

Bożena MAŁYSIAK-MROZEK, Dariusz MROZEK, Łukasz KOŁKOWSKI  
Politechnika Śląska, Instytut Informatyki

## EFEKTYWNA REPREZENTACJA MOLEKULARNYCH STRUKTUR BIAŁKOWYCH STOSOWANA W PROCESIE ICH PORÓWNANIA

**Streszczenie.** Porównanie molekularnych struktur białkowych często jest ważnym procesem towarzyszącym poszukiwaniu podobieństwa strukturalnego białek, identyfikacji ich funkcji, a także badaniu ewolucji organizmów żywych. Efektywna reprezentacja struktur białkowych jest niezwykle istotna dla powodzenia procesu porównania oraz szybkości jego prowadzenia. W niniejszym artykule przedstawiono rozważania na temat wyboru cech reprezentatywnych opisujących struktury białkowe w procesie ich porównywania. Zaprezentowano również badania dotyczące porównania struktur białkowych za pomocą macierzy odległości międzyrezydualnych, transformowanych następnie do macierzy odcieni szarości i opisanych przez współczynnik jakości obrazu Q dla szybszego porównania i wyszukiwania w bazie danych.

**Słowa kluczowe:** bioinformatyka, białka, struktura, podobieństwo, porównanie struktur

## EFFECTIVE REPRESENTATION OF PROTEIN MOLECULAR STRUCTURES APPLIED IN THEIR COMPARISON

**Summary.** Comparison of protein, molecular structures is often an essential component process of protein structure similarity searching, identification of protein functions, and investigation of the evolution of living organisms. Effective representation of protein structures in the comparison process is then very important for its successfulness and swiftness. In the paper, we present considerations on using various, representative features describing protein structures in their comparison. We also show our research on protein structure comparison with the use of intra-residual distance matrices, which are transformed to the grayscale images and described by means of the Universal Image Quality Index for faster comparison and database retrieval.

**Keywords:** structural bioinformatics, protein structure, similarity, structure comparison

## 1. Wprowadzenie

Istnienie, właściwa konstrukcja i aktywność cząstek białkowych leżą u podstaw funkcjonowania wszystkich organizmów żywych. Białka są niezwykle ważnymi cząsteczkami, pełniącymi kluczową rolę we wszystkich reakcjach biochemicznych zachodzących w komórkach organizmów [1]. Nie dziwi zatem fakt, że cząsteczki te stały się obiektami badań naukowych prowadzonych przez naukowców wielu dziedzin, takich jak: biochemia, biologia molekularna, fizyka, bioinformatyka i in.

W badaniach prowadzonych przez różne zespoły naukowe białka bardzo często porównuje się ze sobą, aby znaleźć odpowiedź na rozmaite pytania, w tym m.in.:

- Czy otrzymana w eksperymencie naukowym cząstka białkowa ma zbliżoną budowę do znanych już cząstek?
- Jaką potencjalnie funkcję może pełnić wyizolowane w badanym organizmie białko?
- Jak obecna struktura białka świadczy o jego aktywności komórkowej?
- Czy wytworzona syntetycznie struktura białka może pełnić w organizmie zamierzoną rolę?
- Jak zmiana genetyczna przekłada się na strukturę białka i jakie są tego konsekwencje biologiczne?
- Jaką informację badana struktura białkowa może wnieść w odniesieniu do klasyfikacji ewolucyjnej badanego organizmu?

Porównanie białek może być prowadzone na różnych poziomach opisu czy organizacji ich struktury, począwszy od struktury pierwszorzędowej aż po strukturę czwartorzędową. Każdy z tych poziomów dostarcza innego rodzaju informacji, a sama budowa białka może być różnie opisana lub reprezentowana w systemach komputerowych. W zależności od celu prowadzenia porównania naukowców będzie interesował opis bardziej lub mniej szczegółowy. W każdym jednak przypadku analizie i porównaniu będzie podlegała wewnętrzna budowa cząsteczek białkowych. Efektywna reprezentacja struktur białkowych jest natomiast niezwykle istotna dla powodzenia procesu porównania oraz szybkości jego prowadzenia.

W niniejszym artykule, przyglądając się czterem poziomom organizacji struktur białkowych, przedstawiono rozważania na temat wyboru cech reprezentatywnych opisujących te struktury w procesie porównania. Zaprezentowano także ogólnie akceptowane sposoby reprezentacji deskryptorów struktur białkowych w systemach komputerowych. Ponadto autorzy zaprezentowali własne badania dotyczące porównania struktur białkowych zgodnie z przyjętym przez siebie deskryptorem w postaci macierzy odległości.

## 2. Cztery poziomy opisu budowy białek i ich reprezentacja w systemach komputerowych

Analizując ogólną budowę białek, można powiedzieć, że są to makrocząsteczki o masie cząsteczkowej powyżej 10 kDa ( $1 \text{ Da} = 1,66 \times 10^{-24} \text{ g}$ ), zbudowane z aminokwasów (>100 aminokwasów) połączonych w łańcuchy wiązaniami peptydowymi [2]. W budowie białek wyróżnia się cztery poziomy opisu lub reprezentacji, niezwykle istotne z punktu widzenia prowadzonej analizy – są to: struktura pierwszo-, drugo-, trzecio- i czwartorzędowa [3], [4]. Trzy ostatnie poziomy definiują tzw. konformację białka lub jego strukturę przestrzenną [2]. Porównanie białek prowadzi się zazwyczaj na jednym lub kilku wybranych poziomach – można np. prowadzić porównanie wyłącznie na poziomie struktury pierwszorzędowej wówczas, gdy staramy się wykryć i przeanalizować mutacje powstałe w sekwencji białka na skutek zajścia mutacji w genie kodującym to białko.

### 2.1. Struktura pierwszorzędowa

Strukturę pierwszorzędową białek określa tzw. sekwencja aminokwasów białka, czyli kolejność aminokwasów w łańcuchu białkowym, stąd też często zamiast pojęcia struktura pierwszorzędowa używa się właśnie pojęcia sekwencja. W porównaniu struktur białkowych lub klasyfikacji białek, np. prowadzonych za pomocą systemu CATH [5] lub SCOP [6], struktura pierwszorzędowa często spełnia rolę posilkową. Przyjęto bowiem założenie, że prawdopodobieństwo, iż dwa białka o bardzo zbliżonej sekwencji będą miały zbliżoną strukturę przestrzenną, jest bardzo wysokie. Sekwencja aminokwasów może być różnie reprezentowana w procesie przetwarzania i analizy danych biologicznych. Niemniej jednak najczęściej stosowanym sposobem jest reprezentacja w formacie FASTA [7]. Przykład sekwencji białek mioglobiny i hemoglobiny w formacie FASTA przedstawiono na rys. 1. Pierwsza linia zbioru dla każdej cząsteczki ma charakter opisowy, informacyjny, natomiast począwszy od drugiego wiersza, poszczególne aminokwasy w liniowym łańcuchu są reprezentowane przez litery alfabetu (np. V to aminokwas waliny, L – lizyny, S – seryny itd.).

```
>1MBN:A|PDBID|CHAIN|SEQUENCE
VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASEDLKKHGVTVLTALGA
ILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQQGAMNKALELFRKDIAAKYKE
LGYQG

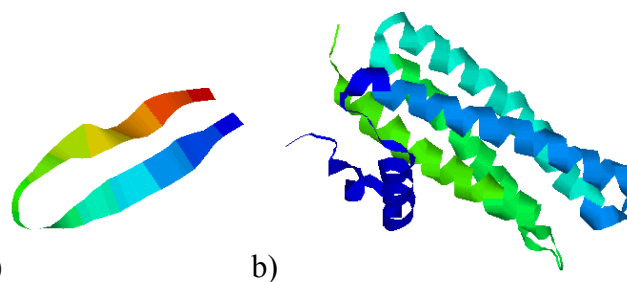
>4HHB:A|PDBID|CHAIN|SEQUENCE
VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVND
DMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR
```

