

DAMIR MOGUT, ANNA IWANIAK, MAŁGORZATA DAREWICZ

**ZASTOSOWANIE ANALIZY GŁÓWNYCH SKŁADOWYCH DO BADANIA  
ZALEŻNOŚCI MIĘDZY STRUKTURĄ A AKTYWNOŚCIĄ  
PRZECIWUTLENIAJĄCĄ DIPEPTYDÓW POCHODZĄCYCH  
Z BIAŁEK ŻYWNOSCI**

Streszczenie

W pracy zastosowano model statystyczny utworzony za pomocą analizy głównych składowych (PCA) do określenia wpływu struktury dipeptydów na ich aktywność przeciwutleniającą. Sekwencje 47 peptydów pobrano z bazy danych sekwencji białek i bioaktywnych peptydów BIOPEP-UWM (<https://biochemia.uwm.edu.pl/biopep-uwm/>). Wybrane deskryptory (inaczej atrybuty) opisujące właściwości fizykochemiczne aminokwasów wchodzących w skład dipeptydów były wyrażone liczbowo i zaimplementowane z ogólnodostępnych programów lub baz danych: Peptide Property Calculator, Biological Magnetic Resonance Data Bank, ProtScale, Molar Polarizability Values oraz ImMunoGeneTics. PCA wykonano za pomocą programu Statistica. Liczbę składowych głównych istotnych w interpretacji wpływu struktury dipeptydów na ich aktywność przeciwutleniającą określono na podstawie procentowego wyjaśnienia ogółu wariancji. Wynosił on 79,9 %, co odpowiadało czterem składowym. Pierwsza i czwarta składowa decydowały o wpływie aminokwasu N-końcowego na aktywność przeciwutleniającą dipeptydu, natomiast druga i trzecia dotyczyły wpływu reszty C-końcowej. Za pomocą PCA wykazano, że modelowy peptyd o aktywności przeciwutleniającej powinien charakteryzować się obecnością N-końcowego aminokwasu cyklicznego lub aromatycznego albo reszty aminokwasu z apolarnym łańcuchem bocznym. C-koniec peptydu powinien być zbudowany z proliny, histydyny, leucyny czy waliny. Uzyskane wyniki są zgodne z rezultatami literaturowymi dotyczącymi badań na temat zależności między strukturą a aktywnością przeciwutleniającą peptydów, oszacowanymi metodą regresji wielorakiej. Na podstawie przeprowadzonych badań stwierdzono, że zastosowana metoda chemometryczna może być przydatna w projektowaniu peptydów pochodzących z żywności, w tym o badanej aktywności.

**Słowa kluczowe:** aktywność przeciwutleniająca, dipeptydy, baza danych BIOPEP-UWM, chemometria, PCA

---

*Dr inż. D. Mogut, prof. dr hab. inż. A. Iwaniak, prof. dr hab. inż. M. Darewicz, Katedra Biochemii Żywności, Wydz. Nauki o Żywności, Uniwersytet Warmińsko-Mazurski w Olsztynie, Pl. Cieszyński 1, 10-719 Olsztyn. Kontakt: damir.mogut@uwm.edu.pl*

## Wprowadzenie

Białka żywności pełnią w organizmie wiele funkcji, m.in. enzymatyczną, budulcową, immunologiczną [18], a także mogą być źródłem peptydów o wielu aktywnościach biologicznych [12], w tym przeciwutleniającej [32]. Według dostępnej literatury źródłem peptydów o aktywności przeciwutleniającej są m.in. białka ryżu, soi, słonecznika, gryki czy kukurydzy, a także pochodzenia zwierzęcego, w tym jaja kurzego, wielu ryb i owoców morza czy fermentowanych produktów mlecznych [25].

Bez względu na rodzaj aktywności biologicznej, jaką charakteryzują się peptydy pochodzące z białek żywności, istnieją trzy kierunki badawcze ich analizy [8]. Pierwszy nazywany jest kierunkiem klasycznym i obejmuje takie etapy, jak: wybór materiału do badań (białka), jego hydrolizę enzymatyczną, oznaczenie aktywności hydrolizatu oraz identyfikację peptydów za pomocą spektrometrii mas. Drugi kierunek, bioinformatyczny, wspomaga badania eksperymentalne i jest związany ze stosowaniem metod *in silico* w przewidywaniu właściwości peptydów, symulowaniu hydrolizy białek czy badaniu zależności aktywności biologicznej peptydów od ich struktury. Kierunek ten jest ściśle powiązany z tworzeniem oraz wykorzystywaniem baz danych, w tym baz danych peptydów bioaktywnych [30]. Trzeci kierunek to tzw. analiza hybrydowa (pojęcie zintegrowane), która jest połączeniem obu ww. kierunków [8].

