

УДК 577.21:577.3

ИССЛЕДОВАНИЕ ПЕРВИЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ГЕНА *gapA* И БЕЛКА GapA БАКТЕРИИ *MYCOPLASMA GALLISEPTICUM*

Л.Л. Фролова, А.Я. Хидиятуллина

Аннотация

В статье рассматриваются структуры гена *gapA* и его продукта – белка GapA бактерии *Mycoplasma gallisepticum*, полученные из баз данных первичных последовательностей ДНК, РНК и белков и обработанные специализированными программными пакетами TandemScan, SOPMA, TMHMM. Особое внимание уделено нахождению областей в структуре белка, участвующих в процессе клеточной адгезии.

Ключевые слова: структура гена *gapA*, структура белка GapA, базы данных GenBank, GenPept, PFam, Prosite, программы TandemSWAN, SOPMA, TMHMM, CPNmodels.

Введение

Вид *Mycoplasma gallisepticum* относят к роду *Mycoplasma*, семейству *Mycoplasmataceae*, порядку *Mycoplasmatales*, отряду *Tenericutes (Mollicutes)*, царству *Procarrioteae* [1]. Микоплазмы – одни из мельчайших прокариот, способных к самостоятельному воспроизведению. Представители класса *Mollicutes* широко распространены в природе и выявляются у человека и других млекопитающих, птиц, рыб, моллюсков, насекомых, растений. Микоплазмы могут вызывать различные заболевания в клетках хозяина, однако возможно их длительное сохранение в организме хозяина без развития каких-либо патологий.

Критериями патогенности микоплазм при использовании культур клеток являются адгезивные свойства штаммов, способность вызывать цитопатическое действие (ЦПД), размножаться и персистировать в клетках.

Патогенные микоплазмы обычно не разрушают организм, а колонизируют эпителиальные поверхности тканей хозяина. При проникновении в организм бактерия *M. gallisepticum* сначала устанавливает специфическую и прочную связь с клеткой благодаря процессу клеточной адгезии, который необходим, чтобы избежать быстрого разрушения бактерии врожденными механизмами защиты клетки-хозяина. Молекулярные основы клеточной адгезии *M. gallisepticum* до конца не изучены.

Однако недавно было показано, что клеточная адгезия является сложным многофакторным процессом, в котором координировано действуют белки GapA, CtnA, PvpA и другие молекулы клеточной адгезии [2].

Ген *gapA* кодирует белок GapA с молекулярной массой 121.5 кДа. Сравнительный анализ нуклеотидной последовательности *gapA* показал 45%-ную го-

мологию с *P1*-геном *Mycoplasma pneumoniae*, 46%-ную – с *MgPa* *Mycoplasma genitalium* и 47%-ную гомологию с геном, кодирующим P1-подобный белок у *Mycoplasma pirum*. В гене *gapA* содержится 64% АТ-пар по сравнению с 46%, 60% и 72% соответственно для *P1*, *MgPa* и гена, кодирующего P1-подобный белок. Как гены *P1* и *MgPa*, *gapA* является первым геном в мультигенном опероне, но, в отличие от них, он представлен только одной копией в геноме. Продуктом гена является трипсин-чувствительный поверхностный белок GapA [3].

Для изучения роли GapA в клеточной адгезии сравнили белковые профили *M. gallisepticum* штамма R_{low}, полученный в результате 15 пассажей, и неvirulentного штамма *M. gallisepticum* R_{high}, полученного в результате 164 пассажей. Было показано, что у последнего отсутствует ряд белков, в том числе участвующих в клеточной адгезии: GapA, 116-кДа белок (p116) и 45-кДа белок (p45). Путем транскрипционного анализа удалось выяснить, что *gapA* не транскрибируется у *M. gallisepticum* R_{high}. Сравнительный анализ последовательности ДНК *gapA* у штаммов *M. gallisepticum* R_{low} и R_{high} показал, что в позиции 105 происходит инсерция аденина. В результате чего образуется стоп-кодон (TAA), и таким образом происходит преждевременное завершение трансляции. Принимая во внимание возможность образования белка с молекулярной массой около 4 кДа, дополнительные исследования, выполненные авторами, не выявили продукт данного размера. По мнению Папазиси и др. [4], редуцированный белок либо не синтезируется, либо подвергается посттрансляционному гидролизу. Восстановление синтеза GapA у *M. gallisepticum* R_{high} не приводило к восстановлению вирулентности штамма для цыплят. Авторы предполагают, что в клеточной адгезии *M. gallisepticum* кроме GapA определенную роль играют продукты генов p116 и/или p45. Ген, кодирующий белок p116 (*crmA*), следует за геном *gapA*, образуя единый оперон, транскрипционный анализ которого показал отсутствие транскрипции *crmA*-гена в клетках *M. gallisepticum* R_{high}.

