



<http://dx.doi.org/10.15407/dopovidi2016.06.103>

УДК 577.322.4:5:7

Т. С. Лиманська¹, О. Ю. Нипорко¹,
член-кореспондент НАН України **О. І. Корнелюк^{2,1}**

¹Інститут високих технологій, Київський національний університет ім. Тараса Шевченка

²Інститут молекулярної біології і генетики НАН України, Київ

E-mail: dfnalex@gmail.com

Аналіз неупорядкованих ділянок білка АІМР1/р43 мультисинтетазного комплексу людини методами біоінформатики

*Особливості вторинної структури та розташування неупорядкованих ділянок у структурі білка АІМР1 — компонента мультисинтетазного комплексу людини — досліджено за допомогою ряду *in silico* підходів. Виявлено, що найдовша ділянка білка, яка тягнє до формування неупорядкованої структури, розташовується з 103 по 148 амінокислотний залишок. Цікаво, що в межах цієї ділянки (приблизно з 121 по 140 залишок) також можлива наявність α -спіралі. Це удаване протиріччя досить просто пояснюється тим, що відповідна ділянка насичена як залишками лізину (маркер неупорядкованих ділянок білка), так і залишками глутамату (типовий маркер саме α -спіральних елементів). З великою часткою ймовірності можна припустити, що ця спіраль є метастабільною, тобто такою, що переходить до неупорядкованого стану і назад внаслідок природних флуктуацій молекули білка.*

Ключові слова: АІМР1/р43, вторинна структура, неупорядковані ділянки, біоінформатика.

Протягом тривалого часу в молекулярній біології панувала думка, що білки виконують свої природні функції, лише набуваючи певної стабільної просторової структури. Проте впродовж останніх років з'явилося багато відомостей, що ставлять під сумнів таку точку зору. Зокрема, відкрито білки, які не мають стабільної третинної структури в умовах живої клітини і знаходяться в розчині в стані, що традиційно вважається денатурованим, не втрачаючи при цьому своїх функцій. Такі природно неупорядковані білки (intrinsically disordered

© Т. С. Лиманська, О. Ю. Нипорко, О. І. Корнелюк, 2016

proteins)¹ експериментально виявлено методами мультимірної ЯМР спектроскопії при дослідженні просторових структур білків у розчині [1]. Виявилось також, що близько 70% усіх еукаріотичних білків мають неупорядковані ділянки [1]. Важливо зазначити, що неупорядковані ділянки часто є висококонсервативними за своїм амінокислотним складом і характеризуються певними патернами первинної структури. Отже, наявність таких ділянок у складі білків може бути передбачена за допомогою методів класичної біоінформатики [2].

Білок АІМР1/р43 — некаталітичний компонент високомолекулярного мультиаміноацил-тРНК-синтетазного комплексу — у нативному стані виявляє різноманітну поза- та внутрішньоклітинну активність, яка забезпечується різними ділянками його поверхні. Білок є гомодимером, містить 312 амінокислотних залишків (а. з.) та зв'язується з аргініл-тРНК синтетазою (ArgPC). Його С-кінцевий домен (166 а. з.) містить олігонуклеотидзв'язувальний мотив, який взаємодіє з тРНК. Роль N-кінцевого (146 а. з.) домену остаточно не з'ясована, але доведено, що він залучений до специфічної білок-білкової взаємодії з ArgPC і до утворення димеру АІМР1 [3].

АІМР1 виконує певні функції в регуляції клітинного метаболізму, зокрема бере участь у запальних процесах, гомеостазі глюкози та ін. АІМР1 є процітокіном — відомий цитокін ЕМАР II є С-кінцевим доменом АІМР1 (ділянка 147–312 а. з.) [4]. N-кінцева частина АІМР1 (ділянка 1–146 а. з.) необхідна для асоціації з мультисинтетазним комплексом. Цей домен АІМР1 не є гомологічним до жодного з відомих білків. Він взаємодіє з N-кінцевим доменом ArgPC людини. Взаємодія з АІМР1 стимулює аміноацилювальну активність ArgPC. Поодинокі N- і С-кінцевий домени АІМР1 такої активності не виявляють [3, 4].