Badania nad zależnością między strukturą peptydów a ich aktywnością biologiczną w obrębie kierunku bioinformatycznego są ściśle powiązane z analizą chemometryczną. Wówczas zasoby bioinformatycznych baz danych służą do tworzenia własnych zbiorów danych, które są następnie analizowane za pomocą metod chemometrycznych, wywodzących się z technik analizy statystycznej [11]. Zaliczana jest do nich m.in. analiza głównych składowych (ang. *Principal Component Analysis*, PCA) [28]. Metoda ta została zastosowana do badania zależności „struktura-aktywność” (ang. *structure-activity relationship*, SAR) inhibitorów enzymu konwertującego angiotensynę (tzw. inhibitorów ACE; EC 3.4.15.1) [27], inhibitorów dipeptydylopeptydazy IV (EC 3.4.14.5) [16] czy peptydów o smaku gorzkim [10]. W literaturze przedmiotu nie ma natomiast wielu danych na temat zastosowania ww. metod w analizie zależności między strukturą a aktywnością peptydów o aktywności przeciwutleniającej. Przeszukiwanie bazy NCBI (tj. *National Centre for Biotechnology Information*; <http://www.ncbi.nlm.nih.gov>; dostęp: kwiecień 2021) za pomocą słów kluczowych „QSAR antioxidant peptides” umożliwiło wykazanie 1922 rekordów, w tym 1165 rekordów z ostatnich 5 lat. W jednej z publikacji podjęto próbę analizy peptydów przeciwutleniających pochodzących z hydrolizatów białek metodą najmniejszych cząstkowych kwadratów (ang. *Partial Least Squares*, PLS) [31]. Za celowe uznano więc zastosowanie PCA do badania wpływu struktury peptydów na ich aktywność przeciwutleniającą. Wśród analizowanych tytułów oraz treści streszczeń rekordów dostępnych w NCBI żaden nie dotyczył zastosowania tej metody w odniesieniu do badań poświęconych ww. zagadnieniu.

Celem pracy było określenie zależności między strukturą a aktywnością dipeptydów o aktywności przeciwutleniającej za pomocą analizy głównych składowych.

### Material i metody badań

Informacje dotyczące sekwencji 47 dipeptydów o aktywności przeciwutleniającej pobrano z bazy danych BIOPEP-UWM (dostęp: kwiecień 2021) oraz z literatury (tab. 1). Dane dotyczące źródła pochodzenia analizowanych peptydów, potwierdzonej aktywności czy cytowanej literatury są dostępne w BIOPEP-UWM po kliknięciu w link, będący jednocześnie numerem akcesyjnym sekwencji peptydu w bazie danych (BIOPEP-UWM ID) [20].

Tabela 1. Wykaz sekwencji dipeptydów o aktywności przeciwutleniającej pobranych z bazy danych BIOPEP-UWM oraz z literatury

Table 1. List of antioxidative dipeptides' sequences acquired from BIOPEP-UWM database and reference literature

Sekwencja Sequence	BIOPEP-UWM ID	Sekwencja Sequence	Literatura References
Leu-His	3305	Ala-Ala	[1]
His-Leu	3317	Gly-Gly	[1]
His-His	3319	His-Glu	[1]
Ala-Tyr	7866	His-Phe	[1]
Leu-Tyr	7872	His-Ile	[1]
Ile-Tyr	7873	His-Lys	[1]
Ala-His	7886	His-Ser	[1]
Glu-Leu	7888	Trp-Glu	[1]
Trp-Tyr	7898	Trp-Phe	[1]
Met-Tyr	8090	Trp-Ile	[1]
Lys-Asp	8134	Trp-Lys	[1]
Pro-Trp	8190	Trp-Ser	[1]
Arg-Trp	8214	Trp-Trp	[1]
Ile-Arg	8215	Trp-Pro	[17]
Leu-Lys	8217	Ala-Leu	[22]
Lys-Pro	8218	Glu-Lys	[22]
Thr-Tyr	8219	Phe-Leu	[22]
Val-Tyr	8224	Gly-Leu	[22]
Thr-Trp	8459	Ser-Phe	[22]
Ala-Trp	8460	Ser-Leu	[22]
Val-Trp	8461	Val-Ala	[22]
Leu-Trp	8462	Trp-Val	[22]
Trp-Gly	9082	-	-
Met-Met	9086	-	-
Phe-Cys	9342	-	-

Zmiennymi opisującymi właściwości dipeptydów (tzw. deskryptorami) były wyrażone liczbowo atrybuty przypisane aminokwasom wchodzącym w skład peptydu. Pobrano je z następujących baz i/lub programów komputerowych:

- a) AAIndex ([www.genome.jp/aaindex](http://www.genome.jp/aaindex)) – [15],
- b) Peptide Property Calculator ([biotools.nubic.northwestern.edu/proteincalc.html](http://biotools.nubic.northwestern.edu/proteincalc.html)),
- c) Biological Magnetic Resonance Data Bank ([www.bmrb.wisc.edu/ref\\_info/aadata.dat](http://www.bmrb.wisc.edu/ref_info/aadata.dat)) – [5],
- d) ProtScale ([web.expasy.org/protscale](http://web.expasy.org/protscale)) – [6],
- e) Molar Polarizability Values – Brigham Young University ([bio.groups.et.byu.net/Molar\\_Polarizability\\_alpha.phtml](http://bio.groups.et.byu.net/Molar_Polarizability_alpha.phtml)),
- f) ImMunoGeneTics ([imgt.org/IMGTeducation/Aide-memoire/\\_UK/aminoacids/abbreviation.html](http://imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/abbreviation.html)).