Целью работы является исследование нуклеотидной последовательности гена *gapA*, аминокислотной последовательности его продукта – белка GapA и выделение областей, участвующих в процессе клеточной адгезии.

1. Материалы и методы исследования

Для поиска и анализа нуклеотидной последовательности гена *gapA* *M. gallisepticum* R использовалась база данных (БД) GenBank и аминокислотной последовательности белка GapA – БД GenPept на сайте NCBI (<http://www.ncbi.nlm.nih.gov/>), для просмотра организации доменов в белке GapA – БД Pfam (<http://www.sanger.ac.uk/Software/Pfam/index.shtml>) и БД PROSITE (<http://www.expasy.org/prosite/>).

Поиск тандемных повторов проведен программой TandemSWAN (<http://strand.imb.ac.ru/SWAN/index.html>), для определения вторичной структуры белка использована программа SOPMA (<http://npsa-pbil.ibcp.fr/cgi-bin/>), для поиска внутри- и внеклеточных участков белка, а также его трансмембранных регионов использована программа ТМНММ (<http://www.cbs.dtu.dk/services/TMHMM>).

Пространственная структура белка построена с помощью программы CPNmodels 2.0. (<http://www.cbs.dtu.dk/services/CPNmodels>).

2. Результаты исследований и их обсуждение

2.1. Анализ нуклеотидной последовательности гена *gapA* *Mycoplasma gallisepticum* R. В геноме *M. gallisepticum* R_{low} ген *gapA* расположен в области 223214..226585 пар оснований (рис. 1).

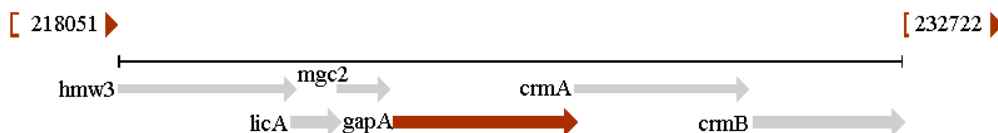


Рис. 1. Расположение гена *gapA* в геноме *M. gallisepticum* R

В БД GenBank ген *gapA* зарегистрирован под уникальным номером AE015450. Длина нуклеотидной последовательности составляет 3372 пар оснований с содержанием Г+Ц пар 35.82%.

В гене *gapA* обнаружено 47 тандемных повторов длиной от 3 до 15 пар нуклеотидных оснований (табл. 1).

2.2. Анализ аминокислотной последовательности белка GarA. Продукт гена *gapA* – белок GarA (AP56530) называют адгезин P1, а также в его названии используются синонимы: цитадгезин P1, адгезивный белок, сцепленный белок. GarA является мембраносвязанным белком длиной 1123 аминокислотных остатка и молекулярной массой равной 121.5 кДа. Процентное содержание аминокислот представлено в табл. 2. Как видно из таблицы, процентное содержание аминокислот белка GarA колеблется в пределах от 0% (цистеин) до 9.2% (аспарагин). Аминокислоты с наибольшим процентом содержания – аспарагин (9.2%), серин (9.1%), аланин (8.5%) – относятся к заменимым аминокислотам.

2.3. Структурный анализ GarA белка. Вторичная структура GarA определена с помощью программы SOPMA. В пределах аминокислотной последовательности на альфа-спирали приходится 210 аминокислотных остатков (18.70%), на протяженные нити – 276 (24.58%), на бета-изгибы – 121 (10.77%), на случайные витки – 516 (45.95%).

