANIE MATEMATYCZNE W SŁUŻBIE NAUK BIOLOGICZNYCH I MEDYCZNYCH

Obecnie modelowanie zachowania się układów biologicznych odgrywa ogromne znaczenie w biologii molekularnej, syntetycznej biochemii i medycynie. Wiele metod matematycznych znalazło zastosowanie do symulowania zachowania działania układów badanych przez te dyscypliny, jak: przemiany metaboliczne zachodzące w komórce, wpływ kancerogenów chemicznych na stabilność helisy DNA, projektowanie nowych leków związanych z chorobami autoregeneracyjnymi. Dostępnych jest wiele darmowych narzędzi programistycznych umożliwiających prowadzenie takich symulacji. Niestety wraz ze wzrostem złożoności układu drastycznie wzrasta złożoność obli-

zeniowa symulacji oraz stopień nieprzewidywalności zachowania się układu. Dodatkowym utrudnieniem w tworzeniu poprawnie działających modeli jest duża liczba czynników zewnętrznych, które zakłócają jego działanie a często nie są należycie uwzględniane (bądź nie są znane) podczas konstruowania takich modeli. Tworzenie matematycznych modeli opisujących projektowany układ pozwala na przewidzenie niektórych jego cech, co może posłużyć do jego udoskonalenia w celu wyeliminowania wad lub wprowadzenia pewnych usprawnień. Obecnie zostały rozwinięte wielkoskalowe modele sieci genetycznych tworzone z myślą o zastosowaniach w sekwencjonowaniu metagenomów ssaczy i bakteryjnych. Pozwalają one na odtwarzanie podstawowych procesów

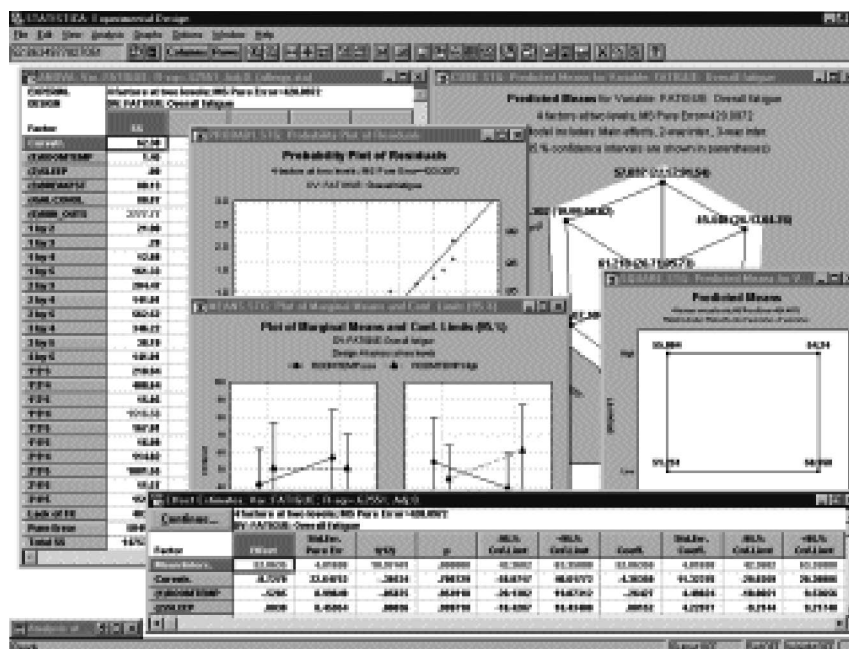
biologicznych, takich jak transkrypcja, translacja oraz wpływ czynników transkrypcyjnych oraz sygnałów środowiskowych na działanie całych sieci genetycznych. Do prowadzenia takich symulacji można wykorzystać różnorodne oprogramowanie komputerowe. Do często stosowanych programów należą: Matlab – nazwa programu pochodzi od angielskich słów *MATRIX LABORATORY*, gdyż początkowo program ten był przeznaczony do numerycznych obliczeń macierzowych. Cechuje go duża liczba funkcji bibliotecznych oraz duże możliwości rozbudowy przez użytkownika za pomocą pisania własnych funkcji. Posiada on swój język programowania, co umożliwia pisanie w pełni funkcjonalnych programów działających w środowisku Matlab. W zakresie grafiki Matlab umożliwia rysowanie dwu- i trójwymiarowych wykresów funkcji oraz wizualizację wyników obliczeń w postaci rysunków statycznych i animacji. Możliwe jest pobieranie danych pomiarowych z urządzenia zewnętrznego przez porty w celu ich obróbki. Wszystko to powoduje, że program ten znajduje bardzo szerokie zastosowanie w medycynie i biologii molekularnej. Istnieją także alternatywne odpowiedniki tego programu, takie jak Scilab, Promot, CellDesigner czy Octave – służą do prostych obliczeń przez macierze, algebrę liniową, przetwarzanie sygnałów, przy wykorzystaniu środowiska graficznego są w stanie rysować grafy i wykresy dwu- i trójwymiarowe, a nawet tworzyć animacje. Do tego typu obliczeń statystycznych służy pakiet *STATISTICA*, który zawiera bardzo dużą liczbę metod obliczeniowych do analizowania danych zebranych w eksperymentach, jak i do dopasowywania planów do ciągłych i skategoryzowanych zmiennych z wykorzystaniem kompletnej analizy wariancji typu (ANOVA). Użytkownik sprawuje pełną kontrolę nad tym, które efekty i interakcje będą włączone do analizowanego modelu; może przeglądać korelacje pomiędzy kolumnami macierzy planu (X) i zapoznać się z