Rys. 1. Sekwencje białek: mioglobiny (PDB ID: 1MBN) i hemoglobiny (PDB ID: 4HHB, łańcuch A) w formacie FASTA

Fig. 1. Sequences of proteins: myoglobin (PDB ID: 1MBN) and hemoglobin (PDB ID: 4HHB, chain A) in the FASTA format

## 2.2. Struktura drugorzędowa

Struktura drugorzędowa białek opisuje wzajemne przestrzenne ułożenie reszt aminokwasowych, sąsiadujących ze sobą w sekwencji liniowej. Ten poziom opisu w strukturze przestrzennej wyróżnia pewne charakterystyczne, regularnie powtarzane regiony, np.: helisy  $\alpha$ , harmonijki  $\beta$  ( $\beta$ -kartki), pętle (zakręty). Struktury drugorzędowe białek są cennym źródłem informacji o budowie cząstek białkowych. Pozwalają one ocenić ogólny kształt białka, ogólny sposób uformowania łańcucha aminokwasów, a także wskazać rodzaje lub elementy struktur drugorzędowych (SSE) występujących w jego budowie – czy są to struktury tylko jednego rodzaju, np. tylko  $\alpha$ -helisy lub tylko  $\beta$ -kartki, czy struktury te są rodzajowo zróżnicowane. Ten sposób reprezentacji umożliwi również stwierdzenie, czy struktury danego typu są mocno posegregowane, czy występują naprzemiennie. Obserwacja trójwymiarowej budowy białka, reprezentowanej na poziomie struktur drugorzędowych, ujawnia wzajemną organizację przestrzenną poszczególnych fragmentów białka i dostarcza informacji dotyczących formowania różnego rodzaju motywów strukturalnych (zwanymi inaczej superstrukturami drugorzędowymi), np. popularna  $\beta$ -spinka (ang.  *$\beta$ -hairpin*) składa się z dwóch antyrównoległych  $\beta$ -nici (ang.  *$\beta$ -strand*) połączonych krótką pętlą (ang. *loop*), równie często występującym motywem jest np. motyw  $\beta$ - $\alpha$ - $\beta$ . Obserwacja białka w postaci struktury drugorzędowej pozwala również określić miejsca występowania domen funkcyjnych, czyli stabilnych konstrukcyjnie fragmentów białka, które mogą fałdować się niezależnie od siebie i pełnią zwykle określoną rolę w procesach komórkowych [8]. Na rys. 2a przedstawiono fragment struktury formujący charakterystyczną  $\beta$ -spinkę, na rys. 2b można natomiast obserwować struktury typu  $\alpha$ -helisa w przestrzennej budowie przykładowego białka z bazy Protein Data Bank (PDB) [9].



Rys. 2. Przykłady struktur przestrzennych białek reprezentowanych na poziomie struktury drugorzędowej: a) motyw  $\beta$ -spinka, b) struktury  $\alpha$ -helikalne w białku PDB ID: 1X91

Fig. 2. Examples of spatial structures represented by secondary structures: a)  $\beta$ -hairpin motif, b)  $\alpha$ -helices in the protein PDB ID: 1X91 (crystal structure of mutant form A of a pectin methylesterase inhibitor from *Arabidopsis*)

Struktury drugorzędowe, podobnie jak sekwencje aminokwasów, stanowią podstawę klasyfikacji białek i używa się ich w procesie porównywania budowy białek. Przyjmuje się tutaj biologicznie uzasadnione założenie, że białka różniące się na poziomie sekwencji aminokwa-

sów mogą pozostać podobne w sensie struktury przestrzennej, w tym struktury drugorzędowej. Struktury przestrzenne są bowiem bardziej konserwatywne. Na etapie ewolucji organizmów, na skutek zmian zachodzących w łańcuchach DNA sekwencje białek uległy zróżnicowaniu, jednakże nie spowodowało to wielokrotnie tak dużych zmian w strukturach przestrzennych białek.

Jednak z punktu widzenia prowadzonej analizy obliczeniowej często istotne jest odpowiednie reprezentowanie struktury drugorzędowej i jej właściwe przechowywanie np. w bazie danych. Dobrym sposobem opisu budowy przestrzennej białek przez struktury drugorzędowe jest reprezentowanie jej przez sekwencje elementów struktury drugorzędowej (SSE). Każdy element struktury drugorzędowej odpowiada jednemu elementowi struktury pierwszorzędowej, a zatem pojedynczemu aminokwasowi. Na rys. 3 zaprezentowano sekwencję aminokwasów białka o nazwie *6-phosphogluconolactonase* w organizmie *Escherichia coli* oraz odpowiadającą jej sekwencję elementów struktury drugorzędowej. Poszczególne symbole w sekwencji elementów struktury drugorzędowej mają następujące znaczenie: H odpowiada  $\alpha$ -helisie, E odpowiada  $\beta$ -kartce, a C odpowiada pętli (dla pętli stosuje się również oznaczenie L). Taki sposób reprezentacji został wykorzystany przez autorów w implementacji języka PSS-SQL [10, 11].

```
A7ZY23
6PGL_ECOHS
6-phosphogluconolactonase OS=Escherichia coli O9:H4 (strain HS) GN=pgl PE=3 SV=1

MKQTVYIASPESQQIHVWNLNHEGALTLTQVVDVPGQVQPMVVS PDKRYLYVGV RPEFRVLAYRIAPDDGALTF AAESAL
PGSPTHISTDHQQGFV FVGSYNAGNVSVTRLEDGLPVGVVDVVEGLDGCHSANI SPDNRTLWVPAL KQDRICLFTVSDDG
HLVAQDPAEVTTV EAGPRHMFHPNEQYAYCVNELNSSVDVWELKDPHGNI ECVTLDMM PENFSDTRWAADIHITPDG
RHLIACDR TASLITVF

CCCCCCCCCEEEEEEECCCCCEEEEEEEEEEEEECCCCCEEEEECCCCCEEEEECCCCCEEEEEEEEECCCCCHHHHHHHHCC
CCCCCEEEEECCCCCEEEEECCCCCEEEEEEEEECCCCCEEEEEEEEECCCCCCCCCCCCCEEEEECHHHHHHEEEEECCCCC
CEEECCCCCEEEEECCCCCEEEEECCCCCEEEEEEEEECCCCCEEEEEEEEECCCCCEEEEECCCCCCCCCCCCCEEEEECCCCC
CEEECCCCCCCCCEEE
```