Інформація стосовно просторової організації АІМР1 є неповною. Незважаючи на численні спроби протягом 20 останніх років, отримати кристали повнорозмірного АІМР1 не вдалося, що є непрямим, але досить вагомим свідченням на користь наявності в структурі цього білка великих неупорядкованих ділянок. На сьогодні є дані щодо просторової організації ЕМАР II, структура якого отримана на основі даних рентгенівської кристалографії і депонована в ряді статей Міжнародного банку білкових структур Protein Data Bank [5], зокрема 1E7Z, 1EUI та 1FL0 з розділенням 2,05, 1,80 та 1,50 Å відповідно [4]. Крім того, нещодавно була отримана низькороздільна структура комплексу глутамінової та аргінінової аміноацил-тРНК синтетаз з N-кінцевою ділянкою АІМР1 (5–80 а. з., розділення 4,05 Å, код доступу 4R3Z) [6].

Таким чином, просторова організація АІМР1 на ділянці 81–146 залишається нез'ясованою. Зважаючи на істотну збагаченість цієї ділянки, особливо в межах Lys121–Lys137, залишками лізину, можна висунути ґрунтовне припущення стосовно природної неупорядкованості саме цього регіону АІМР1. Цілковімовірно, що (потенційна) насиченість АІМР1 неупорядкованими ділянками має важливе функціональне значення для забезпечення як канонічних, так і неканонічних властивостей цього білка.

Матеріали і методи. Амінокислотну послідовність АІМР1 для аналізу структури було отримано з банку даних амінокислотних послідовностей UniProt [7], код доступу Q12904.

Вторинну структуру ділянок АІМР1 з 5 по 80 (4R3Z) та з 147 по 312 (1FL0) амінокислотні залишки, які депоновані в Міжнародному банку білкових структур (PDB), визначено згідно із записами HELIX і SHEET у відповідних координатних файлах з використанням програмного забезпечення Accelrys DS Vizualizer 4.0. Передбачення вторинної структури за

¹ До 2007 р. разом з терміном “природно неупорядковані білки” також використовувався термін “природно неструктуровані білки” (intrinsically unstructured proteins).

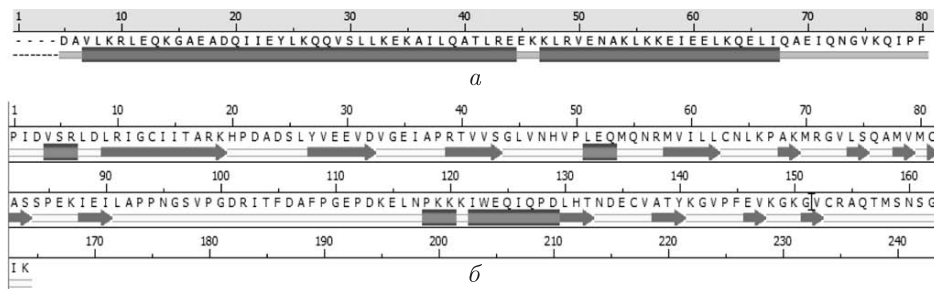


Рис. 1. Діаграма вторинної структури AIMP1 людини за результатами рентгеноструктурного аналізу: *a* — N-кінцевий домен; *б* — C-кінцевий домен. Нумерація амінокислот C-кінцевого домену відповідає нумерації ЕМАР II

амінокислотною послідовністю здійснено з використанням мультипрограмного веб-сервісу NPS@, а також веб-сервісів PSIPRED [8] та PredictProtein [9].

Для передбачення неструктурованих ділянок у послідовності AIMP1 використано 10 спеціалізованих біоінформаційних веб-сервісів, які застосовують різні алгоритми передбачення, зокрема PrDOS [10], GlobPlot [11], FoldIndex [12], DisEMBL1.5 [13], DISOPRED2 [14], DisPROT [15], IUPred, RONN2. За допомогою деяких з наведених сервісів можна проводити передбачення за кількома різними алгоритмами.

Аналіз із застосуванням веб-інструменту DisProt проводили методами VL2, VL3, VL3H та VL3P, сервісу DisEMBL — за визначеннями Remark-465, Hot-loops та Loops/coils, а сервісу GlobPlot 2.1 — Russell/Linding. Використовуючи веб-інструменти DisEMBL і GlobPlot2, можна порівнювати задану послідовність з базою даних, у якій знаходяться зразки різних типів неупорядкованих білків. База даних DisProt зберігає інформацію про білки, що в нативному стані є неупорядкованими на рівні цілої молекули або містять ділянки з неупорядкованою структурою.

Для візуалізації даних використано програмне забезпечення Accelrys DS Vizualizer версій 2.0 та 4.0, а також засоби візуалізації, імplementовані у відповідних веб-сервісах передбачення вторинної структури та неупорядкованих ділянок.

