

УДК 577.322

ТЕОРЕТИЧНИЙ АНАЛІЗ АМІЛОІДОГЕННОГО ПОТЕНЦІАЛУ ЛІЗОЦИМУ, ЦИТОХРОМУ С ТА АПОЛІПОПРОТЕЇНУ А-I

В.М. Трусова

*Харківський національний університет імені В.Н. Каразіна, пл. Свободи, 4, Харків 61022, Україна
e-mail: valeriyam.trusova@gmail.com*

Надійшла до редакції 29 вересня 2016 року

Прийнята 17 жовтня 2016 року

З використанням 8 веб-алгоритмів, що включали Pasta2, AmylPred2, Tango, MetAmyl, Waltz, Aggrescan, BetaScan та FoldAmyloid, було проведено теоретичний аналіз амінокислотних послідовностей лізоциму, цитохрому *c* та N-термінального фрагменту аполіпопротеїну А-I та ідентифіковано їх амілоїдогенні фрагменти. Фрагмент вважався амілоїдогенним, якщо він був ідентифікований щонайменше чотирма алгоритмами. Порівняльний аналіз амілоїдогенних фрагментів нативного та мутантного білків, проведений з використанням різних методів прогнозування, виявив, що усі мутанти характеризуються тими ж амілоїдогенними ділянками, що й нативний білок, однак для мутантних білків агрегаційний потенціал кожної ділянки вище. Показано, що агрегаційні сегменти усіх білків характеризуються збагаченістю гідрофобними алифатичними (Ile, Val, Leu, Ala) та ароматичними (Trp, Phe, Tyr) амінокислотними залишками. Висловлено припущення щодо ключової ролі гідрофобних взаємодій у процесі агрегації білків.

КЛЮЧОВІ СЛОВА: амілоїдні фібрили, агрегаційний потенціал, комп'ютерне моделювання, лізоцим, цитохром *c*, аполіпопротеїн А-I.

THEORETICAL ANALYSIS OF AMYLOIDOGENIC POTENTIAL OF LYSOZYME, CYTOCHROME C AND APOLIPOPROTEIN A-I

V.M. Trusova

V.N. Karazin Kharkiv National University, Svobody Sq., 4, Kharkiv 61022, Ukraine

Using 8 web-algorithms, including Pasta2, AmylPred2, Tango, MetAmyl, Waltz, Aggrescan, BetaScan та FoldAmyloid, theoretical analysis of amino acid sequences of lysozyme, cytochrome *c* and N-terminal fragment of apolipoprotein A-I has been carried out, and amyloidogenic fragments of the proteins have been identified. The fragment was identified as amyloidogenic if it was determined by at least four algorithms. Comparative analysis of aggregation-prone regions of native and mutant proteins showed that that all mutants are characterized by same amyloidogenic segments as native proteins with the amyloidogenic potential being more pronounced for mutated proteins. It was shown that aggregation-prone regions of all proteins analyzed here, were rich in hydrophobic aliphatic (Ile, Val, Leu, Ala) and aromatic (Trp, Phe, Tyr) amino acid residues. Hydrophobic interactions were supposed to play key role in protein aggregation process.

KEY WORDS: amyloid fibrils, aggregation potential, computer modelling, lysozyme, cytochrome *c*, apolipoprotein A-I.

ТЕОРЕТИЧЕСКИЙ АНАЛИЗ АМИЛОИДОГЕННОГО ПОТЕНЦИАЛА ЛИЗОЦИМА, ЦИТОХРОМА С И АПОЛИПОПРОТЕИНА А-I

В.М. Трусова

Харьковский национальный университет имени В.Н. Каразина, пл. Свободы, 4, Харьков 61022, Украина

