

УДК 57.081 23 : 681.3.06

© Коллектив авторов

## КОМПЬЮТЕРНЫЙ АНАЛИЗ ДИСТАНТНЫХ И КОНСЕРВАТИВНЫХ ДОМЕНОВ В СТРУКТУРЕ ФИБРОНЕКТИНОВ

А.В. Федотов<sup>1</sup>, И.Н. Овчарук<sup>2</sup>

<sup>1</sup>Московский государственный университет прикладной биотехнологии, г. Москва.

эл. почта: avfedotov@yandex.ru

<sup>2</sup>Кемеровская государственная медицинская академия, г. Кемерово.

Сообщается о результатах компьютерного анализа дистантных и консервативных доменов в структуре полифункционального матричного гликопротеина - фибронектина и функционально связанных с ними макромолекул тканей и биологических жидкостей. Исследование проведено с целью установления тех видов структурно-функциональных и сигнальных связей, которые могут представлять интерес с точки зрения теории эволюции доменов, а также практики поиска алгоритмов функций, связанных с их активностями. Такие связи могут быть "замаскированы" и проступают более явно после уточнения их транскриптомной или сигнальной трансформации на границе фибронектин/интегриновой передачи информации. Дальнейшее пополнение аннотированных баз данных новой информацией, а также договоренности о методологии и терминологии могут способствовать более быстрому решению этой задачи.

**Ключевые слова:** фибронектин, дистантные и консервативные домены, фибронектиновые повторы, RGD-мотив, интегрины, доменная эволюция.

**ВВЕДЕНИЕ.** Фибронектины (Фн) - семейство полифункциональных гликопротеинов (ГП), модулирующих сигнальные пути взаимодействия клеток с окружающей их средой через фибронектин/интегриновый кластер [1]. Молекула Фн взаимодействует с коллагенами, фибрином, гепарином, ДНК, актинами и др. молекулами, вовлекая таким образом клетки в регуляцию специфической биологической адгезии, регенераторные процессы, морфогенез, клеточную подвижность, репаративные и эмбриогенетические физиологические реакции. Молекула Фн (молекулярная масса 220-250 кДа) представлена одной полипептидной цепью, состоящей из 2300-2500 аминокислотных последовательностей (а.п.) Фибронектины *in vivo* присутствуют в виде димерной молекулы в растворимом состоянии во всех биологических жидкостях и в виде нерастворимой формы, как поперечно-сшитые димерные и мультимерные формы на поверхности практически всех видов клеток. Макромолекула Фн содержит аминокислотные последовательности (АП) из "тильда" 2300-2500 единиц с молекулярной массой примерно 220-250 кДа. С-концы макромолекул Фн связаны двумя дисульфидными связями. Структурно - функциональной особенностью этих молекул является то, что они содержат три типа АП с длиной по 40-90 аминокислотных остатков (а.о.) называемых типами Фн1, -2 и -3. Все три типа повторов отличаются по взаимодействию с другими макромолекулами или их участками и т.о. на основании субдоменной интеграции домены объединяются еще в 6 компактных модуля. На уровне четвертичной структуры они выглядят как глобулярно-ленточные образования с соответствующими уровнями взаимодействия, которые в настоящее время и представляют наиболее актуальный научный интерес [1,2].

Общей для адгезивных ГП внеклеточного матрикса, фибринолитической, свертывания крови а также систем комплемента, сосредоточенной в тех функциональных участках, которые обеспечивают им аффинное узнавание сайтов и биологическую активность, является АП -RGDX- (Арг-Гли-Асп-Х). Этот тетрапептид получен синтетически и в ряде случаев имитирует такое действие целой молекулы, как биоспецифическая адгезия. Фн третьего типа различаются также по наличию

альтернативно сплайсированных внутридоменных сегментов, обозначаемых как ExtraDomen (ED): V (ШСС), ЕША(ED-A) и ЕШВ(ED-B). Причем сегменты ЕША и ЕШВ, преимущественно активируются в ходе эмбрионального развития. Активность V(ШСС) сегмента обнаружена в модельных экспериментах по индукции пролиферации клеток в интима сосудов *in vitro*. При этом в плазме крови выявляются изоформы Фн, содержащие специфические для этого вида активности участки Фн типа 3, а именно: V120+, V95+ и V0. Исследование механизма такой активности представляет значительный теоретический и практический интерес.