Результати та їх обговорення. Розподіл елементів вторинної структури в нативній структурі N-кінцевого та C-кінцевого (білок ЕМАР II) доменів AIMP1 за даними рентгеноструктурного аналізу наведено на рис. 1. Передбачення вторинної структури AIMP1 з використанням алгоритмів, реалізованих у веб-сервісів NSP@, PSIPRED та PredictProtein, наведені на рис. 2. За даними більшості алгоритмів, так само як і за даними з координатних файлів, депонованих у Міжнародному банку білкових структур PDB, C-термінальний домен AIMP1 (білок ЕМАР II) має α -спіралі в ділянці Leu51–Gln53 (відповідно ділянка 199–201 AIMP1), Ile122–Asp139 (ділянка 270–288 AIMP1). α -Спіраль у ділянці Pro118–Lys120 (ділянка 266–268 AIMP1) присутня у кристалографічній структурі ЕМАР II (згідно з координатними файлами), але не передбачена жодним з наведених алгоритмів. β -Елементи (β -смуги) згідно з координатними файлами та за передбаченнями більшості веб-сервісів знаходяться в межах ділянок Leu9–Lys19 (ділянка 155–167 AIMP1), Tyr27–Asp33 (ділянка 175–180 AIMP1), Arg39–Ser43 (ділянка 187–191 AIMP1), Met58–Leu62 (ділянка 206–210 AIMP1), Leu78–Ser83 (ділянка 225–230 AIMP1), Ile88–Ile90 (ділянка 236–238 AIMP1), Ala138–Tyr140 (ділянка 286–288 AIMP1), Glu146–Lys148 (ділянка 294–296 AIMP1).

У випадку N-термінальної ділянки AIMP1 результати передбачення вторинної структури є абсолютно узгодженими — усі використані сервіси однозначно детектують наявність