С использованием 8 веб-алгоритмов, которые включали Pasta2, AmylPred2, Tango, MetAmyl, Waltz, Aggrescan, BetaScan та FoldAmyloid, был проведен теоретический анализ аминокислотных последовательностей лизоцима, цитохрома *c* и N-терминального фрагмента аполіпопротеина А-I, а также идентифицированы их амилоидогенные фрагменты. Сравнительный анализ амилоидогенных фрагментов нативного и мутантного белков, проведенный с использованием различных методов прогнозирования, показал, что все мутанты характеризуются теми же амилоидогенными областями, что и нативный белок, однако для мутантных белков агрегационный потенциал выше. Показано, что агрегационные сегменты всех белков характеризуются насыщенностью гидрофобными алифатическими (Ile, Val, Leu, Ala) и ароматическими (Trp, Phe, Tyr) аминокислотными остатками. Высказано предположение относительно ключевой роли гидрофобных взаимодействий в процессе агрегации белков.

КЛЮЧЕВЫЕ СЛОВА: амилоидные фибриллы, агрегационный потенциал, компьютерное моделирование, лизоцим, цитохром *c*, аполіпопротеин А-I.

Дослідження молекулярних механізмів агрегації біологічних макромолекул, зокрема білків, є одним із фундаментальних напрямків сучасної біофізики [1,2]. Міжмолекулярні нековалентні взаємодії між білковими молекулами призводять до формування широкого спектру надмолекулярних ансамблів, структура яких коливається від невпорядкованих аморфних агрегатів до кристалів із чітко визначеною трансляційною симетрією у трьох напрямках. В останні десятиріччя невпинно зростаюча увага вчених прикута до дослідження процесу одномірної кристалізації білків та пептидів з утворенням високовпорядкованих фібрилярних структур, які зветься амілоїдами [3]. Ці агрегати являють собою лінійні β -складчасті ансамблі мономерних субодиниць із трансляційною симетрією у повздожньому напрямку, структура яких стабілізується, головним чином, мережею водневих зв'язків [4]. Згідно з сучасними уявленнями, акумуляція фібрилярних агрегатів в тканинах організму є ключовим фактором в етіології цілої низки так званих конформаційних хвороб, включаючи хвороби Альцгеймера, Паркінсона, Хантінгтона, діабет другого типу, ревматоїдний артрит, губчасті енцефалопатії (пріонні хвороби), тощо [5]. Численні дослідження свідчать про те, що здатність утворювати амілоїдні фібрили не обмежується амілоїдогенними білками, які пов'язані із конформаційними хворобами, а є фундаментальною властивістю різних структурно незалежних білків [1, 3]. Все більшого теоретичного та експериментального обґрунтування набуває ідея про те, що здатність білка до фібрилізації закодована у його амінокислотній послідовності. Рядом авторів була висунута гіпотеза, згідно з якою за утворення амілоїдних фібрил відповідають короткі фрагменти амінокислотної послідовності поліпептидів [6]. Однак, у більшості випадків експериментальне визначення амілоїдогенних фрагментів білків такими методами, як кристалографія чи ЯМР-спектроскопія, утруднюється внаслідок низької розчинності фібрил. З огляду на це, для ідентифікації ділянок з високим агрегаційним потенціалом та з'ясування загальних принципів і закономірностей патологічної агрегації білків наразі широко застосовуються комп'ютерні методи прогнозування амілоїдогенного потенціалу поліпептидного ланцюга на основі його амінокислотної послідовності. Основною метою даної роботи була ідентифікація агрегаційних ділянок лізоциму, цитохрому *c* та N-термінального фрагменту аполіпопротеїну А-I (апоА 83) з використанням різних методів прогнозування. Варто відзначити, що список розроблених до теперішнього часу алгоритмів не обмежується лише тими методами, які будуть описані нижче. При розв'язанні поставлених задач ми використовували найпоширеніші алгоритми, які відрізняються за принципом імплементації.

МАТЕРІАЛИ І МЕТОДИ

Для ідентифікації амілоїдогенних сегментів лізоциму, цитохрому *c* та N-кінцевого фрагменту аполіпопротеїну А-I були використані алгоритми, що включали Pasta2, AmylPred2, Tango, MetAmyl, Waltz, Aggrescan, Betascan та FoldAmyloid. При аналізі припускали, що фрагмент є агрегаційним, якщо він ідентифікований як мінімум чотирма методами прогнозування.