Atrybutami były: masa cząsteczkowa (M) [b], liczba atomów węgla (C) [c], sferyczność (S) [a], polarność (P) [a], polaryzowalność (POL) [e], tęgość (T) [a], rozmiar (R) [a], hydrofobowość według Kyte'a-Doolittle'a (H) [f], skłonność do formowania helisy (SFH) [a], powierzchnia reszt aminokwasowych ukrytych dla rozpuszczalnika (%PR) [a], prawdopodobieństwo występowania danego aminokwasu w białku (WAK) [d]. W nawiasach okrągłych podano symbole zmiennych, jakimi się posługiwano tworząc macierz danych, zaś w nawiasach kwadratowych pochodzenie zmiennych (tj. jeden z zastosowanych wyżej programów bądź baz komputerowych). W celu odróżnienia zmiennych charakteryzujących C- lub N-końcowe aminokwasy do każdego symbolu dodano oznakowanie „P1” lub „P2”. Przykładowo zmienną o symbolu M P1 należy rozumieć jako masę cząsteczkową aminokwasu N-końcowego, a M P2 – C-końcowego badanego peptydu. Łącznie uzyskano 22 zmienne, po 11 dla każdego aminokwasu wchodzącego w skład dipeptydu.

Przed przystąpieniem do analizy PCA jej zasadność w odniesieniu do utworzonej macierzy danych sprawdzono za pomocą testu KMO (Kayser-Meyer-Olkin). Jego wynik może przyjmować wartości od 0 do 1, a wartość KMO powyżej 0,50 uzasadnia słuszność zastosowanej metody analizy danych [10]. Ponadto przeprowadzono test sferyczności Bartletta. Pierwszy test miał na celu wskazanie prawidłowości (lub nieprawidłowości) zastosowanej analizy chemometrycznej do badania peptydów o aktywności przeciwutleniającej, natomiast drugi decydował o przyjęciu/odrzućeniu hipotezy o istnieniu właściwości wywierających wpływ na aktywność badanych sekwencji. Analizę wykonano za pomocą programu Statistica 10 [10].

Do wykonania analizy PCA zastosowano moduł analityczny o nazwie „Wielowymiarowe techniki eksploracyjne” z pakietu Statistica 10. Po wybraniu polecenia „Analiza składowych głównych i klasyfikacja” otrzymywano wyniki w postaci kart zapisu (raportów). Raporty zawierały informacje o ładunkach czynnikowych oraz wartościach własnych, wyrażanych zarówno liczbowo, jak i graficznie. Ładunki czynni-

kowe są to korelacje między składowymi (im wyższa wartość ładunku, tym bardziej jest skorelowana z daną składową). Wartość własna określa natomiast udział procentowy w wyjaśnianiu wpływu poszczególnych składowych [28].

