Mathematical Biology and Bioinformatics. 2013. V. 8. № 2. P. t12–t26.
URL: http://www.matbio.org/2013/Panyukov_8_t12.pdf.

Original Russian text
Panyukov V.V., Kiselev S.S., Shavkunov K.S., Masulis I.S., Ozoline O.N., 2013 published in Matematicheskaya biologiya i bioinformatika, 2013, V. 8. № 2. P. 432–448.
URL: http://www.matbio.org/2013/Panyukov_8_432.pdf.

========================= **BIOINFORMATICS** ==========================

UDC: 579.252

Mixed promoter islands as genomic regions with specific structural and functional properties

Panyukov V.V.[1], Kiselev S.S.[2,3], Shavkunov K.S.[2,3], Masulis I.S.[2,3], Ozoline O.N.[2,3*]

[1]*Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Pushchino, Moscow region, 142290, Russia*

[2]*Institute of Cell Biophysics, Russian Academy of Sciences, Pushchino, Moscow region, 142290, Russia*

[3]*Pushchino State Institute of Natural Sciences, Moscow region, 142290, Russia*

Abstract. The study is aimed at unusual genomic regions of *E. coli* – *mixed promoter islands*, which possess high density of potential transcription initiation points. They were predicted by promoter finder PlatPromU, without the consideration of conservative elements, recognized by σ-subunits of RNA polymerase, thus they are supposed to interact with polymerase, which has different σ-factors. We found that the structural and functional properties of *mixed promoter islands* are very similar to those of *promoter islands*, previously revealed by the algorithm PlatProm, which was adjusted to the search for σ^{70}-specific promoters of *E. coli*. Thus, both types of *islands* are preferentially located in the regulatory regions of genes acquired by horizontal transfer. Their double helix is characterized by a higher degree of curvature and twisting, compared to normal promoters. Both types of *islands* have a higher capacity to form complexes with RNAP, than normal promoters, but lower ability to initiate productive transcription. This suppression can be largely explained by the advanced ability of *islands* to interact with the histone-like inhibitory protein H-NS. However, RNA synthesis from the *promoter island* associated with the gene *appY*, increased upon its transfer into the plasmid pET28b-eGFP and showed dependence not only from the presence of H-NS binding sites, but also from the spatial configuration of the neighboring areas. That means that conformational changes in the genome can increase the transcriptional activity from the *islands*, delivering the products of adjacent genes for metabolic needs of the bacterial cell.

Key words: *promoter islands, DNA structure, transcription initiation, H-NS, horizontal gene transfer, bacterial evolution.*

INTRODUCTION

Bacterial gene expression is carried out by RNA polymerase, containing exchangeable σ-factors, and is controlled by regulatory proteins that cooperate in the transcription initiation by interacting with specific genomic regions – promoters. Most promoters are located in front of genes, though transcription initiation signals can also be found within the gene coding

*ozoline@rambler.ru

sequences, where they may be involved in the synthesis of antisense or alternative RNAs [1–4]. Though the biological significance of many of such RNAs is not yet clear, their synthesis from intragenic regions is confirmed by modern experimental techniques, which make it possible to map the transcription start points (TSPs) directly in the living cell [5–7]. Except for the two abovementioned types of promoters, having one or several transcription initiation points, all the bacterial genomes investigated so far contain areas anomalously enriched by potential transcription starts [1, 8, 9]. Thus, for instance, our promoter finder PlatProm, configured to search for σ^{70}-promoters in the chromosome of *E. coli*, found 78 regions, which are at least 300 bp long and contain no less than 8 potential TSPs in every running window of 100 bp. We named them "*promoter islands*" or PIs [1, 9]. Analysis of the published genome-scale data on RNAP binding [10, 11] and transcription initiation sites [6] indicated that all PIs interact with RNAP [1] and initiate the synthesis of short abortive RNAs [9]. I. e. they are able to perform the first stage in the process of transcription initiation, but the synthesis of full-sized RNA products is interrupted, and RNAP does not leave the promoter. Moreover, the structural analysis of 3D geometry, carried out by the software package aSHAPE [12], revealed higher curvature for *promoter islands* as compared to normal promoters, which formed anisotropic bends of the DNA double helix more often than non-promoter DNA fragments [9]. This excessive curvature may stabilize RNAP-promoter complexes in the *islands*, holding the enzyme in the state of abortive transcription.

In addition, we found that an average twist angle was similar for non-promoter DNAs and normal promoters, but in the case of *promoter islands* it was noticeably higher [9]. That means that PIs are evolutionary optimized for some additional function besides the interaction with RNAP. Guided by this assumption, we analyzed the available profiles of distribution for nucleoid proteins along the bacterial chromosome [13–17], and found that PIs provide preferential target sites for interaction with H-NS [9]. It is well known that H-NS is one of the main proteins, participating in the condensation of the bacterial genome. At the same time, H-NS specifically represses the expression of horizontally acquired genes through binding with their regulatory regions. Hence, we checked the positional association of PIs with foreign DNA and found that 75 out of 78 *promoter islands* (96%) are indeed located nearby alien genes. However, the expected number of horizontally acquired genes in the genome of *E. coli* is an order of magnitude higher, than the amount of PIs. For example, Lawrence *et al.* [18] have found 706 foreign genes in this genome, while the evaluations of Nakamura *et al.* [19] and Price *et al.* [20] assume that the number exceeds one thousand. But if the high density of potential promoters is a specific feature of horizontally acquired genes, it might be achieved by the presence of σ^{70}-promoters together with promoters of other σ-factors. Therefore, in order to estimate the contribution of additional promoters, we scanned the *E. coli* genome by a unified version of promoter finder PlatPromU [21], which ignores the weight matrices, accounting the σ^{70}-specific modules, and can find promoters, recognized by all the seven σ-subunits of *E. coli*. If our expectation is true, then the difference between the number of *promoter islands* and alien genes may decrease. In present work we identified 434 *mixed promoter islands* (MPIs), estimated the degree of their association with foreign genes and characterized their structural and functional properties.

METHODS

Genomic DNA of *E. coli*, promoters and *promoter islands*

The nucleotide sequence of the genomic DNA of *E. coli* K12 MG1655 and the corresponding gene map were taken from the NCBI GenBank (NC_000913). Both strands were scanned by the promoter finder PlatPromU [21], revealing genomic regions with typical for promoters distribution of structure-specific modules and predicting TSPs. Potential start points were considered as reliable sites of transcription initiation if they have PlatPromU scores exceeding the background level for at least 4 standard deviations (StD), while the background level and StD were estimated as described in [21]. The criteria used to select *mixed promoter islands* were the same, as suggested previously for PIs (see above), except for the minimal length, which was increased up to 360 bp. As a result we obtained a novel set, containing 434 MPIs, including all the PIs analyzed previously. The properties of MPIs were compared with those of 351 experimentally studied regulatory regions of the *E. coli* genome, containing single or multiple promoters, which do not overlap with MPIs and contain at least one promoter recognized by σ^{70}-RNAP. Coordinates of their transcription start points were taken from RegulonDB [7].

Association of MPIs and normal promoters with horizontally transferred genes

To analyze the relative disposition of MPIs or normal promoters in respect to foreign genes, we used the predictions made by 5 independent research groups [18–20, 22, 23]. Promoter regions of both categories were considered as associated with foreign genes, if they lie within long "*genomic islands*" predicted by GIST [22] or IslandViewer [23], as well as if they are located within the regulatory regions of single foreign genes [19–21]. MPIs were also considered as associated with foreign genes, if they lie within their coding sequences or overlap with them for at least 100 bp.