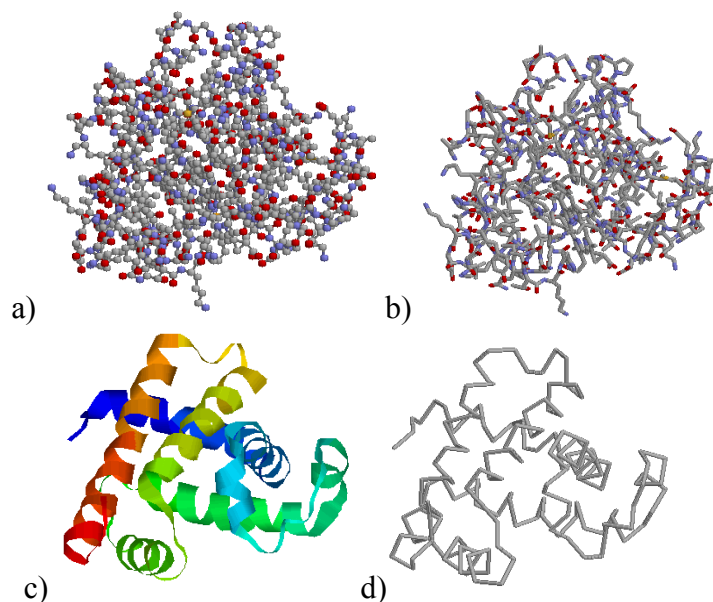
Rys. 3. Przykładowa sekwencja aminokwasów białka *6-phosphogluconolactonase* w organizmie *Escherichia coli* oraz odpowiadająca jej sekwencja elementów struktury drugorzędowej, wygenerowana za pomocą programu Predator [12]

Fig. 3. Sample amino acid sequence of the protein *6-phosphogluconolactonase* in the *Escherichia coli* with the corresponding sequence of secondary structure elements generated by means of the Predator program [12]

### 2.3. Struktura trzeciorzędowa

Struktura trzeciorzędowa odnosi się do powiązań przestrzennych i wzajemnego ułożenia reszt aminokwasowych zarówno tych oddalonych od siebie w sekwencji liniowej, jak i sąsiadujących ze sobą (rys. 4). Opisuje zatem ukształtowanie struktury spowodowane dodatkowymi, wewnętrznymi oddziaływaniami elektrostatycznymi, wodorowymi oraz ewentualnymi

kowalencyjnymi mostkami dwusiarczkowymi<sup>1</sup>. Ten poziom opisu oddaje biologicznie aktywną przestrzenną konformację białka [2]. Biorąc pod uwagę, że białka zbudowane są z setek aminokwasów, a zatem tysięcy atomów, strukturę trzeciorzędową białek można wizualnie prezentować na różne sposoby (rys. 4).



Rys. 4. Struktury trzeciorzędowe mioglobiny (PDB ID: 1MBN): a) reprezentacja atomowa, b) reprezentacja wiązań kowalencyjnych między atomami, c) reprezentacja wstążkowa, d) reprezentacja szkieletowa (wizualizacja RasMol [13])

Fig. 4. Tertiary structures of myoglobin (PDB ID: 1MBN): a) atomic representation, b) sticks representation showing covalent bonds between atoms, c) ribbon representation revealing secondary structures, d) backbone representation (visualized by RasMol [13])

Na rys. 4 przedstawiono cztery sposoby wizualnej reprezentacji struktury: atomową (rys. 4a), ukazującą wiązania międzyatomowe (rys. 4b), wstążkową, ujawniającą elementy struktury drugorzędowej (rys. 4c) i szkieletową, ukazującą szkielet białka przez połączenie charakterystycznych węgli  $C_{\alpha}$  występujących w każdym aminokwasie (rys. 4d).

Tak duża liczba atomów w strukturze przestrzennej białek sprawia, że ich analiza i wzajemne porównywanie stają się zadaniami niezwykle skomplikowanymi. Szczególnie jeśli odbywa się to przez porównanie zadanego białka do całej bazy danych struktur przestrzennych białek, np. bazy Protein Data Bank. W procesie poszukiwania podobieństwa białek, np. algorytmem DALI [14] lub CTSS [15], bardzo często porównuje się zatem położenia atomów węgla  $C_{\alpha}$  i ich wzajemne odległości w poszczególnych aminokwasach. Strukturę trzeciorzędową białka można zatem zredukować do postaci przedstawionej na rys. 4d, tym samym redukując przestrzeń poszukiwań. Czasami brane są również pod uwagę położenia

<sup>1</sup> Mostki dwusiarczkowe są wiązaniami kowalencyjnymi wytworzonymi przez grupy sulfhidrylowe (-SH), pełnią funkcję stabilizacyjną dla konstrukcji białka [4].

węgla  $C_{\beta}$ , co dodatkowo pozwala uwzględnić w porównaniu orientację łańcuchów bocznych poszczególnych reszt aminokwasowych [16].

Z punktu widzenia sposobu reprezentacji struktur trzeciorzędowych powszechnie akceptowanym formatem jest format PDB [17] zaprezentowany na rys. 5, a także jego pochodne mmCIF [18] oraz PDBML [19].

```

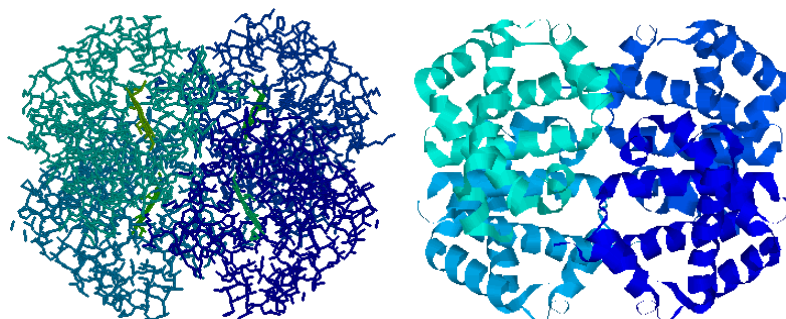
HEADER      TOXIN                               05-JUL-99   1QUZ
TITLE      SOLUTION STRUCTURE OF THE POTASSIUM CHANNEL SCORPION TOXIN
TITLE      2 HSTX1
...
FORMUL     2  NH2      H2  N
HELIX      1  1  THR  A    5  ASP  A    8  5                               4
HELIX      2  2  CYS  A    9  THR  A   17  1                               9
SHEET      1  A 2  GLY  A   22  CYS  A   24  0
SHEET      2  A 2  CYS  A   29  CYS  A   31 -1  N  LYS  A   30  O  LYS  A   23
...
ATOM       1  N  ALA  A    1    -0.160 -2.742  1.881  1.00  0.00                N
ATOM       2  CA ALA  A    1    -0.518 -1.295  1.723  1.00  0.00                C
ATOM       3  C  ALA  A    1    -1.469 -0.812  2.802  1.00  0.00                C
ATOM       4  O  ALA  A    1    -2.260 -1.596  3.325  1.00  0.00                O
ATOM       5  CB ALA  A    1    -1.160 -1.097  0.320  1.00  0.00                C
ATOM       6  HA ALA  A    1     0.397 -0.723  1.804  1.00  0.00                H
ATOM       7  HB1 ALA  A    1    -1.378 -0.023  0.113  1.00  0.00                H
ATOM       8  HB2 ALA  A    1    -0.478 -1.441 -0.488  1.00  0.00                H
ATOM       9  HB3 ALA  A    1    -2.118 -1.658  0.210  1.00  0.00                H
ATOM      10  H1  ALA  A    1     0.608 -3.000  1.227  1.00  0.00                H
ATOM      11  H2  ALA  A    1    -1.010 -3.298  1.659  1.00  0.00                H
ATOM      12  H3  ALA  A    1     0.117 -2.962  2.868  1.00  0.00                H
ATOM      13  N  SER  A    2    -1.405  0.499  3.146  1.00  0.00                N
ATOM      14  CA  SER  A    2    -2.214  1.175  4.166  1.00  0.00                C
ATOM      15  C  SER  A    2    -3.281  2.067  3.554  1.00  0.00                C
ATOM      16  O  SER  A    2    -3.196  2.444  2.388  1.00  0.00                O
ATOM      17  CB  SER  A    2    -1.329  2.014  5.138  1.00  0.00                C
ATOM      18  OG  SER  A    2    -0.539  1.133  5.932  1.00  0.00                O
...