Один з широко застосовуваних алгоритмів, що має назву FoldAmyloid, був запропонований групою Галзитської та співавторів [7]. Цей метод базується на принципі ентальпійно-ентропійної компенсації, який постулює, що для компенсації втрати конформаційної ентропії при переході білка у більш впорядковану конформацію необхідна достатня кількість контактів між амінокислотними залишками, які визначають ентальпійний внесок у вільну енергію стабілізації білка. З огляду на те, що ентальпія залежить від енергії короткодійних взаємодій, було зроблено припущення,

що у випадку, коли очікувана середня густина пакування, яка визначає усереднену кількість контактів між залишками на окремій ділянці, нижче граничної, тобто густини пакування для глобулярних білків, білок залишиться незгорнутим. На противагу, якщо очікувана середня густина пакування перевищує граничну, то ймовірність утворення фібрилярних агрегатів досить висока. Фактично, висока термостабільність амілоїдних фібрил, їх стійкість до дії протеаз та насиченість β -структурами і передбачає той факт, що ці самоасоціати містять ділянки з високою густиною пакування. Результати верифікації описаної гіпотези показали, що здатність поліпептидного ланцюга утворювати амілоїди дійсно корелює із числом ділянок з високою густиною пакування.

Алгоритм Tango прогнозує ймовірність утворення тієї чи іншої вторинної структури [8]. В основі методу лежить статистичний аналіз емпіричних властивостей амінокислот з метою розрахунку відносної ймовірності згортання окремих ділянок поліпептидного ланцюга в одну з наступних конформацій: α -спіраль, β -петля, β -лист та невідповідна структура.

Метод Waltz використовує базу даних пептидів з експериментально визначеними амілоїдогенними властивостями для ідентифікації агрегаційних ділянок білків [9]. База даних включає пептиди, що складаються з шести амінокислотних залишків. При цьому, як мінімум три залишки утворюють амілоїдогенну ділянку. Автори використовують так звану позиційно-специфічну оцінювальну матрицю, яка віддзеркалює ймовірність того, що певна амінокислота буде розташовуватись в амілоїдогенному гексапептиді.

Серед розмаїття методів прогнозування особливе місце посідають AmylPred2 та MetAmyl, мета-сервери для консенсус-аналізу, які комбінують декілька методів аналізу, включаючи FoldAmyloid, Tango та Waltz [10, 11]. На основі коефіцієнту кореляції Метьюса ці алгоритми пропонують усереднені результати прогнозування амілоїдних фрагментів білка.

На основі функції розподілу Больцмана алгоритм Pasta2 визначає, які фрагменти білка стабілізуються крос- β -структурою [12]. Даний аналіз базується на прогнозуванні здатності двох залишків розташовуватись у сусідніх β -тяжах у паралельній чи антипаралельній орієнтації. Далі на основі регресійного аналізу передбачається переважна локалізація агрегаційних фрагментів та їх відносна орієнтація.

Принцип алгоритму BetaScan подібний до Pasta2, однак BetaScan розглядає лише паралельну орієнтацію β -тяжів з огляду на те, що ця орієнтація є більш поширеною у порівнянні із антипаралельною [13].

Сервер Aggrescan базується на припущенні, що короткі фрагменти (5-7 залишків) можуть ініціювати агрегацію в пептидах та білках, та що амілоїдогенний потенціал цих фрагментів залежить від їх амінокислотної послідовності [14]. Шкала агрегаційної здатності амінокислот була розрахована на основі експериментальних даних. Програма аналізує короткі фрагменти досліджуваного білка (5 чи 7 залишків) та визначає середній амілоїдогенний потенціал кожного фрагменту на основі цієї шкали. Фрагменти з найвищим середнім потенціалом вважаються агрегаційними ділянками білка.