Chromatin immunoprecipitation data analysis

The ability of RNAP and H-NS to interact with MPIs was assessed using published chromatin immunoprecipitation data [10, 13–15]. This method (ChIP-on-chip) allows detection of specific protein binding sites in the entire bacterial genome in the living cell. For this purpose its complexes with DNA were extracted from isolated and sonicated chromatin by specific antibodies (ChIP), while DNAs recovered from precipitated complexes are hybridized with microarrays (chip). In the case of RNAP [10], the ratio of hybridization signals obtained with DNA co-immunoprecipitated with the enzyme by σ^{70}-specific antibodies and the control DNA recovered from the complexes without immunoprecipitation (experiment B in ref. [10]), was calculated and expressed as $\log_2$. If the average value of this parameter for all probes within MPI or within ±150 bp area nearby the TSP(s) of normal promoter(s) was positive, the genomic region was considered as interacting with RNAP. The efficiency of interaction with H-NS was evaluated using two data sets published by Kahramanoglou *et al*. [13] and Grainger *et al*. [14, 15]. In the former case the analyzed genomic regions were considered as targets for interaction with H-NS, if they overlap with the published binding sites for at least 20 bp, while in the later case – if they contained at least one probe with hybridization signals ratio ≥ 1.5.

Differential expression analysis

Transcriptional activity of MPIs and normal promoters was evaluated exactly as described in [9]. In brief: raw 5'-end-specific RNA-seq data (supplement in ref. [6]), containing 8967903 sequence reads (samples), each 44 bp long, were analyzed by the software RNAMatcher, which ascribes the registered samples to matching genomic regions and determines their total number. Samples fully identical to the sequences of genomic DNA (44 bases in length) were considered as representatives of productive transcription. Samples, which have only 11 or 10 nucleotides at the 5'-end, coinciding with the sequences of genomic DNA, and contain an adapter at the 3'-end, were considered as products of abortive initiation.

Structural analysis

To compare the structural properties of MPIs with those of normal promoters, we used the software package aSHAPE [12, 9], which analyzed the virtual 3D models of 300 bp long DNA fragments, provided by the server *DNA Tools* [24]. The software calculated the structural parameters of the DNA double helix, using two types of conformation chains: *carbon* and *phosphorous*. The midpoints of the straight lines connecting C_6 of pyrimidines with C_8 of purines in each complementary base pair gave vertices of a *carbon* chain, while the midpoints of segments connecting phosphorus atoms of base pairs gave vertices of a *phosphorous* chain (for detailed description see [9, 12]). The stacking energy for DNA fragments was calculated as a sum of stacking energies computed for their constituent dinucleotides. The values of this parameter were taken from the table of DiProDB [25, 26].

Functional analysis of *appY* associated *promoter island*

Three DNA fragments containing different parts of *promoter island*, covering the space between *tfaX* and *appY* genes, as well as the coding sequence of *appY* in the genome of *E. coli* MG1655, were amplified using primers

5′-GATAAGATCTGCAAGTAAAAATGATACTC-3′ (F1),
5′-ATGCAGATCTTCCTGATTATGATTGTG-3′ (F3),
5′-TCCATCTAGAACCTATCATAAAATTA-3′ (R1) и
5′-CCCTTCTAGATTTGTCGCTTACAATAAA-3′ (R2).

Amplicons were digested by *Bgl*II (restriction site AGATCT) and *Xba*I (TCTAGA) and ligated into plasmid pET28b between *Bgl*II and *Xba*I restriction sites upstream of the gene *gfp*, encoding green fluorescence protein. All enzymatic reactions were carried out using Fermentas enzymes according to the manufacturer’s protocols. GFP fluorescence was measured in the colonies of *E. coli* OmniMAX cells transformed with this plasmid and grown on standard Luria-Bertani plates supplied with 60 μg/ml kanamycin. Cells transformed with the plasmid with promoterless *gfp* were used as a control. Fluorescence was measured on fluorescent microscope Leica (excitation/emission wave lengths 480/510 nm) and computed by ImageJ software.

RESULTS

PlatPromU revealed 434 *mixed promoter islands* in the genome of *E. coli*

PlatProm assesses the probability for a given genomic position to be a transcription start point of σ^{70}-specific promoters. It operates with position weight matrices, which score the contribution of hexanucleotides, recognized by σ^{70}, and with more than 50 cascade matrices, which take into account the specific structural modules, favoring adaptive isomerization of promoter DNA in the transcription complex. The evolutionary conservatism of the transcription machinery offers an opportunity to find promoters recognized by different σ-factors using the unified program PlatPromU [21], which operates only with cascade matrices. Using the score threshold, providing $p < 0.00004$ (defined as described in [21]), we found several thousand of potential promoter regions, including those, which were found by PlatProm. Following the criteria described in [1] and [9] (≥ 8 TSPs on either DNA strand in a running window of 100 bp along at least 300 bp) we identified 600 MPIs. All 78 previously defined PIs were found in this set. Their minimum length was 360 bp, i.e. promoters with different σ-specificity can be found in one MPI. To make this set comparable in size to the compilation of normal promoters, we increase the cut-off length for MPIs up to 360 bp, resulting in 434 *mixed islands*. The maximum length of these *islands* was 7598 bp, while an average size - 661 bp (Table 1).

Table 1. Genomic coordinates of *mixed promoter islands* in the genome of *E. coli* K12 MG1655, their length and association with horizontally transferred genes