```

Rys. 5. Fragment dokumentu PDB opisującego przykładową strukturę białkową (PDB ID: 1QUZ). Zaprezentowano współrzędne atomów aminokwasu alaniny i niektórych atomów seryny

Fig. 5. Fragment of the PDB document describing structure of the potassium channel scorpion toxin HSTX1 (PDB ID: 1QUZ). Atomic coordinates of Alanine and some atoms of Serine are visible

#### 2.4. Struktura czwartorzędowa



Rys. 6. Struktury czwartorzędowe hemoglobiny (PDB ID: 4HHB) w różnych reprezentacjach

Fig. 6. Various representations of quaternary structure of hemoglobin molecule (PDB ID: 4HHB)

Strukturę czwartorzędową wyróżnia się w białkach składających się z więcej niż jednego łańcucha aminokwasów. Struktura ta opisuje wzajemne ułożenie podjednostek (łańcuchów)

i rodzaj ich kontaktu, który może mieć charakter kowalencyjny lub niekowalencyjny [8], [20]. W procesie porównania struktur białkowych mających struktury czwartorzędowe każdy z łańcuchów jest zwykle porównywany osobno. Z każdej struktury czwartorzędowej zostają wyekstrahowane poszczególne podjednostki i porównanie przebiega tak jak dla struktur trzeciorzędowych. Na rys. 6 zaprezentowano struktury czwartorzędowe białka hemoglobiny ludzkiej (PDB ID: 4HHB). Kolorami wyróżniono poszczególne łańcuchy struktury.

### 3. Cechy reprezentatywne struktur białkowych w procesie ich porównania

Poszukiwanie podobieństwa białek polega na porównaniu atrybutów dotyczących budowy przestrzennej białka zadanego przez użytkownika (np. lekarza diagnostę) oraz białek referencyjnych pobranych z bazy danych. Poszukiwanie prowadzone jest zatem parami – porównaniu podlega zadana molekula z całą grupą białek przechowywanych w bazie danych. Poszukiwanie podobieństwa struktur molekularnych ma charakter aproksymacyjny. Polega on na tym, że w wyniku takiego poszukiwania interesują nas nie tylko białka o identycznej budowie, ale również takie, które – mimo że nie są takie same – wykazują bardzo duże podobieństwo kształtu.

Samo badanie regionów przestrzennych białka nie należy do zadań prostych, gdyż białka składają się z setek aminokwasów, a te z kolei zbudowane są z od kilku do kilkunastu atomów różnych pierwiastków. Porównywanie układu wszystkich atomów w badanych strukturach byłoby zbyt czasochłonne. Z tego powodu zwykle poszukuje się zestawu cech opisujących dla struktur białkowych, za pomocą których można następnie reprezentować te struktury w procesie ich porównania. Nie ma jednoznacznej odpowiedzi, które cechy najlepiej opisują strukturę białkową. Zależy to od celu poszukiwania podobieństwa białek i zaimplementowanego algorytmu porównania. W światowej literaturze problemowi temu poświęcono wiele uwagi, m.in. w pracach [8, 14, 15, 21-30].

Według I. Eidhammer i in. [16] idealny opis struktury powinien jednak mieć następujące własności:

1. *Niezmiennność* – sposób reprezentacji struktury powinien być odporny na operacje geometryczne, takie jak translacja i rotacja.
2. *Solidność* – opis struktury nie powinien się drastycznie zmieniać przy niewielkich zmianach w strukturze. Jest to istotne, ponieważ zbiory PDB mogą zawierać drobne błędy.
3. Podobne struktury powinny mieć podobne deskryptory.
4. Niepodobne struktury powinny mieć różne deskryptory.



### 3.1. Zastosowanie macierzy odległości jako deskryptorów struktur białkowych

Białka złożone są z ciągu aminokwasów, w których każdy zawiera dokładnie jeden charakterystyczny atom  $C_\alpha$ . Zestawienie w przestrzeni trójwymiarowej ciągu takich atomów tworzy szkielet struktury białka (porównaj rys. 4b i 4d). W związku z tym, że struktury białek mogą składać się z wielu łańcuchów, interesują nas porównania łańcuchów tylko na poziomie jeden do jeden. Algorytmy porównania struktur białkowych, takie jak m.in. DALI [14], często redukują skomplikowane struktury białkowe do tzw. sztywnych rdzeni lub korpusów (ang. *rigid bodies*), utworzonych przez połączenie atomów  $C_\alpha$  kolejnych aminokwasów.

Zatem w przypadku  $k$ -tego łańcucha, składającego się z  $n$  aminokwasów, szkielet jego struktury zostanie zdefiniowany jako:  $\Omega_k = \{C_{\alpha,1}^k, C_{\alpha,2}^k, \dots, C_{\alpha,n}^k\}$ . Przy porównywaniu struktur molekularnych jesteśmy zainteresowani określeniem podobieństwa dwóch takich łańcuchów. W przypadku nałożenia dwóch łańcuchów o wysokim podobieństwie strukturalnym zostanie otrzymane również dobre dopasowanie szkieletów tych struktur. Przestrzenne relacje pomiędzy atomami  $C_\alpha$  w obu białkach są podobne dla białek wykazujących podobieństwo strukturalne. Dla białek, które nie wykazują podobieństwa strukturalnego, przestrzenne relacje są różne.

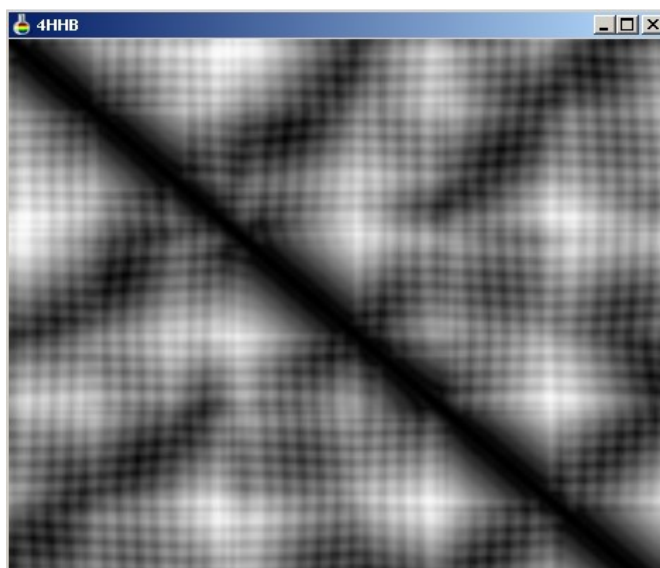
Podczas badań autorów nad opracowaniem odpowiednich deskryptorów struktur białkowych zdecydowano się reprezentować struktury w postaci macierzy odległości międzyrezydualnych. Na podstawie współrzędnych atomów struktury każdego białka obliczano macierze odległości atomów  $C_\alpha$ - $C_\alpha$  w tej strukturze z wykorzystaniem odległości euklidesowej:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (1)$$

gdzie  $x, y \in \mathfrak{R}^n$ .