odwrotnością macierzy iloczynowej $X'X$ (tzn. udostępniane są macierze korelacji i kowariancji estymatorów parametrów). W programie obliczane są estymatory parametrów ANOVA, ich odchylenia standardowe oraz przedziały ufności, a także współczynniki regresji (odchylenia standardowe, przedziały ufności) zarówno dla unormowanych (przedział [-1, +1]), jak i rzeczywistych wielkości wejściowych. Na podstawie wspomnianych estymatorów program umożliwia obliczenie wartości aproksymowanych (oraz odchyłeń standardowych i przedziałów ufności) dla zadanych przez użytkownika wartości wielkości wejściowych. Liczne opcje umożliwiają graficzne przedstawienie uzyskanych wyników: wykresy Pareto efektów, wykresy prawdopodobieństwa normalnego i normalnego połowkowego efektów, wykresy kwadratowe i sześciennie, wykresy wartości średnich i interakcji (wraz z przedziałami ufności dla średnich krańcowych), wykresy przestrzenne i warstwowe powierzchni odpowiedzi (ryc. 1).

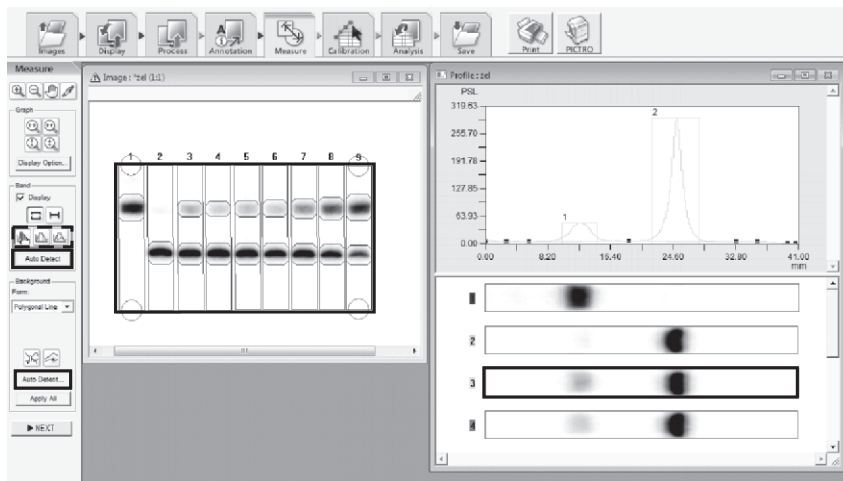
Program Multi Gauge wersja 3.0 wyprodukowany przez firmę FUJIFILM, służy do analizy wyników graficznych w postaci prążków z żeli agarozowych i akrylamidowych wymagających obróbki ilościowej. Do głównych funkcji programu należą: korekta obrazu (Display) umożliwiająca regulację kontrastu, obróbka obrazu (Process): obrót względem osi, redukcja szumu, umieszczenie adnotacji (Annotation): komentarzy i strzałek, pomiar (Measure): obrazu żeli, płytek, kalibrację (Calibration): nanoszenie linii kalibracyjnych dla stężenia i masy molekularnej, analizę wyników (Analysis). Przykładową analizę obrazu żelu przedstawiono na rycinie 2.