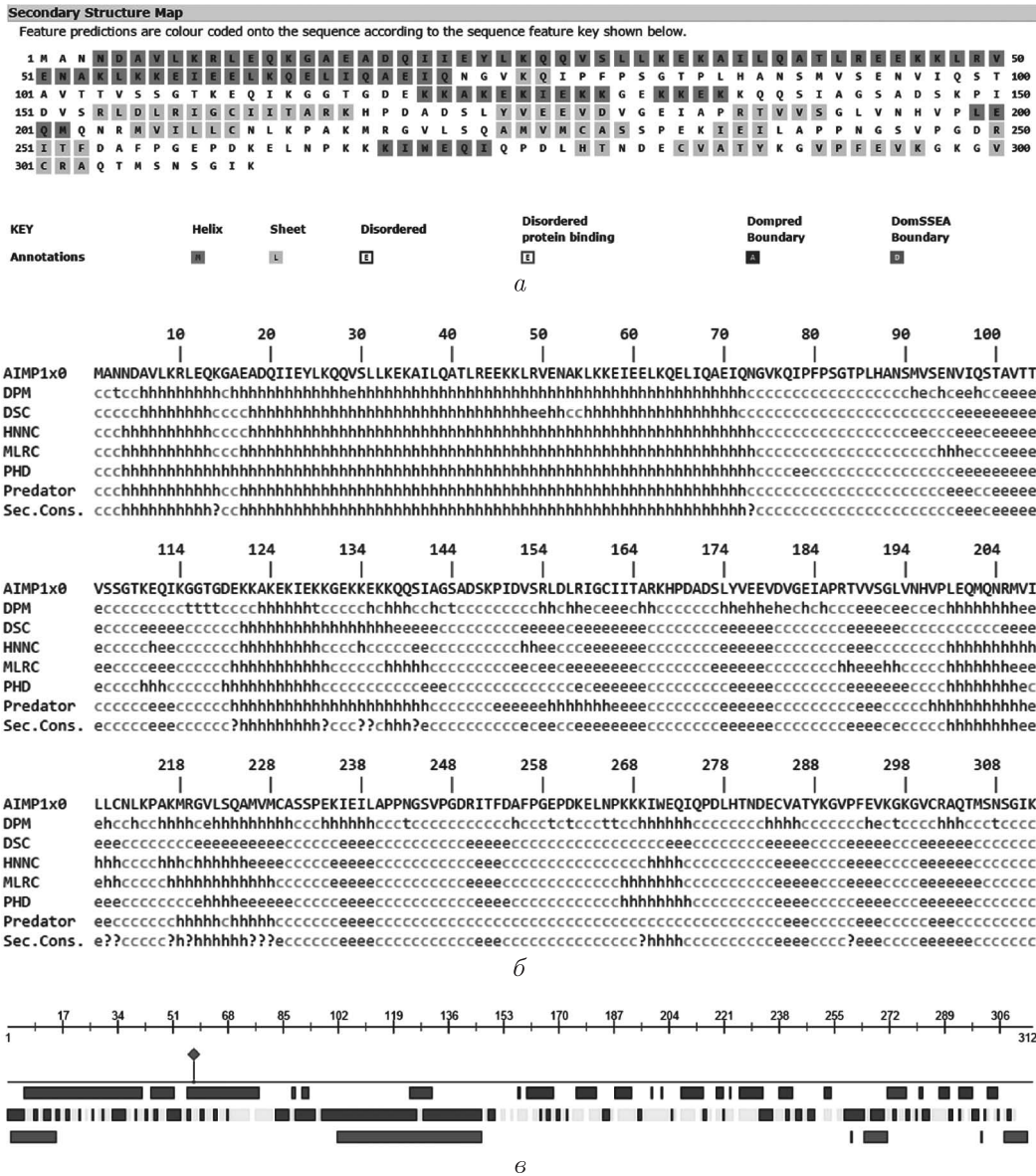


Рис. 2. Вторинна структура AIMP1 людини згідно з даними передбачення за амінокислотною послідовністю: *a* — з використанням веб-сервісу PSIPRED; *b* — за сімома різними алгоритмами з використанням мультипрограмного веб-сервісу NPS@; *b* — з використанням веб-сервісу PredictProtein

великої спіральної ділянки в межах 5–70 а.з., що може містити маленькі неупорядковані вставки (до п'яти залишків). Результати передбачення вторинної структури N-кінця AIMP1 чітко збігаються з даними кристаліграфічних досліджень його комплексу з глютаміною та аргініноюю аміноацил-тРНК синтетазами (див. рис. 1, *a*).

Наявність потенційних ділянок з неупорядкованою структурою було розраховано за допомогою 10 спеціалізованих веб-сервісів (табл. 1). Зведені діаграми передбачення за даними різних сервісів і окремих алгоритмів передбачення наведено на рис. 3. Приклади графічного подання даних від шести веб-інструментів зображено на рис. 4. Використані веб-інструменти розпізнають неупорядковані ділянки, виходячи з їхньої амінокислотної послідовності, на

основі оцінки попарних енергетичних взаємодій між амінокислотними залишками. Постулюється, що неупорядковані ділянки білків не набувають стабільної структури через те, що їх амінокислотний склад не дає сформувати достатню кількість сприятливих взаємодій залишків між собою.

Порівняння передбачень, отриманих за допомогою різних веб-сервісів (див. рис. 4) свідчить про високу ймовірність наявності неупорядкованих ділянок у межах Met1–Asn3 (3 а.з.), Thr103–Lys148 (46 а.з.), Met306–Lys312 (7 а.з.). Найбільш функціонально значущою серед них є велика ділянка з 103 по 148 амінокислотний залишок. Цікаво, що за передбаченнями більшості веб-інструментів для визначення вторинної структури, в межах Lys121–Ser140 потенційно може знаходитись α -спіраль. Це удаване протиріччя досить просто пояснити тим, що відповідна ділянка, крім залишків лізину, характерного для неупорядкованих частин білка, також насичена залишками глютамінової кислоти, яка зазвичай є маркером саме α -спіральних елементів. З великою часткою ймовірності можна припустити, що спіраль у цій області є метастабільною, тобто такою, що внаслідок природних флуктуацій переходить до неупорядкованого стану і назад. Зазначимо також, що ділянка Thr103–Lys148 знаходиться на N-кінці та міждоменному лінкері AIMP1 і частково збігається з ділянками, що відповідають за зв'язування з білком теплового шоку HSP90B1/grp96, антиангіогенез, міграцію та апоптоз клітин ендотелію.

Аналогічний аналіз раніше було проведено для тирозил-тРНК синтетази (TyrPC) ссавців і білка Arg1p, які мають подібну двомодульну структурну організацію, а їхні С-кінцеві модулі є гомологічними до С-кінцевого домену AIMP1 людини [2]. Цим аналізом також показано високу ймовірність неупорядкованого стану міждоменних лінкерів у зазначених білках. Порівнявши результати роботи [2] з отриманими нами, можна зробити висновок, що гнучкий міжмодульний лінкер є характерною ознакою білків, які містять ЕМАРІІ-подібний С-модуль.