Zastosowanie macierzy odległości (zwanymi również odwzorowaniami lub rzutami odległości) w opisywaniu i porównywaniu konformacji białek jest używane już od dłuższego czasu, np. w algorytmie DALI [14]. Najczęściej stosowane są macierze, które zawierają wszystkie odległości pomiędzy centrami reszt aminokwasów, tzw. atomami  $C_\alpha$ . Macierz taka jest reprezentacją dwuwymiarową przestrzeni trójwymiarowej. Powstałe macierze są niezależne od układu współrzędnych i zawierają dość informacji, aby zrekonstruować struktury przestrzenne (z wyjątkiem ogólnej chiralności).

Następnie macierze odległości zostają przetransformowane do macierzy odcieni szarości (rys. 7). Krok ten ma na celu zwiększenie elastyczności procesu porównania struktur. Do tego celu został użyty algorytm hierarchicznego grupowania [31]. W celu wyznaczenia grup arbitralnie przyjęto regularny podział dziedziny (odległości). Liczbę grup określono na 256, czyli tyle, ile przyjęto poziomów szarości powstałego z macierzy odległości obrazu.



Rys. 7. Macierz odcieni szarości dla białka hemoglobiny (PDB ID: 4HHB) z bazy PDB, powstała z transformacji macierzy odległości. Ciemne obszary oznaczają mniejszą odległość, jasne – większą

Fig. 7. Grayscale matrix for hemoglobin molecule (PDB ID: 4HHB) from PDB, generated by transforming the distance matrix. Dark areas indicate smaller distances, bright – higher

Autorzy przyjęli, że każda z porównywanych struktur białkowych jest reprezentowana przez obraz powstały z macierzy odcieni szarości. Wstępna analiza porównawcza struktur białek zapisanych za pomocą obrazów przebiegała zatem przez porównywanie powstałych obrazów. Do badania podobieństwa obrazów została wykorzystana wartość współczynnika jakości obrazu  $Q$  (ang. *Universal Image Quality Index*), która jest określona zgodnie z tym co następuje [32].

Założmy, że  $x = \{x_i | i=1, 2, \dots, N\}$  to sygnał obrazu oryginalnego, a  $y = \{y_i | i=1, 2, \dots, N\}$  to sygnał obrazu testowego. Indeks jakości jest zdefiniowany jako:

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{(\sigma_x^2 + \sigma_y^2)[(\bar{x})^2 + (\bar{y})^2]}, \quad (2)$$

gdzie poszczególne składowe wyrażenia, to:

- średnie arytmetyczne

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{oraz} \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad (3)$$

- wariancje

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad \text{oraz} \quad \sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2, \quad (4)$$

- kowariancja

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}). \quad (5)$$

Zmienna  $Q$  przyjmuje wartości z przedziału  $[-1, 1]$ . Najlepszą wartością jest 1, która jest osiągnięta tylko w przypadku, gdy  $y_i = x_i$  dla każdego  $i=1,2,\dots,N$ . Najgorsza wartość -1 występuje w sytuacji, gdy  $y_i = 2\bar{x} - x_i$  dla każdego  $i=1,2,\dots,N$ . Indeks jakości  $Q$  przedstawia współczynnik dopasowania dwóch obrazów, na który składają się trzy wskaźniki: utrata korelacji, różnica luminancji oraz różnica kontrastu.

Aby to zrozumieć, współczynnik dopasowania  $Q$  można przedstawić jako składową trzech komponentów:

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \quad (6)$$

Pierwszy komponent to współczynnik korelacji pomiędzy obrazami  $x$  i  $y$ , który mierzy stopień liniowego powiązania dla tych obrazów. Wartość zwracana mieści się w przedziale  $[-1, 1]$ . Wartość maksymalna przyjmowana jest w przypadku, gdy  $y_i = ax_i + b$  dla każdego  $i=1, 2, \dots, N$ , gdzie  $a$  i  $b$  są stałymi i  $a > 0$ . Nawet jeśli  $x$  i  $y$  są liniowo zależne, wtedy wciąż mogą wystąpić względne odchylenia pomiędzy nimi. Wówczas wyliczane są one w komponentach drugim i trzecim.

Drugi komponent, którego wartość oscyluje w przedziale  $[0, 1]$ , określa stopień podobieństwa przeciętnej luminancji pomiędzy  $x$  i  $y$ . Osiągnięcie wartości 1 jest możliwe tylko wtedy, gdy  $\bar{x} = \bar{y}$ .  $\sigma_x$  i  $\sigma_y$  mogą być postrzegane jako ocena kontrastu z  $x$  i  $y$ , tak więc trzeci komponent mierzy podobieństwo kontrastów obrazów. Jego wartość również oscyluje w przedziale  $[0, 1]$ , a najlepsza wartość osiągnięta jest tylko w przypadku, gdy  $\sigma_x = \sigma_y$ .

W przypadku obrazów identycznych współczynnik  $Q$  osiąga wartość równą 1. W przypadku, gdy mamy do czynienia z całkowicie odmiennymi obrazami, współczynnik  $Q$  będzie równy -1. Metryka ta została wykorzystana przez autorów do badania podobieństwa pomiędzy fragmentami macierzy odcieni szarości.

### 3.2. Wstępne zestawienie białek (poziom globalny)

Zestawienie wstępne polegało na znalezieniu najlepszego wskaźnika dopasowania dwóch badanych macierzy odległości przez wyliczenie indeksu jakości na podstawie obrazów utworzonych z tych macierzy. Dopasowanie następuje na etapie całych macierzy. W związku z tym, że badane obrazy muszą być takich samych rozmiarów, to dopasowanie może przebiegać w kilku cyklach. W przypadku gdy dwa obrazy są takich samych rozmiarów, wtedy następuje wyliczenie wskaźnika jakości obrazu  $Q$ . Gdy jednak rozmiary badanych obrazów są różne, wtedy liczba dopasowań jest większa o różnicę jednego z wymiarów tych obrazów. Na rys. 8 zostały przedstawione przykładowe macierze odległości dla dwóch różnych białek (schemat poglądowy). Obie macierze są symetryczne wzdłuż przekątnych, które zostały

oznaczone za pomocą symbolu x. W związku z tym szerokość każdej z macierzy odpowiada jej wysokości. W omawianym przykładzie macierz A jest większa od macierzy B. Pierwsza ma rozmiar 12x12, druga 9x9, wobec czego badanie globalnego podobieństwa nastąpi czterokrotnie.