В аминокислотной последовательности были выделены следующие участки:

- 1) по отношению к цитоплазматической мембране (рис. 2):
 - внутриклеточный участок белка (5–27, 1020–1122),
 - внеклеточный участок белка (28–996),
 - трансмембранный спиральный участок (1–4, 997–1019);
- 2) по структуре последовательности:
 - сигнальный пептид,
 - участки низкой сложности (15–35 аминокислот),
 - беспорядочные участки.

В белке GarA с помощью программы PRODOM было найдено 5 доменов: Pfam-B_33296 (1–194), Pfam-B_62070 (215–355), Pfam-B_20861 (356–419), Pfam-B_51660 (420–968), Pfam-B_10685 (987–1088). С помощью программы Pfam-A обнаружить функционально значимые домены в изучаемой последовательности не удалось.

Табл. 1

Тандемные повторы, встречающиеся в гене *gapA*

№	Повторяющаяся последовательность	Количество повторов	Расположение тандемного повтора
1	ARKCY	3	143-158
2	GAW	3	175-184
3	HTMAC	3	187-202
4	TKGWW	3	265-280
5	AYWG	3	352-364
6	TTA	3	459-468
7	SAA	3	486-495
8	XSTGMTRCW	3	534-561
9	CWRSWRGWRGTGSTK	3	583-628
10	HCKAYKDCT	3	693-720
11	TADMRWTYMWAA	3	728-764
12	WXHGAMAGT	3	777-804
13	MASAHT	3	828-846
14	AWTT	3	880-892
15	WGTWAW	3	857-875
16	MTMMTT	3	949-967
17	MATSMT	3	1020-1038
18	WAC	3	1091-1100
19	VTWWAKRTG	3	1192-1219
20	TYA	3	1362-1371
21	RYXTTGKTYMWS	3	1420-1456
22	MCT	3	1593-1602
23	TAR	3	1613-1622
24	STAAAYR	3	1672-1690
25	ATY	3	1790-1799
26	DWWWMWTTA	3	1809-1836
27	WGYTA	3	1872-1887
28	TAW	3	1946-1955
29	CAA	3	1971-1980
30	YTAMTWAKW	3	2092-2119
31	AST	3	2223-2232
32	TYA	3	2283-2292
33	HMRCARXTA	3	2392-2419
34	GAT	3	2457-2466
35	TASW	3	2533-2545
36	CYG	3	2509-2518
37	YTHWASCS	3	2600-2624
38	CAT	3	2701-2710
39	AKT	3	2837-2846
40	THXCGA	3	3001-3019
41	VAT	4	3017-3029
42	YTAGST	3	3039-3057
43	AWT	3	3165-3174
44	WRGA	3	3225-3237
45	MABC	4	3244-3260
46	RRACCADCW	3	3276-3303
47	MMASCAMMARCT	3	3321-3357

Табл. 2

Содержание аминокислот в белке GarA

Название аминокислоты	Символ	Количество аминокислотных остатков	% содержание
Аспарагин	N	103	9.2
Серин	S	102	9.1
Треонин	T	98	8.7
Аланин	A	95	8.5
Глицин	G	87	7.7
Лейцин	L	77	6.9
Валин	V	72	6.4
Пролин	P	69	6.1
Лизин	K	62	5.5
Изолейцин	I	60	5.3
Аспарагиновая кислота	D	54	4.8
Фенилаланин	F	47	4.2
Глутамин	Q	44	3.9
Тирозин	Y	44	3.9
Глутаминовая кислота	E	37	3.3
Аргинин	R	36	3.2
Метионин	M	17	1.5
Триптофан	W	15	1.3
Гистидин	H	4	0.4
Цистеин	C	0	0.0