items 1-109			items 110-218			items 219-327			items 328-434		
position	lenght	data source*	position	lenght	data source*	position	lenght	data source*	position	lenght	data source*
17025	501	--n--	1278619	452	-----	2283999	462	-----	3530410	425	----p
29065	639	-----	1292968	425	--nl-	2301499	404	--n--	3537787	379	-----
41821	610	----p	1297308	672	-----	2310651	867	-----	3550522	598	-----
58659	756	--nl-	1298357	680	-----	2341466	522	g-n-p	3579578	1910	g-nlp
83820	563	g-n--	1307512	826	--nl-	2342045	736	g-n-p	3582164	742	g-nl-
85243	443	g----	1314018	515	--nl-	2362246	433	g---p	3595538	462	-----
89011	649	--n--	1332751	426	-----	2362997	370	g-nlp	3620924	1702	g-nlp
121653	405	----p	1341256	362	---lp	2363578	445	g-n-p	3629188	731	g-nl-
154703	620	ginl-	1358825	383	--n--	2380550	690	ginl-	3630044	1023	g-nl-
156811	519	g-nl-	1369587	403	--nlp	2381326	406	ginl-	3631295	994	g-nl-
233981	381	ginl-	1388697	372	----p	2383618	405	ginlp	3632356	734	g-nlp
237008	491	ginl-	1390680	598	--nl-	2384587	506	ginlp	3648820	598	g-nl-
238001	939	ginl-	1393485	497	--nlp	2385450	784	-inlp	3649513	733	g-nl-
252508	367	--nl-	1410853	372	--nl-	2386354	570	-inl-	3651226	827	g-nlp
252957	438	--nl-	1411493	681	-inl-	2403103	645	-----	3652495	539	g-nl-
259058	463	-----	1421610	582	-inl-	2404930	644	----p	3653800	731	g-nlp
262055	441	-----	1422963	403	-inlp	2411059	362	--n-p	3654706	376	g-nlp
284280	377	-i-lp	1431570	1668	g-nl-	2453522	759	g----	3655630	1258	--nl-
291978	892	-inlp	1434606	767	g----	2460800	584	--nl-	3662512	421	--nl-
292911	1453	-inlp	1462958	512	g---p	2461780	627	-inl-	3663561	687	--nl-
296057	553	-in-p	1473307	403	--nlp	2466817	1773	ginlp	3667209	418	-----
310143	921	-in-p	1500175	528	--n--	2474418	389	--n--	3669849	380	--n--
312157	474	-i---	1518777	502	--n--	2478590	648	-inl-	3694146	368	--nl-
312947	678	-inlp	1524817	1053	g-nlp	2479567	521	-inl-	3694645	392	--n--
317665	361	-inlp	1527769	969	g-nlp	2480771	370	-inl-	3718242	504	--nl-
320379	534	--n-p	1529011	425	g-nl-	2481225	1105	-inl-	3735006	416	--n-p
324288	438	--nl-	1529586	478	g-nl-	2482830	438	-inl-	3743785	493	----p
330759	792	--n--	1541737	2674	g-nl-	2483388	999	-inl-	3749670	812	-----
343046	929	--nl-	1565114	468	----p	2488645	405	-inlp	3752332	496	----p
345477	560	--nl-	1570010	526	----p	2491166	563	---lp	3755772	430	g----
347632	374	-----	1577222	1414	ginlp	2492456	376	--nl-	3764123	1099	g-nlp
383775	742	--nlp	1580426	586	ginlp	2494555	496	-----	3765749	1000	g-nl-
389004	483	--nlp	1581411	908	ginl-	2522857	412	----p	3767002	1081	g----
393590	559	--nl-	1584482	418	g-nl-	2531270	420	----p	3787153	479	--n--
400274	418	--n--	1596053	540	--n--	2557957	572	-inl-	3787887	495	--n--
418266	507	----p	1613695	524	--nl-	2559897	383	-inl-	3790880	737	ginl-
450848	492	-----	1622416	502	--n--	2576296	415	-----	3794844	3642	ginlp
479168	1335	--nl-	1630172	1610	--nl-	2588895	523	-----	3798515	2750	ginlp
522049	488	g-nlp	1634966	1101	ginl-	2626549	428	g-n--	3801914	1517	ginlp
526294	897	g-nlp	1636533	609	gi-l-	2627521	404	g-nl-	3805844	756	ginlp
527695	809	g-nl-	1638669	1326	ginlp	2651326	565	-----	3834585	419	--n--
544455	397	----p	1640088	460	-----	2698323	384	-----	3841692	479	----p
545590	425	----p	1644594	386	-inl-	2755495	421	ginl-	3873193	420	--nl-
557105	399	-inlp	1650572	653	--nl-	2757618	1049	ginlp	3886238	583	-----
557872	419	-inlp	1671208	476	-----	2763165	415	ginlp	3903481	399	--n-p
562926	442	-inlp	1702736	415	-----	2771215	2017	ginl-	3904506	410	--n-p
567750	843	ginl-	1710105	803	-----	2781386	816	ginl-	3920483	735	--n--
569055	1132	ginl-	1752558	436	g-n-p	2782245	1079	ginl-	3965542	466	--n--
570511	709	ginl-	1753096	442	g---p	2783819	1101	ginl-	3983964	587	----p
576007	594	ginl-	1762391	516	-----	2785550	638	ginl-	4000416	1011	--nl-
578302	708	ginl-	1768241	504	--n-p	2786386	627	ginl-	4041929	549	--nl-
582252	1804	ginlp	1771642	382	----p	2796362	863	--n-p	4044492	484	-----
584711	518	ginl-	1776585	424	--nlp	2802496	503	-----	4060020	527	--n-p
585353	445	ginl-	1785089	419	-----	2882114	524	ginlp	4067116	514	----p
635732	457	--nl-	1789980	390	--n--	2884888	475	ginlp	4076478	908	--nl-
636502	592	--nl-	1800863	739	--n--	2885512	362	ginlp	4077482	374	--nlp
637951	427	-----	1802156	560	-----	2898106	794	g---p	4090814	387	--n-p
650912	719	----p	1810478	976	--nl-	2901483	610	-----	4158769	373	----p
655045	929	--n-p	1816517	425	----p	2902283	586	g----	4218284	571	g-nl-
659421	423	--n--	1818909	550	----p	2903410	533	g-nlp	4219093	391	g-nl-
660408	703	--n--	1823589	468	-----	2925686	484	-----	4219964	827	g-nl-
675647	439	--nlp	1829942	464	-----	2931729	572	--n-p	4233455	551	--n--

678496	386	--nlp	1842543	494	--nl-	2966921	511	--n--	4240158	413	-----
707074	530	-----	1851755	574	--nlp	2985071	1571	g-nl-	4248582	1975	g-nlp
715656	671	--nlp	1855560	418	--nl-	2986811	7598	g-nlp	4257986	1081	g-nl-
719924	438	--nlp	1868202	750	g-nl-	2996542	498	g-nl-	4259212	577	g-nl-
727973	490	-i---	1876742	368	-----	3003014	538	--nl-	4266343	1135	g---p
732771	637	ginlp	1891663	381	--nl-	3003731	578	--nl-	4273131	396	--n--
735163	1053	ginl-	1903183	360	--n-p	3013557	605	--nl-	4279950	1329	g-nlp
736988	551	ginl-	1905775	445	--nl-	3028904	410	--n-p	4285394	479	g-n-p
751962	520	g-nl-	1943986	454	--nlp	3048753	438	-----	4304478	404	-----
769872	741	-----	1956024	424	--n-p	3065024	392	--n--	4308668	419	----p
799381	712	--nlp	1976097	428	--n-p	3077172	505	--nlp	4310980	501	--nlp
819812	530	--nlp	1977264	482	--nlp	3085926	405	----p	4324734	462	--nlp
848056	446	-----	1984126	905	g-n-p	3098690	453	-----	4335751	656	--nlp
849273	490	-----	1993347	545	-----	3117038	617	----p	4338464	369	----p
871972	643	-----	2009130	1425	--nl-	3131908	427	--nlp	4346775	605	----p
872825	395	--nl-	2021494	489	--nl-	3134231	539	-----	4358053	481	--nl-
915205	531	----p	2022180	452	--nl-	3169846	625	--nl-	4359906	426	--nl-
931240	410	-----	2031411	824	g-n--	3181498	377	-----	4366380	492	----p
953644	517	g----	2037211	378	--nl-	3183123	473	g--lp	4371868	728	-----
985994	818	-----	2039564	542	--nlp	3187814	502	-inl-	4382328	362	--n-p
996712	487	--nl-	2040188	415	--nl-	3188374	629	-inlp	4407987	664	--nl-
1030860	382	----p	2042317	684	--nl-	3189868	543	-inlp	4435491	1309	g-nl-
1049790	450	--n--	2050929	462	g----	3214483	377	----p	4460668	462	--n--
1050283	554	--nl-	2054529	684	ginlp	3217041	447	----p	4473278	2052	ginl-
1063159	690	--nl-	2055487	815	ginlp	3232171	384	--nl-	4477500	1195	g-nlp
1091438	1604	-inl-	2066210	527	-inlp	3250683	426	g-n--	4495848	478	-inlp
1094857	419	-inl-	2083352	399	----p	3255981	383	g---p	4501810	502	ginlp
1095707	407	--nl-	2085039	380	-----	3261326	390	--n--	4502360	1553	ginlp
1102438	908	g-nlp	2100942	1725	ginlp	3264009	385	g-nlp	4504233	479	ginlp
1107417	415	--n-p	2103095	2563	ginlp	3264982	708	g-nl-	4522961	485	-inlp
1118150	513	--nl-	2105951	1900	ginlp	3265752	1971	g-nl-	4523535	388	-inlp
1160698	386	--n-p	2111096	435	ginl-	3272800	721	g---p	4537372	477	ginlp
1165013	367	-----	2112807	373	-----	3281941	363	----p	4538360	681	ginlp
1167982	372	--n--	2128041	443	--n-p	3284977	664	g-nlp	4539534	554	ginl-
1195484	481	ginlp	2135166	769	-----	3352044	528	-----	4540509	690	ginlp
1196064	1971	ginlp	2140999	369	----p	3358701	551	-inl-	4553289	468	g-nl-
1209833	854	ginlp	2148871	468	--nlp	3359881	379	-inlp	4554206	740	g-nl-
1210772	1359	ginl-	2166143	472	--nlp	3372534	441	----p	4569842	730	--nl-
1214579	1748	ginl-	2176533	370	----p	3375472	381	-----	4574875	960	-inlp
1215406	921	ginlp	2185400	550	-inl-	3382306	540	----p	4578345	1119	ginlp
1218508	500	-----	2188911	552	ginl-	3383191	482	---l-	4589314	545	-----
1219391	392	-i---	2189551	1054	ginl-	3411440	518	--n--	4592666	405	--n-p
1222160	538	--nlp	2202149	512	----p	3416118	360	-----	4600907	689	--nl-
1228554	498	--nl-	2226735	446	--n-p	3416614	402	--n--	4638587	410	-----
1229535	406	--n--	2231648	435	--nlp	3453341	545	--nl-	4639554	405	-----
1254990	742	--n--	2257132	653	--n-p	3467648	565	-i--p			
1272774	392	----p	2267416	665	--n--	3510244	373	--n-p			