	1	2	3	4	5	6	7	8	9	10	11	12		1	2	3	4	5	6	7	8	9	
1	x												1	x									
2		x											2		x								
3			x										3			x							
4				x									4				x						
5					x								5					x					
6						x							6						x				
7							x						7							x			
8								x					8								x		
9									x				9									x	
10										x													
11											x												
12												x											

Macierz A

Macierz B

Rys. 8. Zestawienie wstępne dwóch macierzy

Fig. 8. Preliminary juxtaposition of two distance matrices

W każdym z przypadków cała mniejsza macierz jest badana z fragmentem większej macierzy w taki sposób, że ich przekątne się pokrywają (rys. 7, macierz A). W pierwszym etapie element (1,1) macierzy B zestawiany jest z elementem (1,1) macierzy A, element (1,2) macierzy B z elementem (1,2) macierzy A itd. W kolejnym powtórzeniu następuje badanie podobieństwa, gdy macierz B zostaje przesunięta wzdłuż przekątnej macierzy A o wartość jednej komórki. W tym wypadku element (1,1) macierzy B zostaje zestawiony z elementem (2,2) macierzy B. Dla każdego z tych czterech porównań jest zwracany indeks dopasowania globalnego obu macierzy. Jako wynik przedstawiany jest ten, którego wartość jest największa:

$$Q_G = \max\{Q_1, Q_2, \dots, Q_n\}, \quad (7)$$

gdzie:

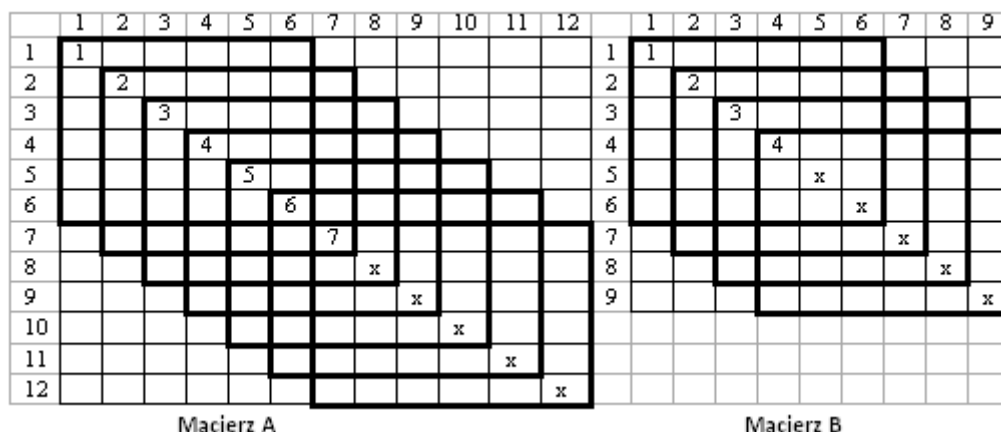
$$n = 1 + |\text{matrixA.width} - \text{matrixB.width}|, \quad (8)$$

$$\text{lub } n = 1 + |\text{matrixA.height} - \text{matrixB.height}|. \quad (9)$$

### 3.3. Szczegółowe zestawienie białek (poziom szczegółowy)

Wyszukanie podobieństw występujących wewnątrz struktury zrealizowano przez zestawienie szczegółowe, ponieważ może się zdarzyć, że białka nie wykazują podobieństwa na całej długości, lecz tylko w określonych obszarach. Porównanie szczegółowe polegało na dopasowywaniu okien o wymiarach 6 x 6 z każdej macierzy odległości, a nie jak w poprzednim przypadku całego okna mniejszej macierzy. Wszystkie okna dobierane były tak, że ich pierwszy element należał do przekątnej danej macierzy. Liczba takich okien dla każdej ma-

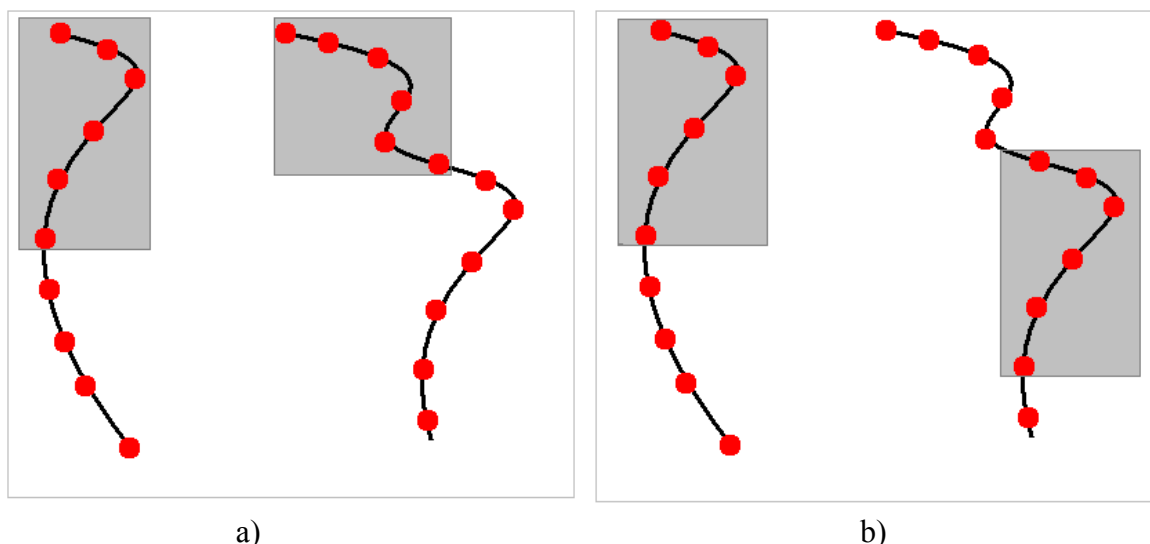
ciery równa jest szerokości (wysokości) takiej macierzy pomniejszonej o wartość 5. Na rys. 9 przedstawiono wszystkie możliwe okna o rozmiarach 6 x 6 w każdej z macierzy z zachowaniem warunku, że pierwszy element znajduje się na przekątnej.



Rys. 9. Generowanie wszystkich par okien 6x6 z poszczególnych macierzy  
 Fig. 9. Generation of all pairs of 6x6 windows in both matrices

Wybierane fragmenty macierzy odpowiadają dopasowaniom, jak to przedstawiono na rys. 10. Każda podmacierz o rozmiarze 6 x 6 odpowiada fragmentowi struktury obejmującemu sześć kolejnych atomów  $C_{\alpha}$ .

W omawianym przypadku (rys. 9) dla macierzy A zostało wygenerowanych 7 okien, natomiast dla macierzy B tylko 4 takie okna. Dopasowanie polegało na zestawieniu każdego okna z macierzy A z każdym oknem z macierzy B (na zasadzie każdym z każdym). W ten sposób, gdy macierz A zawiera  $n$  okien, a macierz B ma  $m$  okien, wtedy całkowita liczba par wynosi  $n \times m$ . W tym przypadku par jest 28.