## Wyniki i dyskusja

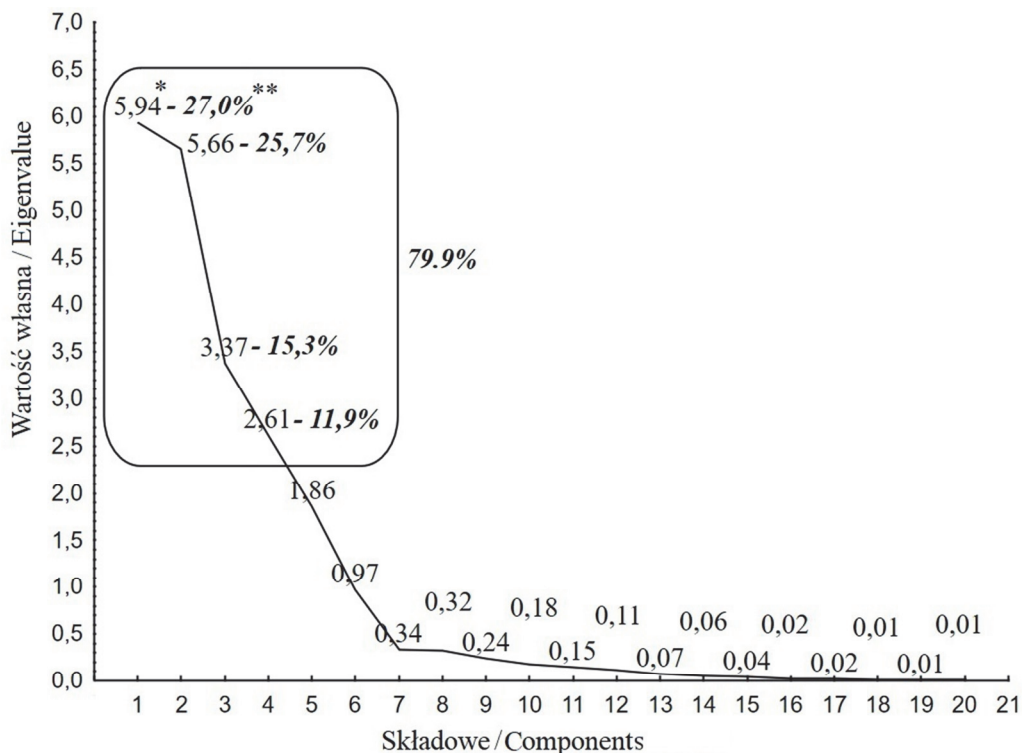
Metody chemometryczne, jak np. analiza składowych głównych, są wciąż przydatne w badaniu cech peptydów istotnych dla ich aktywności biologicznej. Przedstawiona procedura badawcza (konstruowanie modelu na podstawie informacji dostępnych w bazach danych i innych repozytoriach) może służyć badaniu peptydów o innych aktywnościach biologicznych. Została ona zaproponowana po raz pierwszy przez Iwaniak i wsp. [9, 10] do badania zależności między strukturą a gorzkim smakiem peptydów, a następnie wykorzystana przez innych autorów do badania zależności strukturalnych w odniesieniu do peptydowych inhibitorów dipeptydylopeptydazy IV (peptydów antydiabetycznych) [16].

PCA jest najstarszą oraz najlepiej poznaną metodą chemometryczną [10]. Jednym z najważniejszych i najtrudniejszych aspektów stosowania tej metody jest wyszczególnienie liczby składowych pomocnych we wskazaniu zależności między strukturą a funkcją peptydów [10]. Ponadto PCA często jest metodą uzupełniającą badanie zbiorów danych (macierzy) analizowanych za pomocą innych metod chemometrycznych. Wynika to z tego, że w niektórych przypadkach PCA pozwala na „natychmiastowe” wskazanie większej liczby atrybutów strukturalnych wpływających na funkcję biologiczną peptydów niż regresja wieloraka (ang. *multivariate linear regression*, MLR). Przykładem są badania Iwaniak i wsp. [10], w których, w wyniku PCA di- oraz tripeptydów o smaku gorzkim, wskazano więcej deskryptorów decydujących o tym wrażeniu smakowym niż w przypadku zastosowania MLR [9].

Jednym z wyzwań w tworzeniu macierzy danych PCA jest stosunek liczby obiektów (peptydów) do liczby zmiennych. Najczęściej przyjmowana jest wartość 5 : 1, przy czym Shaukat i wsp. [26] postulują, by minimalna liczba obiektów wynosiła 100. Ze względu na specyfikę badań przyrodniczych pozyskiwanie do badań chemometrycznych danych na temat określonych grup związków może być trudne [26], również w odniesieniu do peptydów o aktywności przeciwutleniającej, dlatego niektórzy autorzy postulują przyjęcie liczby obiektów do liczby zmiennych w proporcji 2 : 1, a badania symulacyjne dotyczące „idealnej liczby obiektów” w analizie wielu zmiennych wykazały, że trafne rezultaty otrzymywano już przy 19 obiektach [26]. W związku z powyższym 47 obiektów (dipeptydów o aktywności przeciwutleniającej) uznano za liczbę wystarczającą do przeprowadzenia PCA.

Testy KMO (ang. KMO value = 0,674) oraz sferyczności Bartletta ( $\chi^2 = 654,87$ ) określone dla macierzy dipeptydów oraz p wynoszące 0,05 (dane niezamieszczone) dały podstawę do przyjęcia hipotezy alternatywnej, tj. istnienia wpływu określonych

właściwości na aktywność przeciwutleniającą dipeptydów i kontynuowania badań. Na podstawie wyników PCA wykazano, że pierwsze cztery składowe wyjaśniały 79,9 % ogółu całkowitej wariancji i według przyjętych kryteriów zasobu zmienności uznano je za najważniejsze w wyjaśnieniu ich wpływu na aktywność przeciwutleniającą dipeptydów (rys. 1).



Objaśnienia / Explanatory notes:

(\*) – czcionka zwykła – wartość własna / normal font – eigenvalue; (\*\*) – pogrubiona kursywa – procent wyjaśnienia ogółu wariancji / bold and italic font – percent of cumulative variance explanation.