*The symbols **ginlp** mark data source:
g – «*genomic islands*», found by GIST [22] (474 genes in the gene map of *E. coli* K12 U00096.2 according to the last RegulonDB annotation (ver. 8.0) [7]);
i – «*genomic islands*», predicted by IslandViewer [23] (486 genes of the same genome);
n – foreign genes found by Nakamura *et al.* [19] (1041 genes of the same genome);
l – foreign genes found by Lawrence and Ochman [18] (714 genes of the same genome);
p – foreign genes found by Price *et al.* [20] (1049 genes of the same genome).
Islands that are not associated with foreign genes, as well as *islands* associated with genes, whose horizontal transfer was predicted only in one data set, are bolded.

Mixed promoter islands are associated with foreign genes

It was found that 370 out of 434 *mixed promoter islands* (85%) are associated with potentially foreign genes predicted in at least one study [18, 19, 22, 23], and for 283 *islands* this association is confirmed by at least two data sets (Table 1). The most significant correlation was observed with the list of foreign genes deduced by Lawrence and Ochman [18] on the basis of base composition and codon usage pattern. The number of MPIs associated with foreign genes in this case was 3.6-fold higher than expected by chance. The worst correlation was found for the set predicted by Price *et al.* [20] on the basis of phylogenetic analysis. In this case the set of MPIs overlapping with presumably foreign genes was only 1.8-fold higher compared to random distribution. Since this data set poorly correlated with the other lists of potentially foreign genes, we did not use it in the next experiment, aimed to determine the percentage of foreign genes associated with *mixed promoter islands* and normal promoters. This decreased the portion of MPIs, associated with foreign genes, to 76% (Table 1 and Fig. 1), but raised the degree of overlap in the general set of foreign genes, thus increasing the reliability of comparative analysis.

Fig. 1 schematically demonstrates the mutual overlap between 351 normal promoters or 434 *mixed promoter islands* with 1513 foreign genes revealed by four independent research groups [18, 19, 22, 23]. It is obvious, that normal promoters typically control the expression of cognate genes. Only 23.9% of them are located within the regulatory regions of foreign genes, which is more than 3 times less than in the case of MPIs.

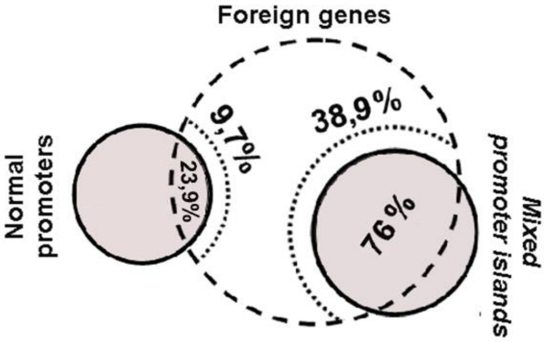

Fig. 1. The relative number of normal promoters and MPIs (shaded circles on the left and right, respectively), as well as their degree of overlap with a common set of foreign genes predicted in [18, 19, 22 and 23] (dashed circle). The dotted segments dissect the portion of foreign genes associated with the test promoter regions (genes transcribed divergently from the common promoter region and genes transcribed as a polycistronic unit were taken into account).

For all that, only 38.9% of foreign genes are associated with *islands*. Thus, even though horizontally acquired genes tend to accumulate promoter-like sites [22], less than half of them are located adjacent to or overlapping with *promoter islands*. Moreover, about 7% of the foreign genes revealed by all research groups [18–20, 22, 23] do not have PlatPromU-predicted promoters at all. Perhaps this unhomogeneity reflects the dynamics of evolutionary adaptation of horizontally acquired DNA to the genome of *E. coli*. If *promoter islands* are indeed an evolutionary tool for assimilation of foreign genetic material, recently acquired genes may be at the stage of their accumulation, while the promoters of genes that are already integrated into the regulatory network of the novel host, in contrast, may be a subject of "purifying" evolution, which removes excessive promoter sites.

Functional properties of 434 *mixed promoter islands* are similar to those of 78 *promoter islands*

Previously, we have evidenced that all 78 *promoter islands* found by PlatProm interact with RNAP [1], but instead of full-length RNA synthesis they usually produce only short oligonucleotides [9]. Circle 4 in Fig. 2 demonstrates that *mixed islands* also form transcription complexes slightly better than normal promoters: interaction with RNAP has been registered for 86% and 82%, respectively. However, only 149 MPIs (34%) produce RNAs ≥ 44 nucleotides long, which is significantly less than in the case of normal promoters from our set (56%). Moreover, the total number of RNA products, matching the MPIs, was 2-fold lower than expected by chance, while in the case of normal promoters the number of samples matching to the ± 50 bp regions around TSPs was, by contrast, 27-fold higher.

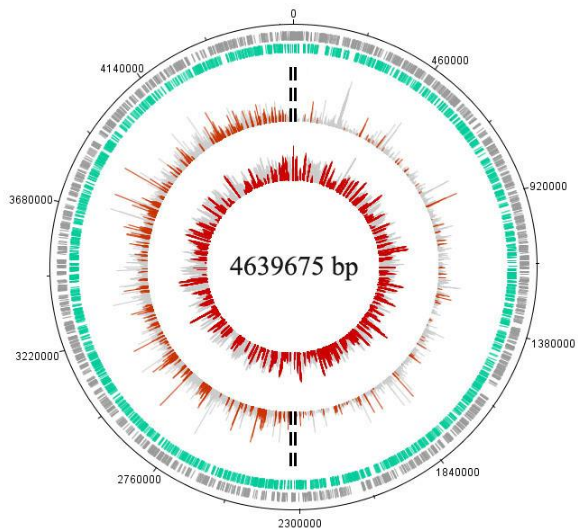

Fig. 2. Distribution of functional sites in the genome of *E. coli* K12 MG1655. **Two outer circles:** gene map of the strain in both strands. **Circle 3:** relative number (N) of 10–11-nucleotide RNAs (left semicircle) and ≥ 44-nucleotide RNAs (right semicircle), registered by Dornenburg *et al*. [6] and plotted as $\log_{10}(N+1)$. **Circle 4:** distribution of RNAP binding sites, as registered by Reppas *et al*. [10], expressed as $\log_2$ of the ratio of specific and control hybridization signals. The number of RNA products and hybridization signals, corresponding to the *mixed promoter islands*, are shown in red.