Відомо, що природно неупорядковані білки (або відповідні ділянки в складі модульних білків) характеризуються великим конформаційним простором, проте здатні набувати цілком певної просторової структури, взаємодіючи з “канонічними” впорядкованими білками/доменами. У випадку AIMP1 вірогідно припустити наявність більш складного і вишуканого явища — неупорядковані ділянки Thr103–Lys148 у складі цих білків безпосередньо

Таблиця 1. Зведені дані передбачення структурно неупорядкованих ділянок у послідовності AIMP1 людини

Веб-сервер	Алгоритм	Передбачені неупорядковані амінокислотні залишки	Кількість ділянок
DisEMBL1.5	Remark-465	103–147, 305–312	2
	Hot-loops	11–18, 41–50, 101–154, 165–174, 289–312	5
	Loops/coils	74–92, 103–120, 140–157, 164–187, 194–201, 230–268, 274–312	7
DISOPRED2	–	1–3, 107–119, 131–132, 138–148, 307–312	5
DisPROT	VL2	1–4, 119–141, 310–313	3
	VL3	28–30, 36–154, 211–232, 241–265, 290–312	5
	VL3H	1–7, 11–18, 26–154, 232–269, 306–312	5
	VSL2P	1–15, 34–160, 230–274, 307–312	4
FoldIndex	–	30–39, 47–51, 103–154, 251–312	4
GlobPlot2.3	Russel/Linding	68–83, 162–173, 195–202, 323–349, 389–394	5
IUPred	Short	1–7, 107–147, 240–273, 306–312	4
	Long	78–93, 97–103, 107–145, 249–258, 264–272	5
PrDOS	–	1–10, 94–147, 306–312	3
RONN2	–	46–66, 76–147, 252–279	3