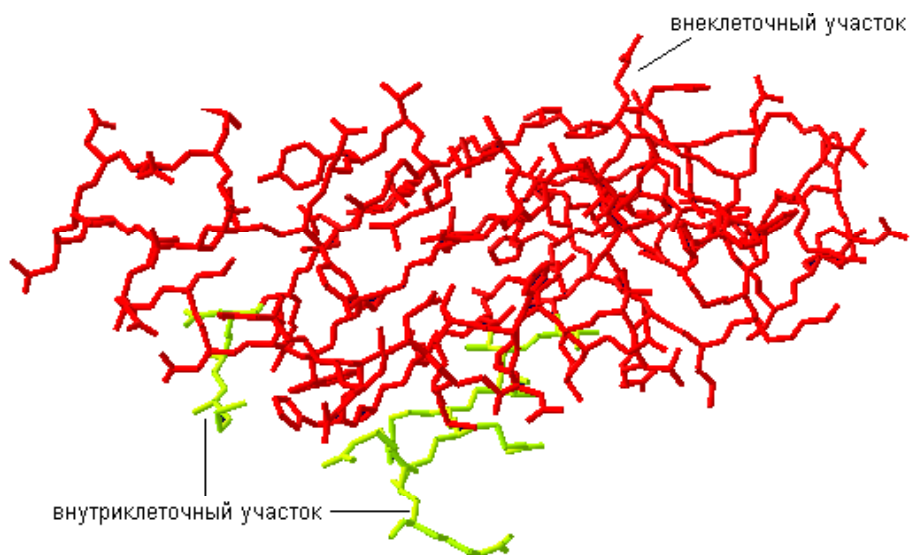


Рис. 2. Пространственная структура белка GarA

В аминокислотной последовательности белка GapA обнаружено 76 мотивов, которые можно сгруппировать в 6 типов по способу модифицирования последовательности – сайты N-гликозилирования, фосфорилирования цАМФ- и цГМФ-зависимой протеинкиназой, фосфорилирования протенкиназой С, фосфорилирования казеинкиназой II, фосфорилирования тирозинкиназой и N-миристоилирования.

Сайт N-гликозилирования (N-{P}-[ST]-{P}) – ASN_GLYCOSYLATION (PS00001, PROSITE documentation PDOC00001, Associated ProRule PRU00498). В последовательности GapA имеется 15 таких сайтов, и они занимают следующие позиции: 64–67 (nfts), 130–133 (nvsd), 415–418 (ngtw), 421–424 (nesa), 436–439 (nrta), 539–542 (nssg), 563–566 (nesi), 652–655 (nfty), 699–702 (nntn), 729–732 (nity), 734–737 (nltr), 773–776 (ngsi), 809–812 (nwtk), 826–829 (nvsk), 1058–1061 (nrts). N-гликозилирование осуществляется по аспарагину, расположенному через один аминокислотный остаток от триптофана. Для сайта N-гликозилирования характерна последовательность Asn-Хаа-Ser/Thr, но присутствие трипептида недостаточно, чтобы началось гликозилирование [5]. Известно, что сворачивание белка играет важную роль в регуляции N-гликозилирования. Присутствие Pro между Asn и Ser/Thr ингибирует N-гликозилирование, это было подтверждено при статистическом анализе сайтов гликозилирования, где было показано, что приблизительно 50% сайтов, имеющих Pro в С-конце около Ser/Thr, не гликозилируются [6]. Есть также данные о случаях гликозилирования сайтов с паттерном Asn-Хаа-Cys [7].

Сайт фосфорилирования цАМФ- и цГМФ-зависимой протеинкиназой ([RK](2)-x-[ST]) – CAMP_PHOSPHO_SITE (PS00004, PROSITE documentation PDOC00004). В последовательности GapA имеется 2 сайта в позициях 639–642 (rkds), 779–782 (kkas). Оба типа киназ отдают предпочтение фосфорилированию серина и остатков треонина, находящихся близко к двум последовательным N-терминальным основным остаткам [8].

Сайт фосфорилирования протенкиназой С ([ST]-x-[RK]) – PKC_PHOSPHO_SITE (PS00005, PROSITE documentation PDOC00005). В последовательности гена GapA подобных сайтов 12 и они занимают позиции 108–110 (tgk), 135–137 (ssk), 218–220 (tar), 494–496 (spr), 578–580 (svr), 638–640 (trk), 681–683 (tak), 701–703 (tnk), 754–756 (ssk), 816–818 (sir), 1038–1040 (snk), 1119–1121 (tkk). В естественных условиях происходит фосфорилирование остатков серина или треонина [9].