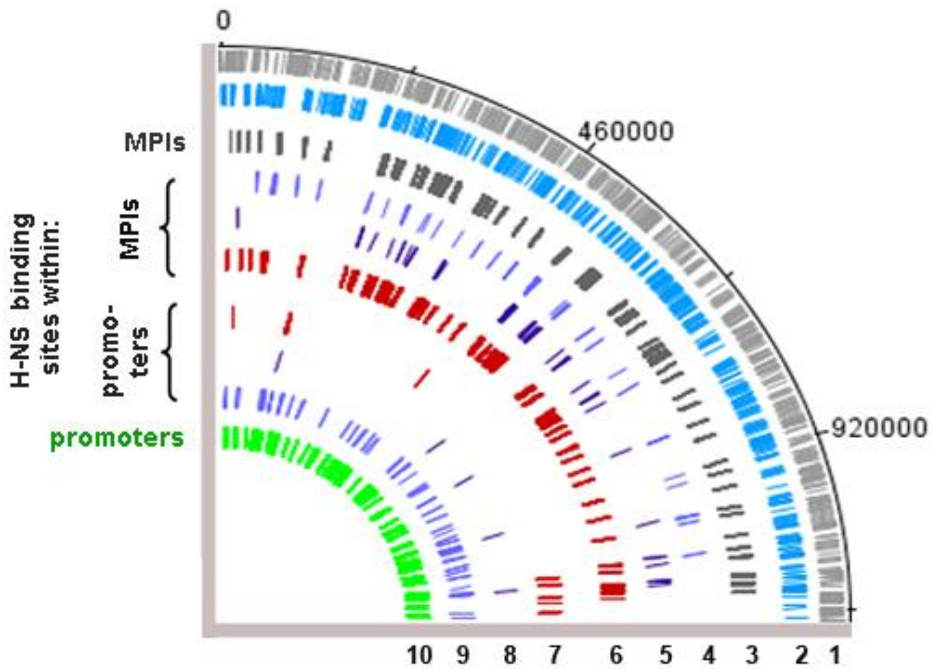

Fig. 3. Distribution of *mixed promoter islands* (quarter-circle **3**) and normal promoters (quarter-circle **10**) in the first quarter of the *E.coli* genome. The gene map on both strands of the bacterial chromosome is shown on quarter-circles **1** and **2**. H-NS binding sites registered within MPIs or nearby normal promoters (± 250 bp around the transcription start point) in the genome of bacterial cells, grown under standard physiological conditions in Luria-Bertani medium [13, 16], are shown on quarter-circles **6** and **7**, respectively. H-NS binding sites overlapping with MPIs or normal promoters within cells grown in M9 medium + fructose [14, 15], are shown on quarter-circles **5** and **8**, respectively, and within cells grown in M9 medium + fructose + salicylic acid [14] – on the quarter-circles **4** and **9**, respectively.

The red and grey bars on the right part of circle 3 in Fig. 2 reflect this difference. At the same time, 97% of MPIs and only 71% of normal promoters produced abortive RNAs (red and gray bars on the left part of the third circle). Thus, the set of 434 MPIs shows the same pattern of transcription output as the set of 78 PIs [1, 9]. The same similarity was observed for interaction with histone-like protein H-NS (Fig. 3), which is known as a specific repressor of horizontally acquired genes [27–29].

Fig. 3 shows H-NS binding sites, which were registered by the ChIP-on-chip technique [13–16] and overlap with *mixed promoter islands* (quarter-circles 4–6) or normal promoters (quarter-circles 7–9). We found that at standard growth conditions 425 of the 434 MPIs (97.9%) interacted with H-NS (quarter-circles 5 and 6). The percentage of normal promoters associated with H-NS at the same conditions was much less (21.5%, quarter-circles 7 and 8). This difference almost disappeared under acid stress, when the bacterial genome is subjected to protective condensation (quarter-circles 4 and 9). I.e. most *promoter islands* are maintained by the cell in a heterochromatin-like state.

Structural properties of *mixed promoter islands* may contribute to their condensed state

It has been observed previously [9] that structural properties of 78 *promoter islands* differ from those of normal promoters. *Islands*, revealed by PlatProm, were more bent; their models contained more non-planar triplets, and the double helix was "overtwisted". These peculiarities may certainly contribute to the specific functional properties of *promoter islands*. In this case, MPIs with very similar to PIs functional behavior also have to possess the same structural properties. In order to verify this possibility, MPIs that contain 78 PIs were removed from the analyzed set. The remaining genomic regions were used to compose a set of 364 non-overlapping DNA fragments 300 bp long. Their 3D-models were created on the *DNA Tools* server [24] and compared with 363 DNA fragments of the same size, containing sequences of normal promoters, and with non-promoter DNAs (the control set used previously in [9]). Fig. 4–6 demonstrate that 364 representatives of MPIs have the same structural properties as the previously characterized 78 *promoter islands*.

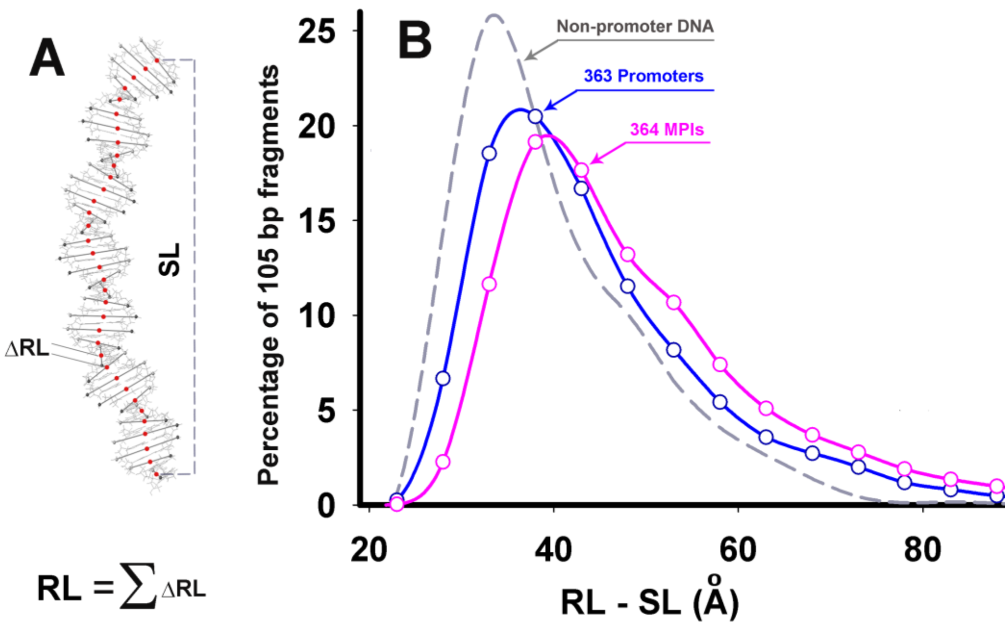

Fig. 4. A: An example of the virtual model created by the *DNA Tools* software [24]. *Phosphorous conformational chain* is shown by red dots. **B:** Histograms of distribution of $(\mathbf{RL\text{-}SL})_{105}$, quantified for all 105 bp sub-fragments for the three analyzed sets (interval of partitioning for $(\mathbf{RL\text{-}SL})_{105}$ - 5Å).

Thus, to assess the global curvature of the DNA double helix we used the conformational parameter **RL-SL**. Its quantification is illustrated in Fig. 4,A. The numerical values were calculated for all fragments 105 bp long taken from each sample with 1 bp shift. The histograms of distribution of these values for all families are shown in Fig. 4,B. They are clearly different and the average values of $(\mathbf{RL\text{-}SL})_{105}$ increases in the following order: non-promoter DNAs (37.7Å) < normal promoters (42.6Å) < MPIs (46.6 Å). This assumes a higher capacity of MPIs to interact with nucleoid proteins, which prefer to bind curved DNAs.