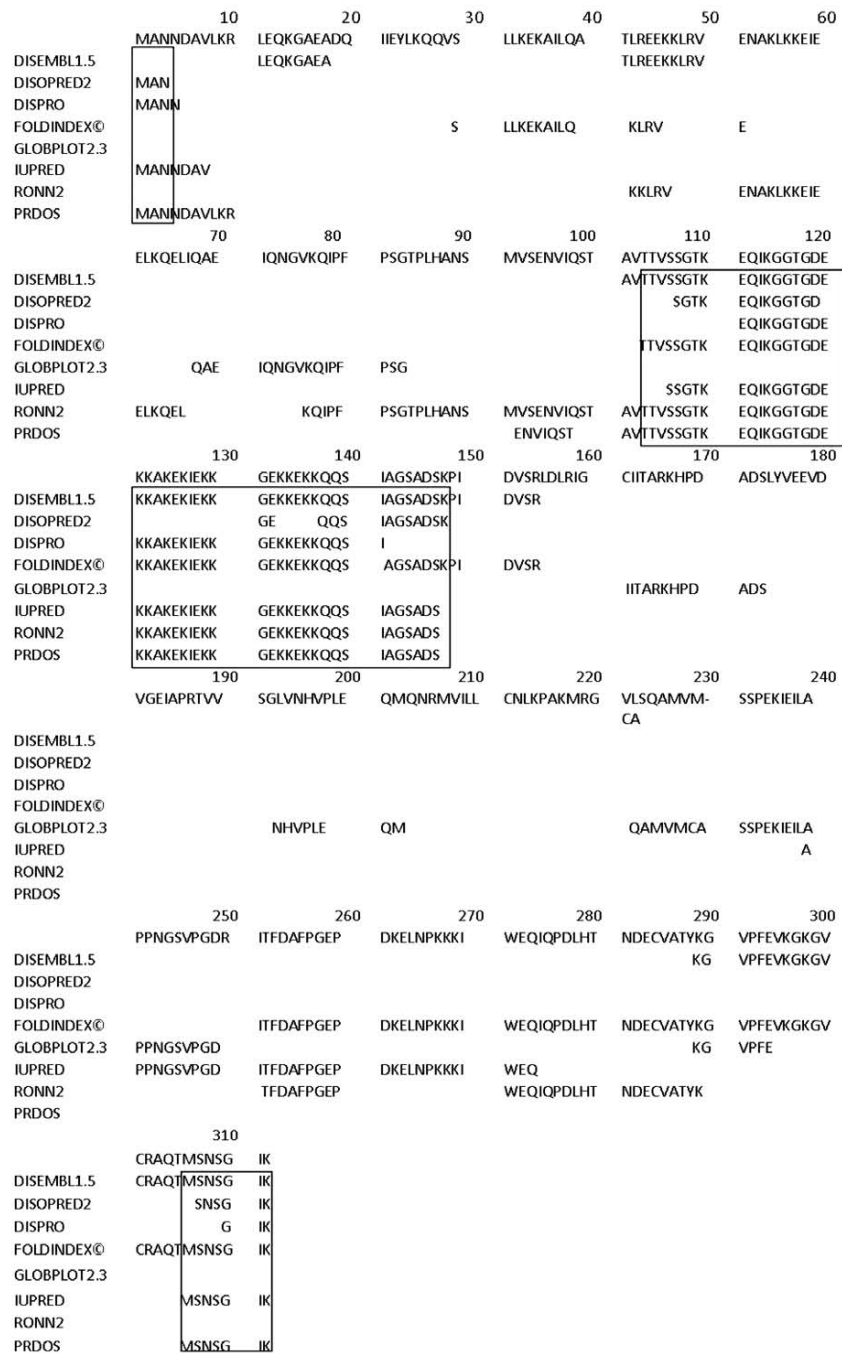


Рис. 3. Зведені результати передбачення неструктурованих ділянок для мономера AIMP1, отримані за допомогою восьми алгоритмів. Ділянки, передбачені одночасно чотирма чи більше алгоритмами, виділено рамкою

залучені в процес димеризації субодиниць AIMP1 і протягом цього процесу індукують взаємний перехід одна одної до певної регулярної структури. Подібна взаємна структуризація є цілком імовірною з огляду на відомі властивості неупорядкованих білків, проте досі не описана в науковій літературі. Очевидно, що неупорядковані ділянки також можуть ві-