Rys. 1. Wykres Cattella przedstawiający istotną liczbę składowych decydujących o aktywności dipeptydów przeciwutleniających

Fig. 1. Cattell's plot showing the number of significant components deciding about the antioxidative activity of dipeptides

Znaczenie czterech składowych jako istotnych w wyjaśnieniu cech struktury decydujących o aktywności przeciwutleniającej dipeptydów zostało również potwierdzone analizą uwzględniającą rotację Varimax znormalizowaną (tab. 2).

Tabela 2. Wytypowane główne składowe wraz ze zmiennymi (z rotacją Varimax znormalizowaną)  
 Table 2. Selected principal components including variables (with standardised Varimax rotation)

Składowa 1 Component 1	Składowa 2 Component 2	Składowa 3 Component 3	Składowa 4 Component 4
(+) M P1	(+) M P2	(-) P P2	(-) P P1
(+) C P1	(+) C P2	(+) T P2	(+) T P1
(+) POL P1	(+) POL P2	(+) H P2	(+) H P1
(+) R P1	(+) R P2	-	-
(-) %PR P1	(-) PR P2	-	-
(-) WAK P1	(-) WAK P2	-	-

Objaśnienia / Explanatory notes:

Objaśnienia skrótów podano w rozdziale „Materiał i metody badań“ / Abbreviations are explained in chapter „Materiał i metody badań“; (+) – korelacja dodatnia / positive correlation; (-) – korelacja ujemna / negative correlation.

Pierwszą składową tworzyły atrybuty dotyczące właściwości reszty N-końcowej dipeptydu, czyli: masa cząsteczkowa (+0,98, M P1), liczba atomów węgla (+0,91, C P1), polaryzowalność (+0,98, POL P1), rozmiar (+0,91, R P1), powierzchnia reszt ukrytych dla rozpuszczalnika (-0,82, %PR P1) oraz prawdopodobieństwo występowania danego aminokwasu w białku (-0,88, WAK P1). W nawiasach podano wartości korelacji między zmiennymi tworzącymi składowe. Składową drugą tworzyły identyczne zmienne wraz z identycznymi tendencjami i charakteryzowały one właściwości aminokwasu

C-końcowego dipeptydu. W skład trzeciej składowej wchodziły zmienne opisujące aminokwas C-końcowy, tj. polarność – korelacja ujemna (-0,87), tęgość oraz hydrofobowość – korelacje dodatnie (+0,78 oraz +0,91). Te same zmienne o identycznym kierunku korelacji, czyli P (-0,90), T (+0,77) oraz H (+0,86) definiowały właściwości aminokwasu N-końcowego i stanowiły czwartą składową.

Zmienne, takie jak powierzchnia reszt ukrytych dla rozpuszczalnika czy prawdopodobieństwo występowania danego aminokwasu w białku, są atrybutami stosowanymi do opisu właściwości fizykochemicznych białek [3, 14]. Z kolei skłonność do formowania helisy może służyć opisowi właściwości zarówno peptydów, jak i białek [23]. Według Kajandera i wsp. [14] powierzchnia reszt ukrytych dla rozpuszczalnika wzrasta wraz ze wzrostem rozmiaru białka. Badania PCA dipeptydów o aktywności przeciwtuleniałej nie potwierdziły tej tendencji. Może to oznaczać, że zmienna ta nie powinna być stosowana do opisu właściwości peptydów lub analizowane peptydy charakteryzowały się zbyt krótkim łańcuchem do stosowania tej zmiennej w konstruowaniu modelu macierzy (ang. *input matrix*).

Prawdopodobieństwo występowania danego aminokwasu w białku jest parametrem często stosowanym do tworzenia algorytmów w celu przewidywania tendencji



białka do krystalizowania, przypisywania białka do klasy strukturalnej [2], czy stabilności białek [7]. Skład aminokwasowy białka jest istotnie związany z hydrofobowością. Białka hydrofilowe nie mogą formować hydrofobowego rdzenia typowego dla białek globularnych i nie wykazują tendencji do fałdowania, ponieważ ich wolna energia wewnętrzna jest niewystarczająca, by kompensować straty entropii niezbędnej w procesie fałdowania [3].

Interpretując uzyskane wyniki w odniesieniu do aminokwasów tworzących badane dipeptydy można stwierdzić, że składały się one przeważnie z N-końcowych aminokwasów zawierających hydrofobowy pierścień aromatyczny (Trp, Phe) lub apolarny łańcuch boczny (Leu, Ile, Ala). W przypadku C-końca dominowały reszty o bardziej zróżnicowanej budowie. Poza wyżej wymienionymi, niektóre z dipeptydów zawierały C-końcowe reszty cysteiny, metioniny czy seryny. Wyniki te są zbieżne z uzyskanymi przez Udenigwe i Aluko [31], którzy zastosowali regresję wieloraką w analizie peptydów o aktywności przeciwutleniającej pozyskanych z różnych hydrolizatów białek żywności. Wymienieni autorzy wykazali, że zależnie od przeprowadzonego testu, obecność w sekwencjach peptydów określonych aminokwasów pozytywnie lub negatywnie przyczyniała się do efektu antyoksydacyjnego analizowanych peptydów, co jest wynikiem działania różnych mechanizmów tworzenia reaktywnych form tlenu [31]. Niemniej na podstawie uzyskanych wyników, dotyczących prawdopodobieństwa występowania danego aminokwasu w białku w powiązaniu z hydrofobowością oraz budową dipeptydów o aktywności przeciwutleniającej, wymienione zmienne można stosować w analizie chemometrycznej peptydów.