In full compliance with a previous analysis [9], the average value of the angle $\mathbf{\Theta}$ between two lines connecting vertex $\mathbf{v_i}$ with $\mathbf{v_{i+1}}$ and $\mathbf{v_{i+1}}$ with $\mathbf{v_{i+2}}$ (insert in Fig. 5) was almost equal in all the three sets (5.6°, 5.6° and 5.5°), while the values of the torsion angle $\boldsymbol{\varphi}$ (insert in Fig. 5) showed specific distributions (Fig. 5,A). Noticeably lowered variability of $\boldsymbol{\varphi}$ in the set of MPIs (Fig. 5,B) is of special interest. We have already observed this feature earlier [9]. It assumes some regularity in the 3D trajectory of DNA, which may be favorable for the assumed condensation within *promoter islands*.

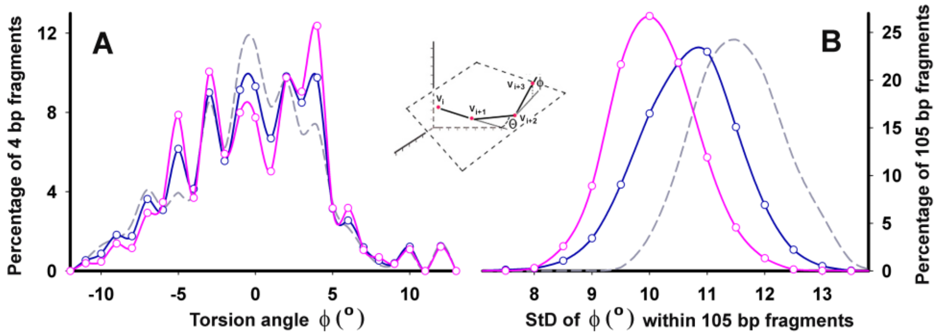

Fig. 5. Distribution (**A**) and variability within 105 bp sub-fragments (**B**) of the torsion angle $\boldsymbol{\varphi}$ (schematically shown in the insert) in the virtual models of 364 MPIs (magenta curves), 363 normal promoters (blue curves), and non-promoter DNAs (dashed gray plots), measured for *carbon* (**A**) and *phosphorous* (**B**) conformational chains. The interval of partitioning for $\boldsymbol{\varphi}$ – 1° (**A**), for StD of $\boldsymbol{\varphi}$ – 0.5° (**B**).

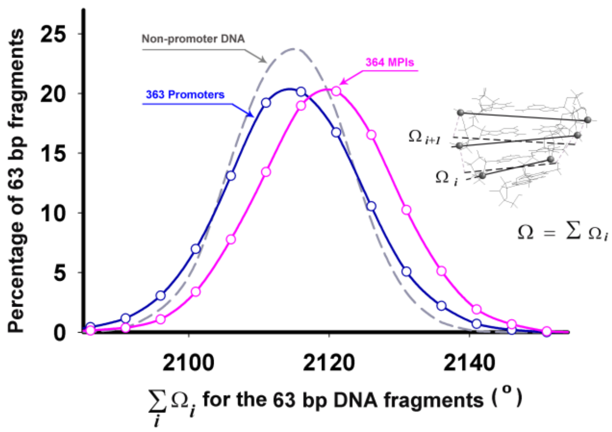

Fig. 6. Distribution of cumulative twist angles $\mathbf{\Omega}$, calculated by *phosphorus chains*, for fragments of 63 bp in the three analyzed families. Partitioning interval was 5°.

One of the most discriminative structural properties of *promoter islands* were the high values of the cumulative twist angle $\mathbf{\Omega}$ (schematically shown in the insert of Fig. 6), which was not typical for normal promoters [9]. Histograms in Fig. 6 show the distribution of the cumulative angles $\mathbf{\Omega}$ for 63 bp long fragments, demonstrating the same feature for MPIs. Even though the cumulative twist angle in this family is only 5° larger than in normal promoters, the excessive twisting of the double helix may impede local DNA melting during transcription complex formation, thus contributing to the suppression of RNA synthesis. If this is the case, transcriptional activity of *promoter islands* should increase in the genomic environment with a higher level of negative supercoiling, compared to the genomic DNA.

Promoter island, associated with the gene *appY*, can initiate synthesis of *gfp*-mRNA

In order to assess the dependence of promoter activity on genomic environment, we chose the *island* associated with the gene *appY* (Fig. 7,A). Its ability to initiate the synthesis of long RNAs has not been registered previously by high throughput techniques [6, 10], assuming that the promoter activity in this PI is suppressed in its natural environment. Three different fragments of this *island* were incorporated into the plasmid pET28b-eGFP in front of the promoterless gene of green fluorescence protein (*gfp*). *E.coli* cells transformed with these plasmids accumulated the reporter protein (Fig. 7,B–D), i.e. promoters of the *appY*-associated *island* can act as normal promoters.

The relative fluorescent intensity for experimental (F_e) and control (F_c) cells was used as a measure of transcription activity of promoter regions integrated into the plasmid. The maximal activation, 87 times exceeding the background expression of the reporter gene, was registered in the case of the construct **F1-R1** (Fig. 7,B), containing all the intergenic promoters of the PI (Fig. 7,A).

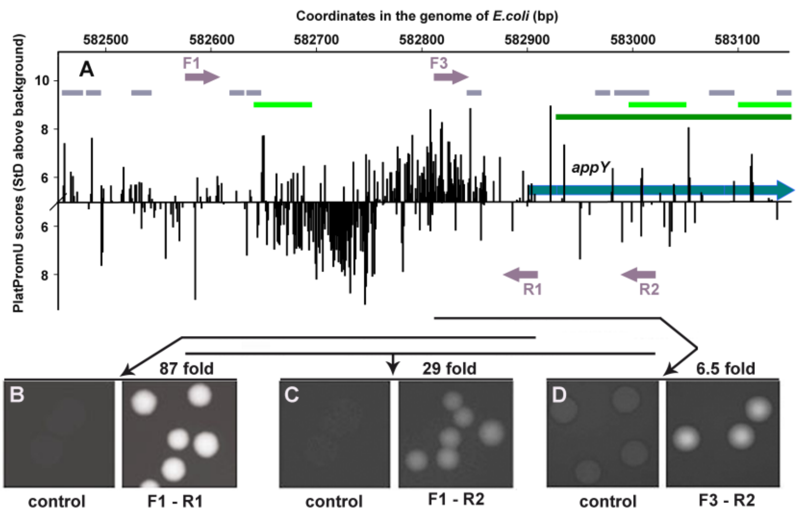

Fig. 7. A: The *appY* gene and the direction of its transcription are marked with blue arrow. Vertical bars above and below the X axis denote transcription initiation points, predicted by PlatPromU for the top or bottom strand, respectively. H-NS binding sites, predicted *in silico* [30], are indicated by thick gray lines, while the H-NS binding sites registered experimentally – by green [14, 15] and dark green [13] lines. Positions of primers used for amplification of the required DNA fragments, integrated into the plasmid pET28b-eGFP, are designated by gray arrows. The relative lengths of the fragments, amplified with primer pairs F1 and R1, F1 and R2 or F3 and R2 are marked by horizontal lines. **B-D:** Colonies of cells transformed with the plasmid carrying promoterless gene *gfp* (control) or by plasmids, which have promoters of the *island* in front of the reporter gene.