ALTERNATYWNE PROCEDURY ANALIZOWANIA DANYCH DOŚWIADCZALNYCH

Obecnie podejmowane są różne inicjatywy polegające na standaryzacji technik biologii molekularnej



Ryc. 1. Przykładowy rzut ekranu.

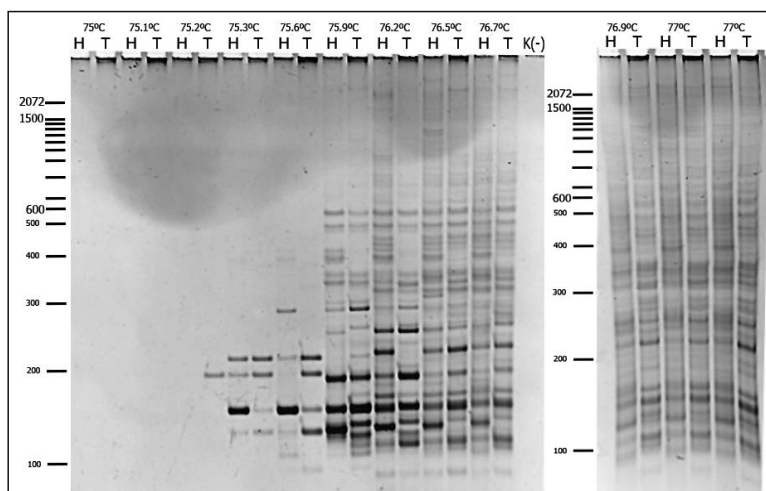


Ryc. 2. Przykładowa analiza obrazu żelu. Doświadczenie miało na celu pomiar aktywności enzymu, którą odzwierciedla procentowy udział prążka o mniejszej masie. W celu uzyskania wyraźnego sygnału i zmniejszenia tła, dopasowano kontrast i jasność obrazu. Aby otrzymać profil każdej ze studzienek przeprowadzono autodetekcję prążków i tła (linia pogrubiona), w razie potrzeby naniesiono własne poprawki (linia przerywana).

stosowanych w biologii syntetycznej. Jedną z nich jest wprowadzenie formatu Biobrick. Jego ideą jest podział sekwencji DNA ze względu na pełnioną funkcję. Przykładowo można wyróżnić sekwencje kodujące białka, promotory lub terminatory. Każdy z takich elementów jest oflankowany z obydwu stron przez ściśle określone sekwencje nazywane prefiksem i sufiksem, zawierające miejsca rozpoznawane przez charakterystyczne enzymy restrykcyjne. Dzięki temu możliwe jest łatwe tworzenie dużych fragmentów DNA z mniejszych elementów składowych. Takie „urządzenia biologiczne” mogą pełnić różnorodne role i stanowić zarazem fragmenty większych syntetycznych obwodów genetycznych lub całych sieci regulacyjnych. Przykładem tego są czynniki fizyczne i chemiczne zdolne do interakcji z DNA, tworząc różnego rodzaju modyfikacje lub addukty (1-4). Modyfikacjom mogą podlegać zasady, deoksyryboza i grupy fosforanowe. Promieniowanie jonizujące, a także stany zapalne oraz infekcje powodują powstawanie reaktywnych form tlenu (RFT) i azotu (RFA), takich jak tlen atomowy, nadtlenek wodoru, rodnik hydroksylowy, tlenek azotu i wiele innych. RFT utleniają zarówno zasady DNA, jak i cukry, przy czym proporcje modyfikacji poszczególnych składników DNA zależą od reaktywności czynnika utleniającego. Rodnik hydroksylowy należy do najbardziej reaktywnych i najbardziej krótkotrwałych RFT. Reagując z zasadami i cukrami DNA, powoduje ich utlenienie, często połączone z fragmentacją. Fragmentacja deoksyrybozy prowadzi do powstawania pęknięć nici DNA. Zasady wolnych nukleotydów i nukleozydów oraz zasady występujące w jednoniciowych fragmentach DNA są znacznie łatwiej modyfikowane (ok. 100-, 1000-krotnie) przez różnorodne czynniki mutagenne, niż wtedy gdy są chronione w podwójnej helisie DNA (5, 6). Zmodyfikowane zasady, których wielkość różni się nieznacznie od zasad niezmodyfikowanych, są usuwane z DNA na drodze naprawy przez wycinanie zasad (ang. *Base Excision Repair* – BER). Niektóre zmodyfikowane przez czynniki mutagenne