Rys. 10. Dopasowanie fragmentów poszczególnych struktur: a) badane fragmenty nie są podobne, b) badane fragmenty są identyczne  
 Fig. 10. Matching fragments of both structures: a) marked fragments are not similar, b) marked fragments are identical

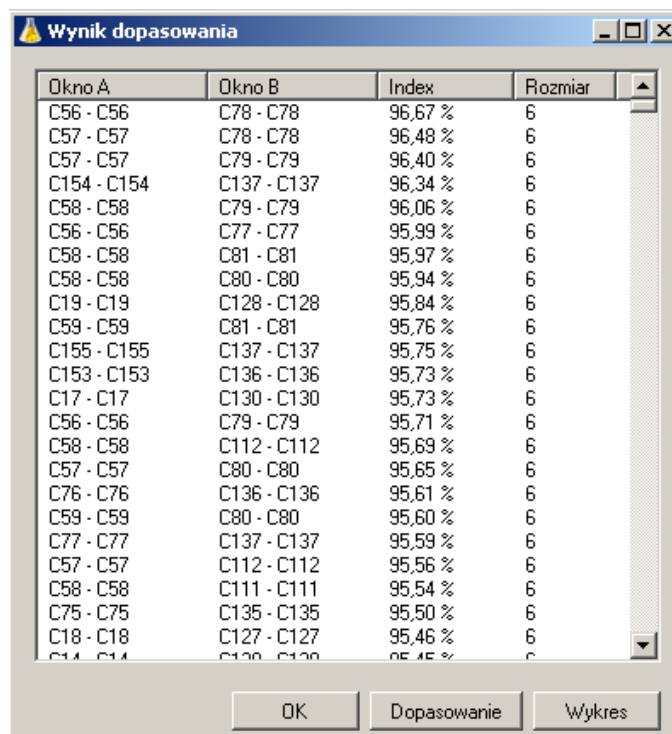
Dla każdej takiej pary obliczany był wskaźnik jakości obrazu Q, którego wartość powinna oscylować w przedziale [0, 2] (w rzeczywistości była to wartość Q+1). Na rys. 11 przedstawiono przykładowe wyniki porównania zadanej struktury białkowej 1N6I z kilkoma innymi strukturami z bazy PDB. Struktury oznaczone identyfikatorami PDB ID: 1N6K, 1N6H i 1N6N są w rzeczywistości tym samym białkiem, co molekula oznaczona PDB ID: 1N6I, w którym wprowadzono drobne mutacje punktowe, które w konsekwencji wprowadzają nieznaczne odkształcenie struktury. Wartość indeksu Q dla każdej struktury jest bliska 2, co oznacza wysokie podobieństwo. Dla innych struktur z bazy danych wartość tego indeksu była znacznie niższa, co sugeruje, że struktury te nie są podobne do zadanej cząstki PDB ID: 1N6I i nie należy dla nich prowadzić szczegółowych obliczeń (zestawienia szczegółowego).



Symbol	Index
1N6K	1,99970890932999
1N6H	1,99964805192564
1N6N	1,99961775754043

Rys. 11. Wartości indeksu Q obliczone podczas porównania przykładowego białka (PDB ID: 1N6I) z grupą białek z bazy PDB (zestawienie globalne)

Fig. 11. Values of the Q index calculated during comparison of sample molecule (PDB ID: 1N6I) to a set of molecules from PDB (global comparison)



Okno A	Okno B	Index	Rozmiar
C56 - C56	C78 - C78	96,67 %	6
C57 - C57	C78 - C78	96,48 %	6
C57 - C57	C79 - C79	96,40 %	6
C154 - C154	C137 - C137	96,34 %	6
C58 - C58	C79 - C79	96,06 %	6
C56 - C56	C77 - C77	95,99 %	6
C58 - C58	C81 - C81	95,97 %	6
C58 - C58	C80 - C80	95,94 %	6
C19 - C19	C128 - C128	95,84 %	6
C59 - C59	C81 - C81	95,76 %	6
C155 - C155	C137 - C137	95,75 %	6
C153 - C153	C136 - C136	95,73 %	6
C17 - C17	C130 - C130	95,73 %	6
C56 - C56	C79 - C79	95,71 %	6
C58 - C58	C112 - C112	95,69 %	6
C57 - C57	C80 - C80	95,65 %	6
C76 - C76	C136 - C136	95,61 %	6
C59 - C59	C80 - C80	95,60 %	6
C77 - C77	C137 - C137	95,59 %	6
C57 - C57	C112 - C112	95,56 %	6
C58 - C58	C111 - C111	95,54 %	6
C75 - C75	C135 - C135	95,50 %	6
C18 - C18	C127 - C127	95,46 %	6
C14 - C14	C120 - C120	95,45 %	6

Rys. 12. Wyniki szczegółowego zestawienia pary białek – zadanego (PDB ID: 1N6I) oraz wybranego białka z bazy danych (PDB ID: 1N6K)

Fig. 12. Results of detailed comparison for a pair of molecules – query (PDB ID: 1N6I) and one of the database molecules (PDB ID: 1N6K)

Na tym etapie badań porównania struktur białkowych były prowadzone w grupie stu losowo wybranych struktur z bazy Protein Data Bank. Porównania dokonano przy użyciu autorskiego programu.

Dopasowanie szczegółowe, przedstawione na rys. 12, prezentuje natomiast listę okien 6 x 6 z porównywanych białek (PDB ID: 1N6I oraz 1N6K) z wyliczonym dla nich procentowym podobieństwem. Poszukując w pokazanym zestawieniu najdłuższej sekwencji odpowiadających sobie okien, można wyznaczyć dopasowanie (ang. *alignment*) porównywanych struktur białkowych.

#### 4. Podsumowanie

Przeprowadzone badania potwierdzają, że reprezentacja struktur białkowych w postaci macierzy odległości, transformowanych następnie do macierzy odcieni szarości, daje możliwość porównania przestrzennych konstrukcji białek. Zastosowanie współczynnika jakości obrazu Q pozwala na porównanie macierzy odległości opisujących te białka i znalezienie regionów obu białek, wykazujących podobieństwo kształtu. Użycie współczynnika Q pozwala również na przeszukanie całej bazy danych w poszukiwaniu białek, których struktury mogą być podobne (identyczne lub zbliżone) do zadanego białka. Wyniki badań potwierdzają zatem, iż przyjęty sposób reprezentacji jest skuteczny w procesie porównania struktur białkowych.

Pomimo redukcji przestrzeni trójwymiarowej (w której opisane są struktury białkowe) do płaszczyzny dwuwymiarowej ten sposób reprezentacji ma jednak istotne wady, są nimi złożoność obliczeniowa pojedynczego porównania i czas potrzebny na jego wykonanie. Czyni to trudnym wykorzystanie takiego sposobu porównania w procesie poszukiwania podobieństwa białek, w którym proces porównania jest wykonywany iteracyjnie dla każdego białka z bazy danych. Z tego powodu w przyszłych badaniach autorzy będą poszukiwać mniej złożonych sposobów reprezentacji struktur, np. w postaci deskryptorów jednowymiarowych, lub postarają się dodatkowo ograniczyć zbiór potencjalnych struktur z bazy danych, które powinny podlegać dalszemu porównaniu.