**МЕТОДИКА.** Нашими образцами для исследования служили данные о АП Фн из баз данных (БД) SWISS-PROT, PDB, NCBI и др. Последовательность компьютерного анализа дистантных и консервативных доменов (ДКД) в нашем случае и в других аналогичных исследованиях сводилось к следующим действиям. Первичный запрос на подобие рестриктов записывали с использованием языка типа Python или SQL. На сайтах многих БД такой запрос можно ввести с помощью соответствующего inquiry (см. [www.exrasy.org/prosite](http://www.exrasy.org/prosite)). Полученные варианты АП подвергали кластерному анализу, по которому можно узнать о их принадлежности соответствующим доменам. На следующем этапе просматривали аннотированные БД типа SBASE. Здесь получали точную текстовую характеристику искомого АП, включая изоформы. Установив степень их гомологичности, приступали к составлению формулы паттернов. По их близости к функциональным доменам и проверке иерархии подобий оценивали степень их сходства. Далее приступали к построению макромолекулярной структуры и/или поиску их сигнальной взаимосвязи. Для решения этой задачи - устанавливали кластер взаимодействия и паттернализованный участок их сигнального взаимодействия. Далее находили наиболее подходящие подходы для решения такой задачи. Наиболее адекватный сайт и соответствующие инструменты для такого действия необходимо подбирать в каждом случае отдельно. Для решения нашей задачи подошел сайт: <http://www.cbs.dtu.dk/services/signal-2.01>. С целью дальнейшего анализа ДКД мы использовали методы выравнивания в уже установленных подобиях, родственных или функционально близких (объединенных паттерном взаимодействия) доменов. Для этой цели, выделенная соответствующим образом доменная группа, анализируется методами мультиплицированного выравнивания, где можно предположить присутствие комплекса последовательностей и в дальнейшем построить иерархию их возникновения. Для этой цели нами использовались соответствующие инструменты протеомного анализа консервативных доменов - CCD. На этом этапе уже возможна фильтрация "замаскированных" и др. АП, которые подтверждают или опровергают рабочую гипотезу.

**РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.** При исследовании наличия идентичных Фн типа 1 FINC\_BOVIN из базы данных SWISS PROT было установлено 94 образца, имеющих гомологичные участки (повторы) в БД Inter Pro с иерархией уровня идентичности: 50, 65 и 80%. Сравнимая автоматически длина полипептидной цепи (ППЦ) равнялась 45 аминокислотным остаткам (a.o.).

Полученная при этом консенсусная последовательность 28-39 (CTCLGGGRGEW) позволила при соответствующем предварительном и сравнительном анализе установить паттерн предполагаемого консервативного домена. Далее мы исследовали его наличие в других макромолекулах для установления его предполагаемых структурно-функциональных связей. Для этой цели был сделан запрос (Query) к БД поиска паттернов представляющей для нас прямой интерес:

\* Pattern: C-x-C-x-G-x-G

\* Motif description: USER\_PATTERN.

\* Databases: swiss [SWISS-PROT Release 40.14 of03-Apr-2002].

· Taxonomic range:

· Bos taurus. ...searching swiss [http://hits.isb-sib.ch/cgi-bin/hits\\_patsearch\\_parser](http://hits.isb-sib.ch/cgi-bin/hits_patsearch_parser)

Полученный в этом случае результат:

>sw:FINC\_BOVIN/92-98 motif=USER\_PATTERN Fibronectin (FN).