Certain contribution to this activation may be given by the high negative supercoiling of the plasmid DNA. In addition, the construct **F1-R1** does not contain the binding sites for the inhibitor H-NS (shown in Fig. 7,A), while the longer insert (**F1-R2**) into the promoter region of *gfp* increased its transcription only 29 fold (Fig. 7,C). That means that H-NS contributes to the suppression of the *appY* transcription in the genome. However, it is important to mention, that the removal of active intergenic promoters from the integrated fragment (construction **F3-R2**) reduced the promoter activity, but maintained it at a rather high level (F_e/F_c = 6.5, Fig. 7,D). Thus, the presence of H-NS binding sites in the regulatory region of *gfp* together with H-NS itself and all the other possible inhibitors in the cytoplasm of the bacterial cells does not preclude the synthesis of *gfp*-mRNA (Fig 7,D). This indicates the importance of the structural factor in the genomic environment of the *island*, which is clearly different in the chromosomal DNA compared to the plasmid at least in terms of supercoiling. Unfortunately, the impact of this unlocalized isomerization on the structural conformation of the promoter insert in the plasmid is not yet possible to model.

Nevertheless, there was a possibility to assess the effect of the plasmid-specific neighborhood on another structural parameter: $(\mathbf{RL\text{-}SL})_{105}$. The examined fragment of the *island* is located between primers **F1** and **R2.** It has two well formed anisotropic bends of the double helix, as indicated by the maxima in the profile of the parameter $(\mathbf{RL\text{-}SL})_{105}$ (the upper plot of Fig. 8,A and animation F1-R2g). Moreover, this fragment lies adjacent to another major bending, which is more distant from the gene *appY*. Depending on the location relative to the transcription start point, anisotropic bends can activate or, conversely, inhibit RNA synthesis. As expected, the formation of the bends takes place in areas with small variations of **φ** (Fig. 8,A, bottom graph), which negatively correlates with the parameter $(\mathbf{RL\text{-}SL})_{105}$.

Fragment **F1-R1** within the plasmid pET28b-eGFP appeared in the flat environment (Fig. 8,B and animation F1-R1) and showed maximal promoter activity, assuming the disposition of the two major bends being acceptable for successful transcription initiation. Incorporation of the longer fragment **F1-R2** into the plasmid led to the formation of the third bend (Fig. 8,C and animation F1-R2p). Since the DNA fragment containing only this moderate bend (Fig. 8,D and animation F2-R2) appeared to be transcriptionally active, it is quite clear, that RNA polymerase interacts with this promoter region. Perhaps it is this interaction, which inhibits RNA synthesis in the construct **F1-R2**, beginning with extremely high efficiency at intergenic promoters located between the primers **F1** and **R1**.

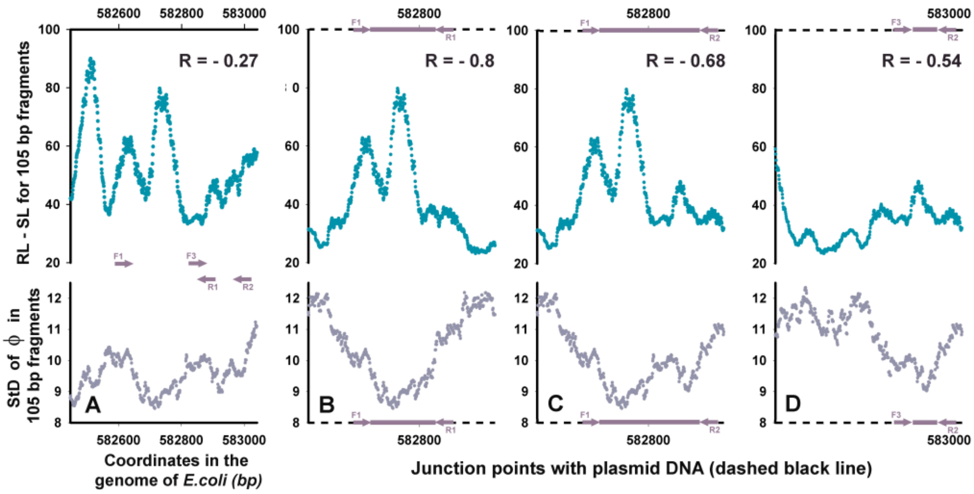

Fig. 8. A: Profiles of conformational parameter, $(\mathbf{RL - SL})_{105}$ (upper graph) and variations of **φ** (bottom graph) within the *promoter island* associated with the *appY* gene in the genomic DNA. Structural parameters were calculated successively for all sub-fragments of 105 bp, with a shift of 1 bp, and the obtained values are indicated at their first positions. The location of primers used for amplification is shown by gray asterisks. **B–D:** The same for different inserts within the plasmid pET28b-eGFP. Inserts in the plasmid DNA are schematically shown on the X-axes by gray cassettes and dashed lines, respectively. Pearson correlation coefficients between parameter $(\mathbf{RL - SL})_{105}$ and StDs are indicated by **R**.

Thus, the structural information obtained in this study allowed suggesting several mechanisms to explain the suppressed transcriptional activity of the studied *promoter island* in the genomic DNA, its activity in the plasmid, as well as the dependence of the reporter gene expression from the peculiarities of different promoter inserts.

DISCUSSION

It has been shown previously that extended (4000 ÷ 48000 bp) "*genomic islands"*, consisting of alien DNA and often containing pathogenic genes, are characterized by high frequency of transcription initiation signals [22]. Analyzing the genomic distribution of the promoters, predicted by the promoter finder PlatProm within the genome of *E. coli* K12 MG1655, we also found 78 unusual regions with an extremely high density of potential transcription start sites but very poor ability to initiate the synthesis of normal mRNAs [1, 9]. Most of them appeared to be located nearby the genes acquired by horizontal transfer, including those which lie outside the long "*genomic islands*". Thus, it was suggested [9], that *promoter islands* are products of accelerated evolution, which is required to assimilate foreign genomic material. Having specific structural properties and enriched by H-NS binding sites, the *islands* may indeed perform contradictory functions: suppress the expression of unwanted "newcomers" and suggest suitable promoters for alien genes, if their expression becomes advantageous under certain growth conditions. This hypothesis requires strong verification. In particular, the first question which came up was: why the number of potential foreign genes predicted by different authors in the genome of *E. coli* K12 MG1655 is much higher (> 1000), than the number of previously discovered *promoter islands* (78)?

This study was undertaken with a hope to fill this gap, if *promoter islands* with mixed σ-specificity will be taken into account. Using the unified promoter finder PlatPromU, which ignores weight matrices accounting σ^{70}-specific modules, we found 434 *mixed promoter islands*. Their functional (Fig. 2 and 3) and structural (Fig. 4–6) properties, as well as specific localization in the bacterial genome (Table 1) appeared to be almost the same as in the case of 78 *islands* revealed previously. However, only ~39% of potentially foreign genes appeared to be associated with the new set of promoter-dense regions (Fig. 1). This percentage is slightly understated, since the selection criteria used to reveal new *islands* were stronger than before. Otherwise we would get additional 166 MPIs, but still only 50% of foreign genes would be associated with the whole set of *promoter islands*. It is clear, that reducing the thresholds for the criteria used, we can further increase the percentage of the assessed overlap. However; approximately 7% of genes with foreign origin, found in all publications [18–20, 22, 23], have no predicted transcription start points at all. It became therefore obvious that some genes acquired by horizontal transfer did not accumulate the transcription signals. Currently it is not clear, whether they are pseudogenes or their expression mode appeared to be suitable for *E. coli* from the very beginning. In any case, the mechanisms, underlying adaptation of foreign genetic material into the regulatory networks of the host cells, remain unknown, and *promoter islands*, as elements specifically associated with alien DNA, provide a basis for the targeted research of adaptive changes in the regulatory regions of transferred genes.