РЕЗУЛЬТАТИ Й ОБГОВОРЕННЯ

Лізоцим з яєчного білка (UniProt ID P00698) містить 129 амінокислотних залишків. У базі даних кристалічних структур білків (PDB) для аналізу було обрано структуру 1AKI. Як видно з табл. 1, білок характеризується п'ятьма амілоїдогенними ділянками, які охоплюють залишки 27-31 (NWVCA), 53-60 (YGILQINS), 81-90 (SALLSSDITA), 106-110 (NAWVA) та 121-125 (QAWIR). При цьому, найвищим агрегаційним потенціалом характеризуються фрагменти 53-60 та 81-90.

Лізоцим людини (UniProt ID P61626) схожий за своїми властивостями та первинною структурою із лізоцимом яєчного білка. Лізоцим людини містить 130 амінокислотних залишків (для аналізу використано структуру 1REX) та згідно з результатами прогнозування з використанням різних комп'ютерних алгоритмів, має три амілоїдогенні фрагменти: залишки 25-33 (LANWMCLAK), 55-62 (GIFQINSR) та 107-112 (RAWVAW). Агрегаційний потенціал збільшується у порядку (25-33) < (107-112) < (55-62). Варто відзначити, що фрагменти з найвищою агрегаційною здатністю аналогічні для людського та курячого аналогу – вони знаходяться майже в однаковому місці в поліпептидному ланцюзі (55-62 та 53-60 для першого та другого випадку, відповідно) та мають схожу амінокислотну послідовність.

Таблиця 1

Агрегаційні ділянки лізоциму, цитохрому *c* та аполіпопротеїну А-I, прогнозовані різними алгоритмами.

Білок/пептид	Амілоїдогенний фрагмент
Лізоцим з яєчного білка	²⁷ NWVCA ₃₁ (5 алг.)*, ⁵³ YGILQINS ₆₀ (4 алг.), ⁸¹ SALLSSDITA ₉₀ (4 алг.), ¹⁰⁶ NAWVA ₁₁₀ (5 алг.), ¹²¹ QAWIR ₁₂₅ (4 алг.)
Лізоцим людини	²⁵ LANWMCLA ₃₂ (4 алг.), ⁵⁵ GIFQINSR ₆₂ (4 алг.), ¹⁰⁷ RAWVAW ₁₁₂ (5 алг.)
Цитохром <i>c</i>	⁹ IFVQK ₁₃ (7 алг.), ⁸⁰ MIFAG ₈₄ (5 алг.)
АпоА 83	¹⁴ LATVYVDVL ₂₂ (8 алг.), ⁵³ VTSTFS ₅₈ (5 алг.)

* У дужках зазначено кількість алгоритмів, які ідентифікували певний фрагмент як амілоїдогенний.

При прогнозуванні амілоїдогенних фрагментів білків має сенс проаналізувати не тільки нативний білок, але і його мутанти, які пов'язані із розвитком конформаційних хвороб. Так, зокрема, відомо, що *in vivo* лізоцим утворює амілоїдні фібрили при мутаціях I56T, F57I, W64R та D67H, F57I/T70N та W112R/T70N [15]. Тому у даному контексті уявлялось доцільним не тільки ідентифікувати амілоїдогенні фрагменти нативного білка, але й порівняти агрегаційні потенціали білка дикого типу та мутантів. Аналіз, проведений з використанням різних методів прогнозування виявив, що усі мутанти характеризуються тими ж амілоїдогенними ділянками, що й нативний лізоцим людини, однак для мутантних білків агрегаційний потенціал кожної ділянки вище у порівнянні з нативним білком (табл. 2).