Poza doborem zmiennych tworzenie macierzy danych do konstruowania modeli wyjaśniających ilościową zależność między strukturą a aktywnością peptydów (ang. *quantitative structure-activity relationship*, QSAR) wymaga zastosowania miary aktywności, co nie jest wymagane do wykonania PCA [10, 11]. W celu przeprowadzenia PCA za informacją wystarczającą do tworzeniu macierzy danych przyjęto adnotację o wykazywaniu przez peptydy aktywności przeciwutleniającej. Z przeglądu literatury wynika, że do chwili obecnej nie ma znormalizowanej metody określania aktywności przeciwutleniającej związków żywności i dlatego zalecane jest, by oznaczenia prowadzić za pomocą testów bazujących na różnych mechanizmach opisujących efekt antyoksydacyjny związku [4]. Aktywność peptydu jest wyrażana w różny sposób i może pochodzić z różnych eksperymentów prowadzonych różnymi metodami, co odgrywa szczególną rolę w badaniach QSAR [21] i ma wpływ na miarodajność tworzonych modeli oraz uzyskanych wyników [24]. Niemniej tworzenie przez naukowców nowych narzędzi bio- oraz chemoinformatycznych sprzyja analizie zagadnień związanych m.in. z aktywnością biologiczną peptydów. Narzędzia te przyczyniają się do pogłębiania wiedzy na temat właściwości biologicznych oraz funkcjonalnych składników żywności, takich jak: skład aminokwasowy białka i występujących w nich peptydów, właści-



wości fizykochemiczne aminokwasów tworzących peptydy czy charakterystyka proteaz uwalniających biopeptydy [8]. Stosując wymienione narzędzia, należy mieć na uwadze, że jakość uzyskanych wyników zależy od poprawności skonstruowanego modelu, na co wpływ mają dane wejściowe [19]. Mimo wskazanych ograniczeń metod *in silico*, ich stosowanie może być przydatne w projektowaniu peptydów pochodzących z żywności, odgrywających ważną rolę w profilaktyce chorób cywilizacyjnych. Do takich peptydów można również zaliczyć sekwencje o działaniu przeciwutleniającym.

Reasumując, można powiedzieć, że mimo przedstawionych ograniczeń uzyskane wyniki z wykorzystaniem PCA pozwalają na zaobserwowanie pewnych prawidłowości między strukturą a aktywnością przeciwutleniającą badanych dipeptydów. Jak uważa Tu [29], nie istnieje idealna metoda analizy chemometrycznej związków chemicznych, ale uzyskane wyniki, jak i zastosowana metoda, mogą przyczynić się do pogłębienia wiedzy na temat badania wpływu struktury na aktywność przeciwutleniającą peptydów. Jest to zgodne z postulatami naukowców rekomendujących stosowanie różnych metod chemometrycznych do badania określonego zjawiska.

## Wnioski

1. Bez względu na lokalizację reszt w sekwencji dipeptydu (tj. N- lub C-koniec), można stwierdzić, że na ich aktywność przeciwutleniającą miały wpływ następujące właściwości aminokwasów: masa cząsteczkowa (+), liczba atomów węgla (+), polaryzowalność (+), rozmiar (+), polarność (-), powierzchnia reszt ukrytych dla rozpuszczalnika (-) czy prawdopodobieństwo występowania danego aminokwasu w białku (-). Były one skorelowane dodatnio (+) lub ujemnie (-).
2. Modelowy dipeptyd przeciwutleniający spełniający powyższe warunki powinien zawierać reszty takich aminokwasów jak: Pro, His, Leu oraz/lub Val.
3. Wykazano, że zmienne stosowane do przewidywania właściwości białek, czyli powierzchnia reszt ukrytych dla rozpuszczalnika oraz prawdopodobieństwo występowania danego aminokwasu w białku można stosować do badań zależności „struktura-aktywność przeciwutleniająca“ dipeptydów.
4. Mimo pewnych ograniczeń metod *in silico*, ich stosowanie może być przydatne w projektowaniu peptydów pochodzących z żywności, odgrywających ważną rolę w profilaktyce chorób cywilizacyjnych m.in. peptydów o działaniu przeciwutleniającym.