zasady DNA są niestabilne i ulegają przekształceniu we wtórne uszkodzenia o innych właściwościach biologicznych. Często są one również rozpoznawane przez inne enzymy naprawcze. Dodatkowo też wykorzystuje się metody badania niestabilności genomu oparte na identyfikacji zmian sekwencji DNA w określonych loci lub w obrębie całego genomu. Badanie mutacji w określonych loci jest podejściem klasycznym, które z jednej strony umożliwia identyfikację czynnika uszkodzającego DNA, gdyż opiera się na pomiarze specyficznych mutacji, ale z drugiej strony czułość tych metod jest ograniczona i wymaga stosowania wysokich dawek mutagenów. Do globalnych metod obliczeniowych należą, np. analiza polimorfizmu pojedynczych nukleotydów czy analiza wymian chromatyd siostrzanych lub mikrojąder. Wykorzystuje się do tego metody biologiczne opartej na technice Ligation Mediated PCR Melting Profiles (LM – PCR) (7-9). Oparta jest ona na amplifikacji w gradiencie temperatury fragmentów restrykcyjnych uzyskanych przez cięcie genomowego DNA jedną restryktazą. Po strawieniu DNA restryktazą, do końców fragmentów restrykcyjnych ligowane są adaptory, które stanowią miejsce przyłączenia starterów. Mutacje punktowe w sekwencjach rozpoznawanych przez restryktazę, a także pomiędzy nimi oraz rearanżacje chromosomalne zmieniają temperaturę topnienia fragmentów restrykcyjnych, a w konsekwencji profil prążków uzyskanych po rozdziale elektroforetycznym produktów reakcji (ryc. 3). Badaniu poddawane są profile topnienia dla tkanki jelita grubego normalnej i nowotworowej uzyskanych podczas terapeutycznych zabiegów operacyjnych. Tkanka nowotworowa charakteryzuje się nabywaniem wielu mutacji w różnych genach. Ocenia się, że mogą to być tysiące mutacji w jednej komórce (10, 11).

Metoda LM-PCR polega na monitorowaniu przyrostu liczby kopii badanej sekwencji w czasie rzeczywistym. Jest to możliwe dzięki znakowaniu produktów amplifikacji cząsteczkami zdolnymi do fluorescencji. Istnieją różne metody znakowania amplifikowanych cząsteczek: niespecyficzne,



Ryc. 3. Rozdział elektroforetyczny produktów reakcji LM-PCR uzyskanych w gradiencie temperatury dla guzów jelita grubego (T) oraz obrzeża niewykazującego zmian histologicznych (H) po cięciu genomu restryktazą Hind III.

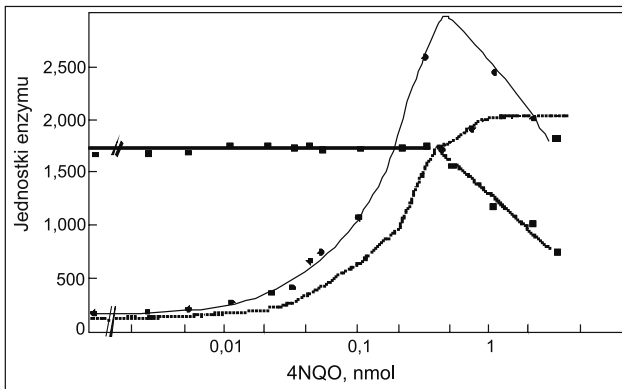
poprzez interkalację fluorochromu do dwuniciowego DNA powstającego w trakcie reakcji PCR, oraz specyficzne przy użyciu znakowanych starterów lub sond. Fluorescencja każdej próbki jest proporcjonalna do ilości zsyntetyzowanego produktu (im silniejsza fluorescencja, tym więcej kopii). Mierzac fluorescencję próbek po każdym cyklu reakcji można monitorować jej przebieg. Do obliczenia liczby badanych cząsteczek obecnych w mieszaninie na początku reakcji wykorzystuje się moment, w którym poziom fluorescencji barwnika przekroczy zdefiniowany próg (CT). Im wcześniej poziom fluorescencji wzrósł do określonego poziomu, tym więcej było kopii badanej sekwencji w próbce na początku reakcji. Najczęściej CT jest momentem wejścia reakcji w fazę logarytmicznego przyrostu ilości produktu. W celu zbadania liczby kopii badanej sekwencji w próbce sporządza się serię rozcieńczeń roztworu o znanym stężeniu i liczbie kopii DNA, wyznacza się wg nich krzywą standardową i mając dany punkt CT, na podstawie krzywej standardowej szacuje się liczbę cząsteczek. Wykorzystuje się tutaj metodę znakowania produktów PCR przy użyciu fluorochromu SYBR green. Genem referencyjnym, względem którego będą przeprowadzone obliczenia ilości produktu PCR, są 18s lub fosfomannozomutaza 1.