Таблиця 2

Агрегаційна здатність нативного лізоциму людини та його амілоїдогенних мутантів за даними прогнозування серверу Pasta2

Номер амінокислотного залишку	Нативний	I56T	F57I	W64R	D67H	F57I/T70N	W112R/T70N
25-33	0,25	0,31	0,26	0,37	0,44	0,46	0,36
55-62	0,66	0,69	0,72	0,74	0,73	0,74	0,7
107-112	0,42	0,47	0,5	0,52	0,57	0,61	0,57

На відміну від лізоциму, у молекулах апоА 83 людини (UniProt ID P02647; PDB ID 2A01) та цитохрому *c* коня (UniProt ID P00004; PDB ID 1HRC) були ідентифіковані два амілоїдогенних сегмента, локалізовані на залишках 14-22 (LATVYVDVL), 52-58 (VTSTFS) (апоА 83) та 9-13 (IFVQK), 80-84 (MIFAG) (цитохром *c*) (табл. 1).

Важливо, що ділянка 14-22 апоА 83 була ідентифікована усіма алгоритмами прогнозування, а також підтверджена експериментальними дослідженнями [16]. Необхідно також зазначити, що аналогічно лізоциму, введення точкової амілоїдогенної мутації G26R у амінокислотну послідовність апоА 83 не призвело до змін агрегаційних фрагментів білка, однак майже у 1,5 рази підвищило його амілоїдогенну здатність.

Згідно з сучасними фізичними теоріями, ключова роль у забезпеченні стабільності фібрилярних структур належить гідрофобності поліпептидного ланцюга [17]. Аналогічно фолдінгу глобулярних білків, самоасоціація білкових мономерів в амілоїдні структури ініціюється гідрофобним ефектом. Конформаційний перехід поліпептидного ланцюга у стан розплавленої глобули при зміні концентрації білка, температури, рН чи іонної сили розчину є початком процесу нуклеації. Мінімізація числа термодинамічно не вигідних контактів між неполярними залишками, які експонуються у ході розгортання білкової глобули, та розчинником призводить до утворення олігомерних інтермедіатів. Зрілі фібрили поряд з гідрофобними взаємодіями стабілізуються також мережею водневих зв'язків. Мережа водневих зв'язків виступає у ролі остову, який утримує усю архітектуру молекули. Численні експериментальні дослідження підтверджують ідею про те, що зростання гідрофобності поліпептидного ланцюга прискорює швидкість фібрилізації та підвищує термодинамічну стабільність агрегатів [1]. Окрім цього, окремий клас гідрофобних взаємодій – ароматичні взаємодії – також відіграють суттєву роль у стабілізації фібрил [18]. У той час як Марек зі співавторами [19] припустили, що π - π взаємодії кінетично важливі лише у ході лаг-фази, Газіт зробив висновок, що роль ароматичних контактів включає два аспекти [20]: і) вони роблять позитивний енергетичний внесок за рахунок π - π стекингу, та ii) їх специфічний орієнтуючий ефект лежить в основі енергетично оптимального пакування β -тяжів у структурі фібрил. Ці положення збігаються з результатами аналізу лізоциму, апоА 83 та цитохрому *c* (табл. 1). Відмінними рисами агрегаційних сегментів усіх досліджених білків є збагаченість гідрофобними аліфатичними (Ile, Val, Leu, Ala) та ароматичними (Trp, Phe, Tyr) амінокислотними залишками.

ВИСНОВКИ

За допомогою алгоритмів FoldAmyloid, Tango, Waltz, AmylPred2, MetAmyl, Pasta2, BetaScan та Aggrescan ідентифіковані амілоїдогенні фрагменти лізоциму, цитохрому *c* та N-термінального фрагменту аполіпопротеїну А-I. Проведено порівняльний аналіз агрегаційних властивостей білків дикого типу та мутантів лізоциму людини та апоА-I. Продемонстровано, що амілоїдогенні мутації лізоциму та аполіпопротеїну А-I не змінюють положення агрегаційно-компетентних ділянок, але підвищують їх агрегаційний потенціал.