Сайт фосфорилирования казеинкиназой II (СК-2) ([ST]-x(2)-[DE]) – CK2_PHOSPHO_SITE (PS00006, PROSITE documentation PDOC00006). В последовательности GapA 18 таких сайтов и они находятся в следующих положениях: 87–90 (srld), 138–141 (tvve), 187–190 (tsae), 238–241 (stse), 242–245 (tvid), 258–261 (tlld), 375–378 (spye), 432–435 (tkfe), 638–641 (trkd), 716–719 (twpe), 736–739 (trtd), 764–767 (stgd), 791–794 (twqd), 896–899 (srnd), 903–906 (sffe), 915–918 (tead), 984–987 (sdqe), 989–992 (tgfd). Активность казеинкиназы II не зависит от циклических нуклеотидов и кальция. СК-2 фосфорилирует различные белки. Специфичность субстрата этого фермента может быть охарактеризована следующим образом [10].

- В определенных условиях серин подобен треонину.

- Кислый остаток (Asp или Glu) должен быть среди трех остатков С-терминального сайта, акцептора фосфата.

- Дополнительные кислые остатки в позициях +1, +2, +4, и +5 увеличивают скорость фосфорилирования. Большинство физиологических субстратов имеют не один кислый остаток в этих позициях.

- Основной остаток в N-конце акцепторного сайта уменьшает скорость фосфорилирования, в то время как кислый увеличивает ее.

Сайт фосфорилирования тирозинкиназой ([RK]-x(2,3)-[DE]-x(2,3)-Y) – TYR_PHOSPHO_SITE (PS00007, PROSITE documentation PDOC00007). В последовательности GapA имеется 2 таких сайта, и они занимают позиции 843–849 (rsldsyy), 920–927 (kevltsky). Субстраты тирозинкиназы характеризуются наличием лизина или аргинина за семь остатков от N-терминального конца до фосфорилируемого тирозина. Кислые остатки (Asp или Glu) часто встречаются за три или четыре остатка с N-терминального конца. Есть множество исключений из этого правила, например, сайт фосфорилирования тирозинкиназой у енолаз и липокортин II [11].

Сайт N-миристоилирования (G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}) – MYRISTYL (PS00008, PROSITE documentation PDOC00008). В аминокислотной последовательности имеется 27 сайтов, и они расположены в следующих участках: 25–30 (glavsg), 30–35 (gakpnn), 72–77 (gtnpga), 197–202 (gggsss), 198–203 (ggssss), 199–204 (gssssa), 225–230 (gnsagm), 326–331 (gsangk), 400–405 (gglyak), 413–418 (ggngtw), 416–421 (gtwvnn), 428–433 (gvfrtk), 443–448 (gNtktv), 453–458 (gillsa), 476–481 (gqdvgy), 497–502 (gangni), 505–510 (gsaitw), 511–516 (gtnggn), 542–547 (galqns), 711–716 (ggaitt), 712–717 (gaittw), 759–764 (gaylss), 797–802 (glttta), 892–897 (gafisr), 1017–1022 (glgigi), 1021–1026 (gipmsq), 1104–1109 (gakpta). Большинство белков у эукариот ацилированы путем ковалентного присоединения миристана (C₁₄-насыщенной жирной кислоты) к остаткам, расположенным в N-конце. Чтобы произошло N-миристоилирование должны соблюдаться определенные условия: N-конец должен начинаться с глицина; во второй позиции должны быть незаряженные остатки; заряженные остатки, пролин и большие гидрофобные остатки не должны присутствовать; в 3 и 4 позиции могут находиться почти все аминокислотные остатки; в позиции 5 должны быть маленькие незаряженные остатки (Ala, Ser, Thr, Cys, Asn и Gly), лучше всего серин; в 6 позиции не должно быть пролина [12].

Заключение

Ген *gapA* в геноме бактерии расположен в пределах 223214...226585 пар оснований, является линейной ДНК длиной 3372 п.о., в пределах последовательности обнаружено 47 тандемных повторов длиной от 3 до 15 пар нуклеотидных оснований.

Белок GapA входит в семейство адгезин P1, имеет длину 1123 аминокислотных остатка, молекулярную массу равную 121.5 кДа и 6 групп по способу модифицирования последовательности из 76 мотивов. Структура белка содержит 5 доменов, области которых характерны для адгезинов, относящихся к семейству ADP1. По отношению к цитоплазматической мембране имеется вне-

клеточный участок белка, присутствие которого необходимо для успешного протекания клеточной адгезии.