CTCIGAG

>sw:FINC\_BOVIN/136-142 motif=USER\_PATTERN Fibronectin (FN).

CVCLGNG

>sw:FINC\_BOVIN/182-188 motif=USER\_PATTERN Fibronectin (FN).

CTCLGEG

>sw:FINC\_BOVIN/227-233 motif=USER\_PATTERN Fibronectin (FN).

CICTGNG

>sw:FINC\_BOVIN/302-308 motif=USER\_PATTERN Fibronectin (FN).  
 CTCLGNG  
 >sw:FINC\_BOVIN/465-471 motif=USER\_PATTERN Fibronectin (FN).  
 CTCVGNĠ  
 >sw:FINC\_BOVIN/512-518 motif=USER\_PATTERN Fibronectin (FN).  
 CTCFGQG  
 >sw:FINC\_BOVIN/556-562 motif=USER\_PATTERN Fibronectin (FN).  
 CYCYGRĠ  
 >sw:FINC\_BOVIN/2112-2118 motif=USER\_PATTERN Fibronectin (FN).  
 CQCLGFG  
 >sw:FINC\_BOVIN/2155-2161 motif=USER\_PATTERN Fibronectin (FN).

Анализируя и сравнивая получаемые данные с другими видами молекул-образцов с соответствующей иерархией идентичности выравниваемых АП или по наличию гомологичных повторов искомого паттерна, переходили к следующему этапу исследования. В окончательном виде паттерн Фн типа 1 выглядит следующим образом в транскрипции IUPAC: C-(8)-[YF]-x(12,5)-C-x-C-x-[GSR]-x-[GRENQ]-x(5)-C-x-(5)-x-. Внутримолекулярное выравнивание типов Фн 1 и 2 FINK\_HUMAN обнаруживает консенсусный паттерн C-x(6,8)-[LFY]-x(5)-[FYW]-x-[RK]-x(8,10)-C-x-C-x(6,9)-C и у FINC\_BOVIN (Type I fibronectin domain bovinum) C-x(6,8)-[LFY]-x(5)-[FYW]-x-[RK]-x(8,10)-C-x-C-x(6,9)-C. Для Фн типа 2 -(Type II fibronectin collagen-binding domain) паттерн (PA) соответственно представлен C-x(2)-P-F-x-[FYWI]-x(7)-C-x(8,10)-W-C-x(4)-[DNSR]-[FYW]-x(3,5)-[FYW]-x-. Паттерн доменов подобных эпидермальному фактору роста -(EGF-like domain signature 1) - соответственно: C-x-C-x(5)-G-x(2)-C.

В связи с вышеизложенным нам представляется важным представить нашу интерпретацию семантики структуры химической связи а.о., а также ее связь с местоположением данного а.о., которое в конечном итоге формирует смыслополагающее действие. Если АП определяет 3D структуру, то местоположение отдельной аминокислоты определяет  $\alpha$ -спиральность и  $\beta$ -повороты. Это также следует трактовать как представленность, а именно семантику и код будущего программирования функции.