*Projekt finansowany w ramach programu Ministra Edukacji i Nauki pod nazwą "Regionalna Inicjatywa Doskonałości" w latach 2019-2022, nr projektu 010/RID/2018/19, kwota finansowania 12.000.000 złotych oraz funduszu Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, nr tematu: 17.610.014-110.*

*Autorzy dziękują dr inż. Monice Hryniewicz za pomoc w tworzeniu macierzy danych oraz wskazówki dotyczące interpretacji wyników*

## Literatura

- [1] Canabady-Rochelle L.L.S., Harscoat-Schiavo C., Kessler V., Aymes A., Fournier F., Girardet J.-M.: Determination of reducing power and metal chelating ability of antioxidant peptides: Revisited methods. *Food Chem.*, 2015, 183, 129-135.
- [2] Carugo O.: Amino acid composition and protein dimension. *Protein Science*, 2008, 17 (12), 2187-2191.
- [3] Deiana A., Shimizu K., Giansanti A.: Amino acid composition and thermal stability of protein structures: The free energy geography of the Protein Data Bank. [on line]. 2010. Dostęp w Internecie [6.08.2020]: <https://arxiv.org/ftp/arxiv/papers/1012/1012.5916.pdf>
- [4] Di Bernardini R., Harnedy P., Bolton D., Kerry J., O'Neill E., Mullen A.M., Hayes M.: Antioxidant and antimicrobial peptidic hydrolysates from muscle protein sources and by-products. *Food Chem.*, 2011, 124, 1296-1307.
- [5] Ellinger J.J., Chylla R.A., Ulrich E.L., Markley J.L.: Databases and software for NMR-based metabolomics. *Curr. Metabolomics*, 2013, 1, 28-40.
- [6] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.: Protein identification and analysis tools on the ExPASy server. In: *The proteomics protocols handbook*. Ed. J.M. Walker. Humana Press, Totowa 2005, pp. 571-607.
- [7] Hormoz S.: Amino acid composition of proteins reduces deleterious impact of mutations. *Sci. Rep.*, 2013, 3, #2919.
- [8] Iwaniak A., Darewicz M., Mogut D., Minkiewicz P.: Elucidation of the role of in silico methodologies in approaches to studying bioactive peptides derived from foods. *J. Funct. Foods*, 2019, 61, #103486.
- [9] Iwaniak A., Hryniewicz M., Bucholska J., Minkiewicz P., Darewicz M.: Understanding the nature of bitter-taste di- and tripeptides derived from food proteins based on chemometric analysis. *J. Food Biochem.*, 2019, 43 (1), #e12500.
- [10] Iwaniak A., Hryniewicz M., Bucholska J., Darewicz M., Minkiewicz P.: Structural characteristics of food protein-originating di- and tripeptides using principal component analysis. *Eur. Food Res. Technol.*, 2018, 244 (10), 1751-1758.
- [11] Iwaniak A., Minkiewicz P., Darewicz M., Protasiewicz M., Mogut D.: Chemometrics and cheminformatics in the analysis of biologically active peptides from food sources. *J. Funct. Foods*, 2015, 16, 334-351.
- [12] Iwaniak A., Dziuba J.: Bioaktywne sekwencje w białkach żywności. W: *Biologicznie aktywne peptydy i białka żywności*. Red. J. Dziuba, Ł. Fornal. WNT, Warszawa 2009, ss. 176-270.
- [13] Jolliffe I.T.: *Principal Component Analysis*. 2<sup>nd</sup> ed. Springer, New York City 2002.
- [14] Kajander T., Kahn P.C., Passila S.H., Cohen D.C., Lehtio L., Adolfsen W., Warwicker J., Schell U., Goldman A.: Buried charges surface in a protein. *Structure*, 2000, 8 (11), 1203-1214.
- [15] Kawashima S., Ogata H., Kanehisa M.: AAindex: Amino acid index database. *Nucleic Acids Res.*, 1999, 27, 368-369.
- [16] Kęska P., Stadnik J.: Structure - activity relationships study on biological activity of peptides as dipeptidyl peptidase IV inhibitors by chemometric modeling. *Chem. Biol. Drug.*, 2020, 95, 291- 301.
- [17] Kleekayai T., Harnedy P.A., O'Keeffe M.B., Poyarkov A.A., CunhaNeves A., Suntornsuk W., Fitz-Gerald R.J.: Extraction of antioxidant and ACE inhibitory peptides from Thai traditional fermented shrimp pastes. *Food Chem.*, 2015, 176, 441-447.