Specyficzność mutacji indukowanych przez różne czynniki uszkodzające DNA można badać poprzez bezpośrednie sekwencjonowanie mutantów. Ze względu jednak na fakt, że wiąże się to ze sporą ilością czasu i środków, a także niemałym nakładem pracy, badacze zajmujący się toksykologią genetyczną opracowali szereg testów bakteryjnych umożliwiających wykrycie typów mutacji indukowanych przez badane czynniki genotoksyczne. Jednym z nich jest test wykorzystujący operon laktozowy, tzw. test Millera. Każdy ze szczepów serii CC101 – CC106 zawiera inną mutację typu podstawienia zasady w genie lacZ, która zmienia kodon 461 β-galaktozydazy (Glu), prowadząc do inaktywacji enzymu. Rewersja specyficznej mutacji przywróci aktywność β-galaktozydazy i umożliwi wzrost na pożywce zawierającej laktozę jako jedyne źródło

węgla. Testy bakteryjne wykrywają zdolność związków chemicznych lub czynników fizycznych do: a) indukcji mutacji (najszerzej stosowany jest test Ames) lub b) wprowadzania uszkodzeń DNA o właściwościach letalnych (SOS chromotest). Stosując te testy, można również uwzględnić wpływ metabolizmu ssaków na aktywność mutagenną/rakotwórczą badanych związków chemicznych (aktywację metaboliczną), dodając do układu bakteryjnego frakcję mikrosomalną z wątroby lub innych organów (tzw. frakcja S9). Błony reticulum endoplazmatycznego zawierają bowiem enzymy aktywujące/detoksykujące niepolarnie prekursorów aktywnych kancerogenów, tzn. system monoooksygenaz i glukuronilotransferazę.

SOS chromotest jest testem kolorymetrycznym opierającym się na zdolności związków chemicznych lub czynników fizycznych do indukcji systemu SOS u *Escherichia coli*. Szczep testowy PQ37 zawiera fuzję genu lacZ z jednym z genów indukowanych w systemie SOS – sfiA (inhibitor podziałów komórkowych) oraz delecję w normalnym rejonie genu lacZ. Tak więc ekspresja β-galaktozydazy jest ściśle zależna od indukcji regulonu SOS (12). Badane związki chemiczne w pewnych stężeniach mogą hamować syntezę białka, co powodowałoby zniżenie wartości aktywności β-galaktozydazy. W celu wyeliminowania tego błędu równolegle z β-galaktozydazą oznacza się poziom alkalicznej fosfatazy, która w szczepie testowym wyrażana jest w sposób konstytutywny. Stosunek obydwu aktywności (β-galaktozydazy do alkalicznej fosfatazy) stanowi aktywność właściwą (R) β-galaktozydazy (ryc. 4). Przy tak podanych wynikach współczynnik indukcji wzrasta wraz ze wzrostem stężenia i osiąga plateau, nawet przy dawkach obniżających syntezę białka.