*Praca naukowa finansowana ze środków na naukę w latach 2008-2011 jako projekt badawczy N N516 265835: Poszukiwanie podobieństwa strukturalnego białek w rozproszonym środowisku wieloagentowym.*

**BIBLIOGRAFIA**

1. Dickerson R. E., Geis I.: *The Structure and Action of Proteins*, 2nd edition. Benjamin/Cummings, Redwood City, Calif. Concise 1981.
2. Bańkowski E.: *Biochemia. Podręcznik dla studentów uczelni medycznych*. Wydawnictwo Medyczne Urban & Partner, Wrocław 2004.
3. Creighton T. E.: *Proteins: Structures and molecular properties*. 2nd ed. Freeman, San Francisco 1993.
4. Branden C., Tooze J.: *Introduction to Protein Structure*. Garland 1991.
5. Orengo C. A., Michie A. D., Jones S., Jones D. T. et al.: CATH – A Hierarchic Classification of Protein Domain Structures. “Structure”, Vol. 5, No. 8, 1997, s. 1093÷1108.
6. Murzin A. G., Brenner S. E., Hubbard T., Chothia C.: SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J. Mol. Biol.*, No. 247, 1995, s. 536÷540.
7. Pearson W. R., Lipman D. J.: Improved Tools for Biological Sequence Analysis. *PNAS*, No. 85, 1988, s. 2444÷2448.
8. Gu J., Bourne P. E.: *Structural Bioinformatics (Methods of Biochemical Analysis)*, 2 edition. Wiley-Blackwell, 2009.
9. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H. et al.: The Protein Data Bank. *Nucleic Acids Res.*, No. 28, 2000, s. 235÷242.
10. Mrozek D., Wiczorek D., Małysiak-Mrozek B., Kozielski S.: PSS-SQL: Protein Secondary Structure – Structured Query Language. 32th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2010, EMBS 2010. IEEE, Buenos Aires, Argentina 2010, s. 1073÷1076.
11. Wiczorek D., Małysiak-Mrozek B., Kozielski S., Mrozek D.: A metod for matching sequences of protein secondary structures. *Journal of Medical Informatics & Technologies*, Vol. 16, 2010, s. 133÷137.
12. Frishman D., Argos P.: Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.*, Vol. 9(2), 1996, s. 133÷142.
13. Sayle R.: *RasMol, Molecular Graphics Visualization Tool*. Biomolecular Structures Group, Glaxo Welcome Research & Development, Stevenage, Hartfordshire 1998.
14. Holm L., Sander C.: Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, Vol. 233(1), 1993, s. 123÷138.
15. Can T., Wang Y. F.: CTSS: A Robust and Efficient Method for Protein Structure Alignment Based on Local Geometrical and Biological Features. *Proceedings of the 2003 IEEE Bioinformatics Conference (CSB 2003)*, 2003, s. 169÷179.



16. Eidhammer I., Inge J., Taylor W. R.: Protein Bioinformatics: An Algorithmic Approach to Sequence and Structure Analysis. John Wiley & Sons, 2004.
17. Westbrook J. D., Fitzgerald P. M. D.: The PDB Format, mmCIF Formats, and Other Data Formats. Structural Bioinformatics, Vol. 44, John Wiley & Sons, Inc., 2003.
18. Bourne P. E., Berman H. M., Watenpaugh K. et al.: The macromolecular Crystallographic Information File (mmCIF). Methods Enzymol., Vol. 277, 1997, s. 571÷590.
19. Westbrook J., Ito N., Nakamura H., Henrick K., Berman H.M.: PDBML: the representation of archival macromolecular structure data in XML. Bioinformatics, Vol. 21(7), 2005, s. 988÷992.
20. Hames B. D., Hooper N. M., Houghton J. D.: Krótkie wykłady – Biochemia. Wydawnictwo Naukowe PWN, Warszawa 2000.
21. Taylor W. R., Orengo C. A.: Protein structure alignment. J. Mol. Biol., Vol. 208, 1989, s. 1÷22.
22. Holm L, Sander C.: The FSSP database of structurally aligned protein fold families. Nucleic Acids Res., Vol. 22(17), 1994, s. 3600÷3609.
23. Gibrat J. F., Madej T., Bryant S.H.: Surprising similarities in structure comparison. Curr Opin Struct Biol, Vol. 6(3), 1996, s. 377÷385.
24. Godzik A.: The structural alignment between two proteins: is there a unique answer? Protein Sci., Vol. 5, 1996, s. 1325÷1338.
25. Shindyalov I. N., Bourne P. E.: Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Engineering, Vol. 11(9), 1998, s. 739÷747.
26. Plewczynski D., Pas J., von Grotthuss M., Rychlewski L.: 3D-Hit: fast structural comparison of proteins. Appl Bioinformatics., Vol. 1, 2002, s. 223÷225.
27. Ye Y., Godzik A.: Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics, Vol. 19(2), 2003, s. 246÷255.
28. Shapiro J., Brutlag D.: FoldMiner and LOCK2: protein structure comparison and motif discovery on the web. Nucleic Acids Res., Vol. 32, 2004, s. 536÷541.
29. Jia Y., Dewey G., Shindyalov I. N., Bourne P. E.: A New Scoring Function and Associated Statistical Significance for Structure Alignment by CE. J Comp Biol., Vol. 11, No. 5, 2004, s. 787÷799.
30. Yang J.: Comprehensive description of protein structures using protein folding shape code. Proteins, Vol. 71(3), 2008, s. 1497÷1518.
31. Kolatch E.: Clustering algorithms for spatial databases: a survey. University of Maryland, 2001.
32. Wang Z., Bovik A. C.: Universal Image Quality Index. IEEE Signal Processing Letters, Vol. 9, No. 3, 2002.

Wpłynęło do Redakcji 31 stycznia 2012 r.

## Abstract

Comparison of protein, molecular structures is often an essential component process of protein structure similarity searching, identification of protein functions, and investigation of the evolution of living organisms. Protein comparison can be carried out on different organization levels of protein structure, beginning from the primary structure, and ending on the quaternary structure. Each of the description levels provides different sort of information, and the construction of a protein itself can be described and represented in a variety of ways in particular IT systems. Depending on the purpose of the comparison process, scientists will be interested in a less detailed or more detailed description of protein structure. Nevertheless, in any case, we have to focus on appropriate features of internal construction during the analysis and comparison. An effective representation of protein structures in the comparison process is then very important for its successfulness and swiftness.

In the paper, taking into account four organization levels of protein structures, we present considerations on using various, representative features describing protein structures in the comparison process. We also show widely accepted representations of protein structure descriptors used by computer systems. Finally, we demonstrate our research on protein structure comparison with the use of intra-residual distance matrices, which we adopted as a primary protein structure descriptor. These distance matrices are transformed to the grayscale matrices and we use the Universal Image Quality Index in order to (1) compare grayscale images that represent protein structures and (2) search a database of protein structures looking for similar molecules.

## Adresy

Bożena MAŁYSIAK-MROZEK: Politechnika Śląska, Instytut Informatyki,  
ul. Akademicka 16, 44-100 Gliwice, Polska, bozena.malysiak@polsl.pl.

Dariusz MROZEK: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16,  
44-100 Gliwice, Polska, dariusz.mrozek@polsl.pl.

Łukasz KOŁKOWSKI: absolwent Instytutu Informatyki, Politechniki Śląskiej,  
ul. Akademicka 16, 44-100 Gliwice, Polska, lukasz.kolkowski@polsl.pl.