- [18] Kostyra H., Kostyra E.: Bioaktywne peptydy uwalniane z białek żywności. W: Żywność prozdrowotna. Wyd. Uniwersytetu Przyrodniczego w Poznaniu, Poznań 2014, ss. 235-249.
- [19] Li-Chan E.C.Y.: Bioactive peptides and protein hydrolysates: Research trends and challenges for application as nutraceuticals and functional food ingredients. *Curr. Opin. Food Sci.*, 2015, 1, 28-37.
- [20] Minkiewicz P., Iwaniak A., Darewicz M.: BIOPEP-UWM database of bioactive peptides: Current opportunities. *Int. J. Mol. Sci.*, 2019, 20 (23), #5978.
- [21] Mooney C., Haslam N.J., Pollastri G., Shields D.C.: Towards the improved discovery and design of functional peptides: Common features of diverse classes permit generalized prediction of bioactivity. *PLoS One*, 2012, 7 (10), #e45012.
- [22] Nongonierma A.B., FitzGerald R.J.: Dipeptidyl peptidase IV inhibitory and antioxidative properties of milk protein-derived dipeptides and hydrolysates. *Peptides*, 2013, 39, 157-163.
- [23] Pace C.N., Scholtz J.M.: A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.*, 1998, 75 (1), 422-427.
- [24] Pripp A.H., Isaksson T., Stepaniak L., Sørhaug T., Ardö Y.: Quantitative structure activity relationship modelling of peptides and proteins as a tool in food science. *J. Agric. Food Chem.*, 2005, 16 (11), 484-494.
- [25] Sarmadi B.H., Ismail A.: Antioxidative peptides from food proteins: A review. *Peptides*, 2010, 31, 1949-1956.
- [26] Shaukat S.S., Rao T.A., Khan M.A.: Impact of sample size on principal component analysis ordination of an environmental data set: Effects on eigenstructure. *Ekologia (Bratislava)*, 2016, 35 (2), 173-190.
- [27] Shu M., Mei H., Liao L., Li Z.: Structural parameter characterization and bioactivity simulation based on peptide sequence. *QSAR Combinatorial Sci.*, 2009, 28 (1), 27-35.
- [28] Stanisław A.: Analiza składowych głównych. Analiza czynnikowa. Przystępny kurs statystyki z zastosowaniem Statistica PL na przykładach z medycyny. Analizy wielowymiarowe. T. 3. StatSoft Polska, Kraków 2007, ss. 165-267.
- [29] Tu J.V.: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medicinal outcomes. *J. Clin. Epidem.*, 1996, 49, 1225-1231.
- [30] Udenigwe C.C.: Bioinformatic approaches, prospects and challenges of food bioactive peptide research. *Trends Food Sci. Technol.*, 2014, 36, 137-143.
- [31] Udenigwe C.C., Aluko R.: Chemometric analysis of the amino acid requirements of antioxidant food hydrolysates. *Int. J. Mol. Sci.*, 2011, 12, 3148-3161.
- [32] Zou T.-B., He T.-P., Li H.-B., Tang H.-W., Xia E.-Q.: The structure-activity of the antioxidant peptides from natural proteins. *Molecules*, 2016, 21(1), #72.

**APPLICATION OF PRINCIPAL COMPONENT ANALYSIS TO STUDY RELATIONSHIP  
BETWEEN STRUCTURE AND ANTIOXIDATIVE ACTIVITY OF DIPEPTIDES DERIVED  
FROM FOOD PROTEINS**

S u m m a r y

In the research study, there was applied a statistical model developed with the use of Principal Component Analysis (PCA) for the purpose of determining the impact of dipeptides' structure on their antioxidant bioactivity. Sequences of 47 peptides were taken from a BIOPEP-UWM database of proteins and bioactive peptides (<https://biochemia.uwm.edu.pl/biopep-uwm/>). The selected descriptors (also called attributes) to describe the physicochemical properties of amino acids present in a dipeptide chain were

numerically expressed and implemented from open-access computer software or databases such as: Peptide Property Calculator, Biological Magnetic Resonance Data Bank, ProtScale, Molar Polarizability Values and ImMunoGeneTics. PCA was carried out using a Statistica software. The number of principal components that were significant for the interpretation of the impact of dipeptides' structure on their antioxidant activity was determined based on the percentage of cumulative variance. It was 79.9 % and it was in line with the four components. The first and the fourth component determined the impact of N-terminal amino acid residue on the antioxidant activity of dipeptide, whereas the second and the third one referred to the effect of C-terminal amino acid residue. By means of PCA it was shown that a model dipeptide possessing antioxidant activity should be characterised by the presence of N-terminal amino acid with a cyclic or aromatic ring or amino acid residue with a non-polar side chain. The C-end of peptide should be composed of proline, histidine, leucine or valine. The results obtained are consistent with the reference literature data that refer to the research on relationship between structure and activity of antioxidative peptides, assessed using a multivariate regression analysis. Based on the research study performed, it was found that the chemometric method applied might be useful when designing food-derived peptides, including those possessing the activity studied.

**Key words:** antioxidant activity, dipeptides, BIOPEP-UWM database, chemometrics, PCA 