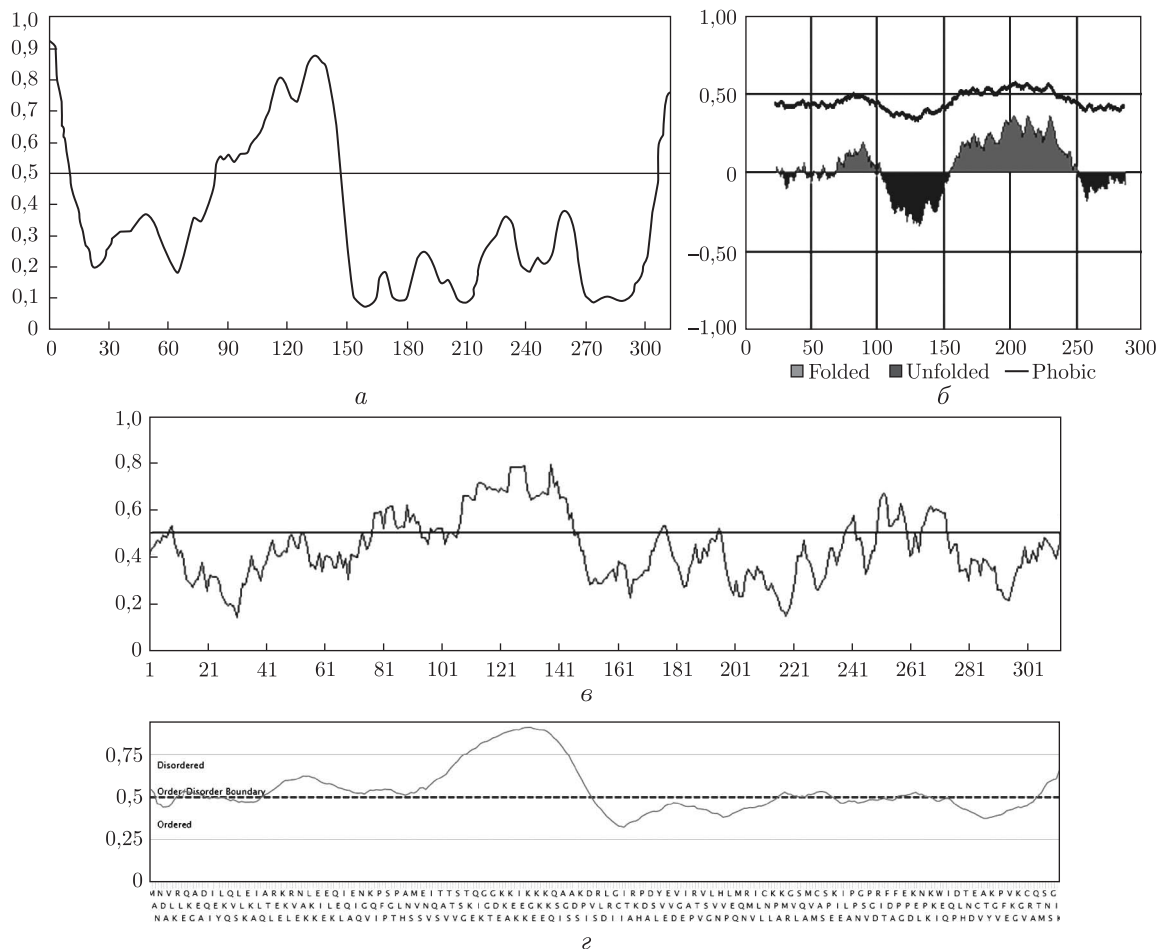


Рис. 4. Приклади передбачення непорядкованих ділянок AIMP1 людини за допомогою деяких серверів у графічному вигляді: *a* — PrDOS; *б* — FoldIndex; *в* — IUPred; *г* — RONN2. По осі абсцис відкладено імовірність наявності непорядкованого стану. На всіх графіках горизонтальною лінією позначено рівень відсікання для ділянок з непорядкованою і впорядкованою структурами

дігравати істотну роль при формуванні комплексів AIMP1 з молекулами тРНК, а також з іншими білками-партнерами. Проте остаточне підтвердження нашої гіпотези потребує детальних досліджень просторової структури і поведінки AIMP1 з використанням, зокрема, методів обчислювальної молекулярної динаміки.

Цитована література

1. Oldfield C. J., Dunker A. K. Intrinsically disordered proteins and intrinsically disordered protein regions // *Annu. Rev. Biochem.* – 2014. – **83**. – P. 553–584.
2. Odynets K. A., Kornelyuk A. I. Analysis of unstructured regions cytoplasmic tyrosyl-tRNA synthetase of human by bioinformatical methods // *Biopolym. Cell.* – 2005. – **21**. – P. 446–453.
3. Kim J. H., Han J. M., Kim S. Protein-protein interactions and multi-component complexes of aminoacyl-tRNA synthetases // *Top. Curr. Chem.* – 2014. – **344**. – P. 119–144.
4. Renault L., Kerjan P., Pasqualato S. et al. Structure of the EMAPII domain of human aminoacyl-tRNA synthetase complex reveals evolutionary dimer mimicry // *EMBO J.* – 2001. – **20**. – P. 570–578.
5. Berman H. M., Westbrook J., Feng Z. et al. The Protein Data Bank // *Nucleic Acids Res.* – 2000. – **28**. – P. 235–242.

6. Fu Y., Kim Y., Jin K. S. et al. Structure of the ArgRS-GlnRS-AIMP1 complex and its implications for mammalian translation // Proc. Natl. Acad. Sci. USA. – 2014. – **111**. – P. 15084–15089.
7. The UniProt Consortium. Activities at the universal protein resource (UniProt) // Nucleic Acids Res. – 2014. – **42**: D191–D198.
8. Buchan D. W. A., Minneci F., Nugent T. C. O. et al. Scalable web services for the PSIPRED protein analysis workbench // Nucleic Acids Res. – 2013. – **41**. – P. W340–W348.
9. Yachdav G., Kloppmann E., Kajan L. et al. PredictProtein – an open resource for online prediction of protein structural and functional features // Nucleic Acids Res. – 2014. – **42**, Web Server issue. – P. W337–W343.
10. Ishida T., Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence // Nucleic Acids Res. – 2007. – **35** (Web Server issue). – P. W460–W464.
11. Linding R., Russell R. B., Neduva V., Gibson T. J. GlobPlot: exploring protein sequences for globularity and disorder // Nucleic Acid Res. – 2003. – **31**. – P. 3701–3708.
12. Prilusky J., Felder C. E., Zeev-Ben-Mordehai T. et al. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded // Bioinformatics. – 2005. – **21**. – P. 3435–3438.
13. Linding R., Jensen L. J., Diella F. et al. Protein disorder prediction: implications for structural proteomics // Structure – 2003. – **11**. – P. 1453–1459.
14. Ward J. J., Sodhi J. S., McGuffin L. J. et al. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life // J. Mol. Biol. – 2004. – **337**. – P. 635–645.
15. Sickmeier M., Hamilton J. A., LeGall T. et al. DisProt: the database of disordered proteins // Bioinformatics. – 2005. – **21**. – P. 137–140.