Биологическое значение, которое необходимо выразить определенной логикой а.о. и их сравнительный анализ - задача чрезвычайной сложности. Подход, который мы применили для решения своей задачи, состоял в том, что аналогичные а.о. в первичной структуре выбранных нами образцов, совпадающие и разделенные вставками различной длины, характеризуют подобие. Процесс такого выбора подобия подбирался в соответствии с оптимизацией или селективностью биологической функции, которую это подобие должно решать, т. е. на основе целеполагания. В современной теоретической биологии это обозначается как "общая биологическая цель". Программа выравнивания матрицы текста, таким образом, содержит только длину последовательности, которая является базисной для написания алгоритма поиска ДП. Следствие и значение обнаружения таких логических текстов в матричных ГП обсуждались в соответствии с типами вставок и их иерархией для интерпретации смысла, если не считать их случайными в таком тексте. Такой смысл был постулирован нашими более ранними исследованиями. Приведем в качестве примера данные, развивающие представления о связи паттернализации и ее смысловых оснований (семантики), а именно, соотношения между последовательными и непоследовательными остатками коллагена, придающими ему свойство гидрофобности I, L, V и F. Все эти буквы текста и их очередность имеют свой смысл отдельно и в определенной последовательности. Наличие во вставках дипептида или трипептида будет нести более высокий уровень обобщений, чем гидрофобность и чаще всего он обозначает принадлежность к характеристикам свойственным глицин-обогатенным или пролин-обогатенным участкам коллагеновых белков. Наиболее привлекательной способностью определенных АП, обеспечивающих 3D-укладку /фолдинг/ протеинов матричного происхождения, обладают ДП с участием цистеина. Очевидной здесь является роль пролина, который может изменять  $\alpha$ -спираль и  $\beta$ -повороты. В исследованиях паттернализации (in situ) эти процессы следует обсуждать в связи с другими модулями белковых структур, которые изменяют их по более тонкому, а следовательно, и более сложному механизму.

В таблице № 1 приведены белки, имеющие мотив, выполняющий функции передачи сигнала из окружающего клетку пространства во внутриклеточные структуры посредством взаимодействия Фн и интегринов.

Таблица 1. Гликопротеины, имеющие гомологичный Фн тип 3 мотив - RGDX.

Образцы ГП	Искомая АП
Фибронектин бовисный	2RGD
Гликопротеин лейшманиозный	RGDA
Дискоидин 1	RGDF
СЗ Комплимент	RGDQ
Интегрин ПМЯЛ	RGDSC
МАГ	RGD
МНС II	RFDS, RYDS
Тенасцин	RGDM
Остеопонтин	RGDS
Фактор фон Вилле Брандта	RGDS, (RGDC), (RYD)
Коллагены *	RGDTPC
Ламинин	RGDN
Фибронектины	RGDS, REDV
Дисинтегрин	RGD

\*Примечание: Последние 4 образца таблицы объединены в структурно-функциональный паттерн агрегации/деагрегации.

Паттернализация (Пт) на основе базисных групп подобия и ее представленность в виде структурных алгоритмов, на наш взгляд, наиболее перспективна в связи с ее наиболее общим для интерпретации смыслом. Ее биологическое значение, а также роль, которую она занимает в других более частных биологических интерпретациях определяются профилем подобия в искомым образцах. Пт часто связаны с ДП. Было разумным предположить что в Фн, различных видов живых организмов, разделенных в эволюционном развитии достаточно большими промежутками времени, и обозначенных как Фн-3, такая связь может существовать. Анализируя роль Пт в доменной структуре Фн, можно получить наиболее полную картину механизма второго каскада взаимодействия, обеспечивающего прикрепление молекулы Фн к соответствующим сайтам интегриновых рецепторов, и сделать практические выводы, которые следуют из понимания этого взаимодействия. Эта область исследований привлекает наиболее пристальное внимание всех групп, работающих в данной области. Особенный интерес представляют те события, которые следуют за таким сигналом биологического взаимодействия. Эти исследования ведутся многочисленными разработчиками этой проблемы, при поддержке фармацевтических компаний и требуют параллельных и совместных исследований с другими научными группами.

**ЗАКЛЮЧЕНИЕ.** Таким образом, согласно нашим данным анализа АП Фн тип 3 и данным других авторов, следует различать два типа повторов - ФнША и ФнШВ, которые соответственно обеспечивают два разных механизма взаимодействия Фн и интегринов, генерирующих различные типы сигнала.