СПИСОК ЛІТЕРАТУРИ

1. Buell K. The physical chemistry of the amyloid phenomenon: thermodynamics and kinetics of filamentous protein aggregation / A. Buell, C. Dobson, T. Knowles // *Essays Biochem.* – 2014. – V. 56. – P. 11 – 39.
2. Volpatti L. Polymer physics inspired approaches for the study of the mechanical properties of amyloid fibrils / L. Volpatti, T. Knowles // *J. Polymer Sci. A.* – 2014. – V. 52. – P. 281-292.
3. Adamcik J. Protein fibrils from a polymer physics perspective / J. Adamcik, R. Mezzenga // *Macromolecules.* – 2012. – V. 45. – P. 1137-1150.
4. Tycko R. Solid state NMR studies of amyloid fibril structure / R. Tycko // *Annu. Rev. Phys. Chem.* – 2011. – V. 62. – P. 279-299.
5. Common core structure of amyloid fibrils by synchrotron X-ray diffraction / M. Sunde, L.C. Serpell, M. Bartlam et al. // *J. Mol. Biol.* – 1997. – V. 273. – P. 729-739.
6. Impact of sequence on the molecular assembly of short amyloid peptides / V. Wagoner, M. Cheon, I. Chang et al. // *Proteins.* – 2014. – V. 82. – P. 1469-1483.

7. Garbuzynskiy S. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence / S. Garbuzhynskiy, M. Lobanov, O. Galzitskaya // *Bioinformatics*. – 2010. – V. 26. – P. 326-332.
8. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins / A. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz et al. // *Nat. Biotechnol.* – 2004. – V. 22. – P. 1302-1306.
9. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices / S. Maurer-Stroh, M. Debulpaep, N. Kuemmerer et al. // *Nat. Methods*. – 2010. – V. 7. – P. 237-242.
10. A consensus method for the prediction of “aggregation-prone” peptides in globular proteins / A. Tsolis, N. Papandreou, V. Iconomidou et al. // *PLoS One*. – 2013. – V. 8. – P. e54175.
11. Emily M. MetAmyl: a META-predictor of AMYLoid proteins / M. Emily, A. Talvas, C. Delamarche // *PLoS One*. – 2013. – V. 8. – P. e79722.
12. PASTA2: an improved server for protein aggregation prediction / I. Walsh, F. Seno, S. Tosatto et al. // *Nucl. Acids Res.* – 2014. – V. 42. – P. W301-307.
13. BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis / A. Bryan, M. Menke, L. Cowen et al. // *PLoS Comput. Biol.* – 2009. – V. 5. – P. 1000333.
14. AGGRESKAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides // O. Conchillo-Sole, N. de Groot, F. Aviles et al. // *BMC Bioinformatics*. – 2007. – V. 8. – P. 65.
15. Identification of the core structure of lysozyme amyloid fibrils by proteolysis / E. Frare, M. Mossuto, P. Polverino de Laureto et al. // *J. Mol. Biol.* – 2006. – V. 361. – P. 551-561.
16. Das M. Amyloid-forming properties of human apolipoproteins: sequence analysis and structural insights / M. Das, O. Gursky // *Springer Subcellular Biochemistry Series: Lipids in Protein Misfolding* – 2015. – V. 855. – P. 175-211.
17. Relative influence of hydrophobicity and net charge in the aggregation of two homologous proteins / M. Calamai, N. Taddei, M. Stefani et al. // *Biochemistry*. – 2003. – V. 42. – P. 15078-15083.
18. Tu L. Role of aromatic interactions in amyloid formation by islet amyloid polypeptide / L. Tu, D. Raleigh // *Biochemistry*. – 2013. – V. 52. – P. 333-342.
19. Aromatic interactions are not required for amyloid fibril formation by islet amyloid polypeptide but do influence the rate of fibril formation and fibril morphology / P. Marek, A. Abedini, B. Song et al. // *Biochemistry*. – 2007. – V. 46. – P. 3255-3261.
20. Gazit E. Self-assembly of short aromatic peptides into amyloid fibrils and related nanostructures / E. Gazit // *Prion*. – 2007. – V. 1. – P. 32-35.