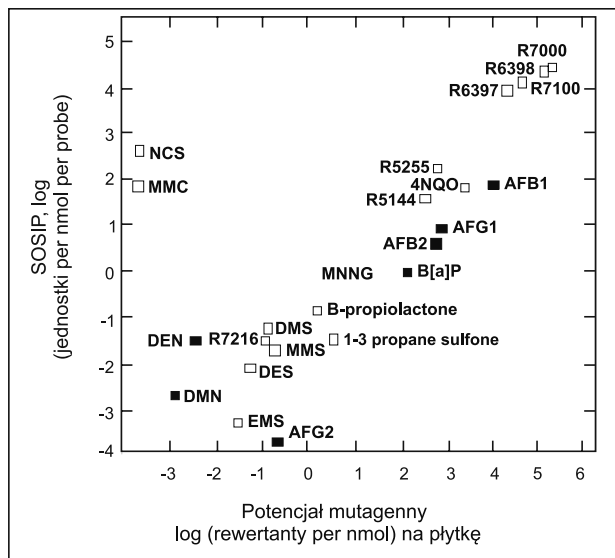
Porównanie wyników otrzymanych w SOS chromotestie (13, 14) z wynikami uzyskanymi w testach na mutagenność, np. w teście Ames wskazuje, że dla znakomitej większości związków chemicznych istnieje bardzo dobra ilościowa korelacja pomiędzy zdolnością do indukcji systemu SOS, indukcji mutacji oraz indukcji nowotworów (ryc. 5). □



Ryc. 4. SOS chromatostest dla różnych stężeń tlenku 4-nitrochinoliny (4NQO): aktywność β -galaktozydazy (\bullet - \bullet), aktywność alkalicznej fosfatazy (\blacksquare); aktywność właściwa β -

Piśmiennictwo

1. Tornaletti S, Pfeifer GP: Slow repair of pyrimidine dimers at p53 mutation hotspots in skin cancer. *Science* 11 Mar 1994; 263 (5152): 1436-1438. 2. Pfeifer GP, Tornaletti S: Footprinting with UV irradiation and LMPCR. *Methods* Feb 1997; 11 (2): 189-196. 3. Pfeifer GP, Denissenko MF, Tang MS: PCR-based approaches to adduct analysis. *Toxicol Lett* 28 Dec 1998; 102-103: 447-451. 4. Pfeifer GP, Dammann R: Measuring the formation and repair of UV photoproducts by ligation-mediated PCR. *Methods Mol Biol* 1999; 113: 213-226. 5. Smith LE, Denissenko MF, Bennett WP et al.: Targeting of lung cancer mutational hotspots by polycyclic aromatic hydrocarbons. *J Natl Cancer Inst* 2000; 92: 803-811. 6. Lau AY, Scharer OD, Samson L et al.: Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: Mechanisms for Nucleotide flipping and base excision. *Cell* 16 Oct 1998; 95: 249-258. 7. Stivers JT, Pankiewicz KW, Watanabe KA: Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. *Biochemistry* 1999; 38: 952-963. 8. Hollis T, Ichikawa Y, Ellenberger T: DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, *Escherichia coli* AlkA. *EMBO J* 2000; 19: 758-766. 9. Dodson ML,



Ryc. 5. Korelacja pomiędzy indukcją systemu SOS a indukcją mutacji w testach bakteryjnych dla znanych mutagenów/karcinogenów.

Michaels ML, Lloyd RS: Unified catalytic mechanism for DNA glycosylases. *J Biol Chem* 1004; 269: 32709-32712. 10. Krokan HE, Standal R, Slupphaug G: DNA glycosylases in the base excision repair of DNA. *Biochem J* 1997; 325: 1-16. 11. Mitra S: DNA glycosylases: specificity and mechanisms. *Prog Nucleic Acid Res Mol Biol* 2001; 68: 189-192. 12. Scharer OD, Jiricny J: Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *Bioessays* 2001; 23: 270-281. 13. Matsumoto Y, Kim K, Bogenhagen DF: Proliferating cell nuclear antigen-dependent abasic site repair in *Xenopus laevis* oocytes: an alternative pathway of base excision DNA repair. *Mol Cell Biol* 1994; 14: 6187-6197. 14. Frossina G, Fortini P, Rossi O et al.: Two pathways for base excision repair in mammalian cells. *J Biol Chem* 1996; 271: 9573-9578. 15. Mohn GR: Bacterial systems for carcinogenicity testing. *Mutation Research* 1981; 87: 191-210.

nadesłano: 31.07.2012
zaakceptowano do druku: 16.08.2012

Adres do korespondencji:
*Paweł Kowalczyk
Interdyscyplinarne Centrum Modelowania
Matematycznego i Komputerowego UM
ul. Pawińskiego 5a, 02-106 Warszawa
tel.: +48 728-864-717
e-mail: pawelk@ibb.waw.pl