Повторы Фн ШВ предположительно являются более поздними в эволюционном происхождении. Количественная и качественная интерпретация получаемых данных затрудняется разнородностью их трактовки разными рабочими группами (РГ) и отсутствием объединяющей гипотезы пользователей разных БД. [4,5]. Всеми работающими в этом направлении РГ выделено 120 повторов типа ФнША и 10 повторов Фн ШВ. Все 130 повторов отчетливо различающихся по признаку ДП, что пытаются интерпретировать в связи с их ролью в специфической биологической адгезии. Загруднения, вызываемые неполнотой "репрезентативной выборки" и имеющихся в БД образцов, будут постепенно преодолены путем взаимодействия исследовательских групп.

В практической плоскости для фармацевтических целей, следует продолжить дальнейшее изучение КДП Фн.

Для нашей области исследований можно выделить три типа решаемых задач:

- создать совокупность функционально близких Фн и структурно родственных ГП;
- установить КДД наподобие тех, которые существуют в повторах Фн типа I и II;
- продолжить идентификацию замаскированных КДД на основе метамотивов и родственных им мотивов возможного взаимодействия.

Число таких задач может увеличиваться и это совершенно очевидно, но очевидно и

то, что они будут отбираться на основе той логики, которую выбирает исследовательская группа или принимает ее в связи с общими интересами.

Очевидным в, связи с вышеизложенным, является создание отдельной единой БД подобий и ДП. Принципы построения таких БД должны учитывать функции, структуру и функции доменов и их рецепторов в соответствующих паттернах, а также многое другое, что реализовано в больших БД типа PRODOM и SWISS-PROT. Таким образом, можно утверждать, что применение компьютерного анализа в расшифровке структуры и функции ГП представляет собой растущую отрасль биоинформатики. Среди компьютерных программ, дающих удовлетворительные результаты, в частности, по сравнительному анализу дистантных подобий, следует упомянуть FASTA и BLAST, PFAM и др. Конкретное же решение и постановка задачи определяют выбор других программ, позволяющих на различных этапах выявлять скрытые или маскированные семантики или мотивы [6].

#### ЛИТЕРАТУРА

1. *Mijamoto S, B-Z Katz, Lafrene R.M. fnd K.Yamada* (2002) *Annals N-Y Academy of Science*, p.119-129.
2. *Craid D., Kramer A, Schuiten K., Vogel V.* (1996) *PNAS* 98.10, 5590-96.
3. *Haughn Jang A., Abraham, R., Ottmers, V., Musher, D.* (1996) *J.Imml.*, 157, N2, p.720-31.
4. *Chi Y. Wu.* (1997) *Austr.Ann.Meet.Prot.Res.Sci*, "Exp.BioL", "New Orleans", Apr.6-9, II. 3, 541.
5. *Xiaoown Li et al.* (2003) *J.Biol.Chem.*, 278, 43, 41646-53.
6. *Van der Loo J.C.* (1998) *J.Clin.Invest.*, 102, 1051-61.

#### THE COMPUTER ANALYSIS OF DISTANT AND CONSERVATIVE DOMAINS IN STRUCTURE OF FIBRONECTINS.

*A.V.Fedotov<sup>1</sup>, I.N.Ovcharuk<sup>2</sup>*

<sup>1</sup>Moscow State University of Applied Biotechnology; Moscow, Russia;

<sup>2</sup>Kemerovo State Medical Academy.

e-mail: avfedotov@yandex.ru

It is informed on results of the computer analysis distant and conservative domains in structure multifunctional matrix glycoprotein - fibronectin and the macromolecules of biological liquids connected by it and fabrics. Research is lead with the purpose of an establishment of those kinds of structurally functional and alarm interactions which can be of interest from the point of view of the theory of evolution of domains and also practice of search of algorithms of functions connected with their activity. Such communications can be masked and appear more obviously after their specification transcriptomics or alarm transformation on border fibronectin/integrin transfers of the information. The further completingot the data on such mechanism and the arrangement in the field of methodology and terminology will promote faster solution of this bioinformatics problem.

**Key words:** fibronectin, distant and conservative domains, fibronectins type 1,2,3, integrin, domain evolution, RGD-motif.