Авторы выражают глубокую благодарность научному сотруднику лаборатории молекулярных основ патогенеза КИББ РАН, кандидату биологических наук О.В. Горшкову за обсуждение результатов исследования.

Summary

L.L. Frolova, A.Ya. Hidiyatullina. The Research of Gene *gapA* and Protein GapA *Mycoplasma gallisepticum* Primary Sequences.

The structures of *gapA* gene and its product, GapA protein of bacteria *Mycoplasma gallisepticum*, have been viewed in the article. Nucleotide and amino acid sequences have been retrieved from DNA, RNA and Protein databases, and both sequences have been processed by program packages TandemSwan, SOPMA, TMHMM. Special attention has been given to regions in the structure of protein which take part in cyte adhesion.

Key words: structure of *gapA* gene, structure of GapA protein, databases GenBank, GenPept, PFam, Prosite, program packages TandemSWAN, SOPMA, TMHMM, CPHmodels.

Литература

1. Razin S. Molecular properties of mollicutes: a synopsis / Eds S. Razin, J.G. Tully. // Molecular and diagnostic procedures in mycoplasmology. – San Diego. 1995. – V. 1. – P. 213–229.
2. Papazisi L., Troy K.E., Gorton T.S., Liao X., Geary S.J. Analysis of Cytadherence-Deficient, GapA-Negative *Mycoplasma gallisepticum* Strain R // Infection and immunity. – 2000. – V. 68, No 12. – P. 6643–6649
3. Goh M.S., Gorton T.S., Forsyth M.H., Troy K.E., Geary S.J. Molecular and biochemical analysis of a 105 kDa *Mycoplasma gallisepticum* cytoadhesin (GapA) // Microbiol. – 1998. – V. 144. – P. 2971–2978.
4. Papazisi L., Frasca S. (Jr.), Gladd M., Liao X., Yogeve D., Geary S.J. GapA and CrmA coexpression is essential for *Mycoplasma gallisepticum* cytoadherence and virulence // Infection and immunity. – 2002. – V. 70, No 12. – P. 6839–6845.
5. Pless D.D., Lennarz W.J. Enzymatic conversion of proteins to glycoproteins // Proc. Nat. Acad. Sci. USA. – 1977. – V. 74, No 1. – P. 134–138.
6. Gavel Y., von Heijne G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering // Protein Engin. – 1990. – V. 3, No 5. – P. 433–442.
7. Miletich J.P., Broze G.Jr. Beta protein C is not glycosylated at asparagine 329. The rate of translation may influence the frequency of usage at asparagine-X-cysteine sites // J. Biolog. Chem. – 1990. – V. 265, No 19. – P. 11397–11404.
8. Glass D.B., Smith S.B. Phosphorylation by cyclic GMP-dependent protein kinase of a synthetic peptide corresponding to the autophosphorylation site in the enzyme // J. Biolog. Chem. – 1983. – V. 258, No 24. – P. 14797–14803.
9. Kishimoto A., Nishiyama K., Nakanishi H., Uratsuji Y., Nomura H., Takeyama Y., Nishizuka Y. Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3':5'-monophosphate-dependent protein kinase // J. Biolog. Chem. – 1985. – V. 260, No 23. – P. 12492–12499.

10. *Pinna L.A.* Casein kinase 2: an 'eminence grise' in cellular regulation? // *Biochim. Biophys. Acta.* – 1990. – V. 1054, No 3. – P. 267–284.
11. *Patschinsky T., Hunter T., Esch F.S., Cooper J.A., Sefton B.M.* Analysis of the sequence of amino acids surrounding sites of tyrosine phosphorylation // *Proc. Nat. Acad. Sci. USA.* – 1982. – V. 79, No 4. – P. 973–977.
12. *Towler D.A., Gordon J.I., Adams S.P., Glaser L.* The biology and enzymology of eukaryotic protein acylation // *Ann. Rev. Biochem.* – 1988. – V. 57. – P. 69–99.

Поступила в редакцию
14.02.08

Фролова Людмила Леонидовна – кандидат технических наук, доцент кафедры генетики Казанского государственного университета.

E-mail: *Lucy.Frolova@ksu.ru*

Хидиятуллина Айгуль Ядкарьевна – выпускник кафедры генетики Казанского государственного университета.

E-mail: *kindness2006@yandex.ru*