References

1. Oldfield C. J., Dunker A. K. Annu. Rev. Biochem., 2014, **83**: 553–584.
2. Odynets K. A., Kornelyuk A. I. Biopolym. Cell., 2005, **21**: 446–453.
3. Kim J. H., Han J. M., Kim S. Top. Curr. Chem., 2014, **344**: 119–144.
4. Renault L., Kerjan P., Pasqualato S. et al. EMBO J., 2001, **20**: 570–578.
5. Berman H. M., Westbrook J., Feng Z. et al. Nucleic Acids Res., 2000, **28**: 235–242.
6. Fu Y., Kim Y., Jin K. S. et al. Proc. Natl. Acad. Sci. USA, 2014, **111**: 15 084–15 089.
7. The UniProt Consortium. Nucleic Acids Res., 2014, **42**: D191–D198.
8. Buchan D. W. A., Minneci F., Nugent T. C. O. et al. Nucleic Acids Res., 2013, **41**: W340–W348.
9. Yachdav G., Kloppmann E., Kajan L. et al. Nucleic Acids Res., 2014, **42**, Web Server issue: W337–W343.
10. Ishida T., Kinoshita K. Nucleic Acids Res., 2007, **35** (Web Server issue): W460–W464.
11. Linding R., Russell R. B., Neduva V., Gibson T. J. Nucleic Acid Res., 2003, **31**: 3701–3708.
12. Prilusky J., Felder C. E., Zeev-Ben-Mordehai T. et al. Bioinformatics, 2005, **21**: 3435–3438.
13. Linding R., Jensen L. J., Diella F. et al. Structure, 2003, **11**: 1453–1459.
14. Ward J. J., Sodhi J. S., McGuffin L. J. et al. J. Mol. Biol., 2004, **337**: 635–645.
15. Sickmeier M., Hamilton J. A., LeGall T. et al. Bioinformatics, 2005, **21**: 137–140.

Надійшло до редакції 28.10.2015

Т. С. Лиманская¹, А. Ю. Ныпорко¹,
 член-корреспондент НАН України **А. И. Корнелюк^{2,1}**

¹Институт высоких технологий, Киевский национальный университет им. Тараса Шевченко

²Институт молекулярной биологии и генетики НАН Украины, Киев

E-mail: dfnalex@gmail.com

Анализ неупорядоченных участков белка AIMP1/p43 мультисинтетазного комплекса человека методами биоинформатики

Особенности вторичной структуры и расположения неупорядоченных участков в структуре белка AIMP1 — компонента мультисинтетазного комплекса человека — исследованы с помощью ряда in silico подходов. Выявлено, что наиболее длинный участок белка,

склонный к формированию неупорядоченной структуры, располагается с 103 по 148 аминокислотный остаток. Интересно, что в пределах этого участка (примерно с 121 по 140 остаток) также возможно наличие α -спирали. Это кажущееся противоречие достаточно просто объясняется тем, что соответствующий участок насыщен как остатками лизина (маркер неупорядоченных участков белка), так и остатками глутамата (типичный маркер именно α -спиральных элементов). С большой долей вероятности можно предположить, что эта спираль является метастабильной, т. е. переходящей к неупорядоченному состоянию и обратно вследствие естественных флуктуаций молекулы белка.

Ключевые слова: AIMR1/p43, вторичная структура, неупорядоченные участки, биоинформатика.

T. S. Limanska¹, A. Yu. Nyporko²,
Corresponding Member of the NAS of Ukraine A. I. Kornelyuk^{2,1}

¹Institute of High Technology, Taras Shevchenko National University of Kiev

²Institute of Molecular Biology and Genetics of the NAS of Ukraine, Kiev

E-mail: dfnalex@gmail.com

Analysis of disordered regions of AIMR1/p43 protein from human multisynthetase complex with bioinformatics methods

Features of the secondary structure and locations of the disordered regions in the structure of AIMR1 protein — component of the human multisynthetase complex are investigated with several in silico approaches. It is revealed that the longest part of the protein disclosed to fold into a disordered structure is located from 103 to 148 amino acid residue. Interestingly, within this region (from about 121 to residue 140), the presence of α -helix is also possible. This seeming contradiction is simply explained by the saturation of the appropriate region by both lysine residues (marker of disordered regions of proteins) and glutamate residues (typical marker of α -helical elements). With high probability, we can assume that this spiral is metastable, i. e., moving to a disordered state and back due to natural fluctuations of the protein molecule.

Keywords: AIMR1/p43, secondary structure, disordered regions, bioinformatics.