The structural data obtained in this study confirmed all previous observations [9] including rather unexpected low variability of the torsion angle **φ** within the molecular models of *promoter islands*. It was suggested that this low variability reflects some regularity in the spatial configuration of the DNA double helix. Evidences, obtained from *appY*-associated *promoter island*, confirm that the bend localization (maxima of the **RL-SL** parameter) corresponds to the position of the DNA regions characterized by small StD values (Fig. 8). The correlation coefficients for these two parameters are very large and for the three fragments incorporated into pET28b-eGFP decrease in the order: $|\mathbf{R}_{(F1\text{-}R1)}| > |\mathbf{R}_{(F1\text{-}R2)}| > |\mathbf{R}_{(F3\text{-}R2)}|$, exactly corresponding to the registered transcriptional activity (Fig. 7,B–D). It is not yet clear, how general is this correlation. But the bending and the lowered variability of the torsion angles are characteristic properties not only for *promoter islands*, but also for normal promoters (Fig. 4 and 5). I.e. the ability to drive active transcription is encoded in the structural organization of *islands* and can be implemented if necessary.

Acknowledgments

The study was supported by Russian Foundation for Basic Research (grants No. 12-07-00530 and 13-04-00997), and by the Ministry of Education and Science of Russian Federation (project “Multiple promoters as indicators of overlapping transcription and alternative coding”, № 01201276858).

REFERENCES

1. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons in genome-scale promoter mapping. *Nucl. Acids Res.* 2009. V. 37. P. 4419–4431.
2. Huerta A.M., Collado-Vides J. Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *J. Mol. Biol.* 2003. V. 333. P. 261–278.
3. Tutukina M.N., Shavkunov K.S., Masulis I.S. Ozoline O.N. Intragenic promoter-like sites in the genome of *Escherichia coli*. Discovery and functional implication. *J. Bioinform. Comput. Biol.* 2007. V. 5. P. 549–560.
4. Tutukina M.N., Shavkunov K.S., Masulis I.S. Ozoline O.N. Antisense transcription within the *hns* locus of *Escherichia coli*. *Molecular Biology.* 2010. V. 44. P. 439–447.
5. Mendoza-Vargas A., Olvera L., Olvera M., Grande R., Vega-Alvarado L., Taboada B., Jimenez-Jacinto V., Salgado H., Juarez K., Contreras-Moreira B., Huerta A.M., Collado-Vides J., Morett E. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE*. 2009. V. 4. Article № e7526.
6. Dornenburg J.E., DeVita A.M., Palumbo M.J., Wade J.T. Widespread antisense transcription in *Escherichia coli*. *mBio*. 2010. V. 1. Article № e00024-10.
7. Salgado H., Peralta M., Gama-Castro S., Santos-Zavaleta A., Muniz-Rascado L.J., Garcia-Sotelo J.S., Weiss V., Solano-Lira H., Martinez-Flores I., Medina-Rivera A., Salgado-Osorio G., Alquicira-Hernandez S., Alquicira-Hernandez K., Lopez-Fuentes A., Porron-Sotelo L., Huerta A.M., Bonavides-Martinez C., Balderas-Martinez Y., Pannier L., Olvera M., Labastida A., Jimenez-Jacinto V., Vega-Alvarado L., del Moral-Chavez V., Hernandez-Alvarez A., Morett E., Collado-Vides J. RegulonDB (version 8.0): omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucl. Acids Res.* 2013. V. 41. P. D203–D213. URL: http://regulondb.ccg.unam.mx/ (accessed 03 September 2013).
8. Shavkunov K.S., Tutukina M.N., Masulis I.S., Ozoline O.N. Promoter islands: the novel elements in bacterial genomes. *J. Biomol. Struct. Dynam.* 2011. V. 28. P. 1128–1129.
9. Panyukov V.V., Ozoline O.N. Promoters of *Escherichia coli* versus promoter islands: function and structure comparison. *PLoS ONE*. 2013. V. 8. Article № e62601.
10. Reppas N.B., Wade J.T., Church G.M., Struhl K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Mol. Cell.* 2006. V. 24. P. 747–757.
11. Herring C.D., Raffaelle M., Allen T.E., Kanin E.I., Landick R., Ansari A.Z., Palsson B.O. Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J. Bacteriol.* 2005. V. 178. P. 6166–6174.
12. Panyukov V.V., Nazipova N.N., Ozoline O.N. Software package as a tool to study the spatial conformation of bacterial genome sites. *Mathematical Biology and Bioinformatics.* 2011. V. 6. P. t36–t52.
13. Kahramanoglou C., Seshasayee A.S.N., Prieto A.I., Ibberson D., Schmidt S., Zimmermann J., Benes V., Fraser G.M., Luscombe N.M. Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucl. Acids Res.* 2011. V. 39. P. 2073–2091.

14. Grainger D.C., Hurd D., Goldberg M.D., Busby S.J.W. Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucl. Acids Res.* 2006. V. 34. P. 4642–4652.
15. Grainger D.C., Aiba H., Hurd D., Browning D.F., Busby S.J.W. Transcription factor distribution in *Escherichia coli*: studies with FNR protein. *Nucl. Acids Res.* 2007. V. 35. P. 269–278.
16. Prieto A.I., Kahramanoglou C., Ali R.M., Fraser G.M., Seshasayee A.S.N., Luscombe N.M. Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucl. Acids Res.* 2012. V. 40. P. 3524–3537.
17. Cho B.K., Knight E.M., Barrett C.L., Palsson B.O. Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res.* 2008. V. 18. P. 900–910.
18. Lawrence J.G., Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA.* 1998. V. 95. P. 9413–9417.
19. Nakamura Y., Itoh T., Matsuda H., Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genetics.* 2004. V. 36. P. 760–766.
20. Price M.N., Dehal P.S., Arkin A.P. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.* 2008. V. 9. Article № R4.
21. Kiselev S.S., Ozoline O.N. Structure-specific modules as indicators of promoter DNA in bacterial genomes. *Mathematical Biology and Bioinformatics.* 2011. V. 6. P. t1–t13.
22. Huang Q., Cheng X., Cheung M.K., Kiselev S.S., Ozoline O.N., Kwan H.S. High-density transcriptional initiation signals underline genomic islands in bacteria. *PLoS ONE.* 2012. V. 7. Article № e33759.
23. Langille M.G, Brinkman F.S. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics.* 2009. V. 25. P. 664–665.
24. Vlahovicek K., Kajan L., Pongor S. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucl. Acids Res.* 2003. V. 31. P. 3686–3687. URL: http://hydra.icgeb.trieste.it/dna/model_it.html (accessed 03 September 2013).
25. Friedel M., Nikolajewa S., Suehnel J., Wilhelm T. DiProDB: a database for dinucleotide properties. *Nucl. Acids Res.* 2009. V. 37. D37–D40. URL: http://diprodb.fli-leibniz.de/ShowTable.php (accessed 03 September 2013).
26. Sponer J., Gabb H.A., Leszczynski J., Hobza P. Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study. *Biophys. J.* 1997. V. 73. P. 76–87.
27. Oshima T., Ishikawa S., Kurokawa K., Aiba H., Ogasawara N. *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res.* 2006. V. 13. P. 141–153.
28. Lucchini S., Rowley G., Goldberg M.D., Hurd D., Harrison M., Hinton J.C.D. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* 2006. V. 2. Article № e81.
29. Dorman C.J. H-NS, the genome sentinel. *Nat. Rev. Microbiol.* 2007. V. 5. P. 157–161.
30. Munch R., Hiller K., Grote A., Scheer M., Klein J., Schobert M., Jahn D. Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics.* 2005. V. 21. P. 4187–4189.

Received July 23, 2013.
Published October 22, 2013.