

# The Jackprot Simulation © 2011

Conceptualized by **Guillermo Paz-y-Miño C.** [gpazymino@umassd.edu](mailto:gpazymino@umassd.edu) (University of Massachusetts Dartmouth), **Avelina Espinosa** (Roger Williams University) and **Chunyan Y. Bai** (Roger Williams University).

The Jackprot Simulation can be accessed online at: <http://faculty.rwu.edu/cbai/JackprotSimulation.htm> which is a self explanatory website.

**What is The Jackprot Simulation?** It is a didactic slot machine model that illustrates how mutation rate coupled with natural selection can interact to generate highly specialized proteins. The simulation also helps us understand why evolution by means of natural selection cannot be a random process.

**How does The Jackprot Simulation work?** When accessing The Jackprot Simulation, the following window opens:

The Jackprot Simulation

# THE JACKPROT

A didactic slot machine simulation that illustrates how mutation rate coupled with natural selection can interact to generate highly specialized proteins.

The slot-machine represents the cellular chemical apparatus required to generate each of the three nucleotides coding for an amino acid.

Run Your Own Sequence

> Enter your own cDNA sequence in the box below (text format only).  
Example: atgtcaacacaacaaactatgactgtgat...  
From any GenBank or nucleotide accession number click on the CDS link (contains the coding sequence, DNA without introns; from start to stop codon)  
Sequences can be obtained from NCBI (<http://www.ncbi.nlm.nih.gov/nucleotide/>)  
Press Submit to run your own sequence.

Run Example Sequence of KcsA, an Ion Channel Protein from Bacteria

KcsA - Run

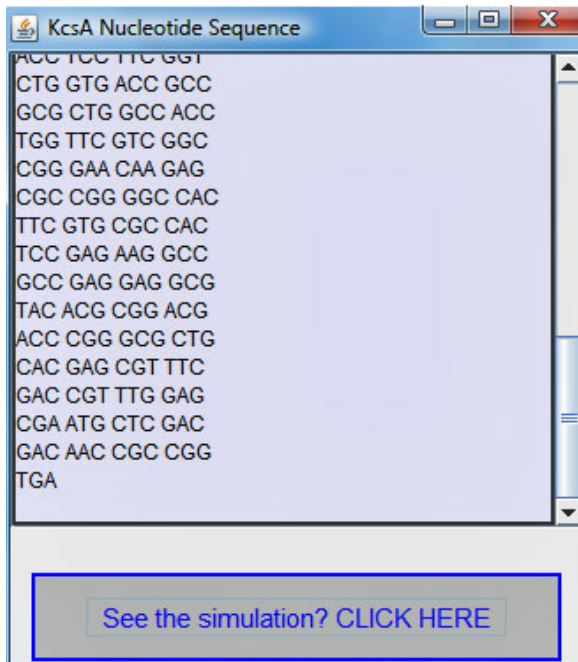
Submit

To run an example protein sequence, in this case KcsA, a bacterial two-transmembrane-domain (2TM) potassium channel, from *Streptomyces lividans* (GenBank Z37969), scroll down the page and click on the box "KcsA - Run," located in the lower left corner of the page, see below:

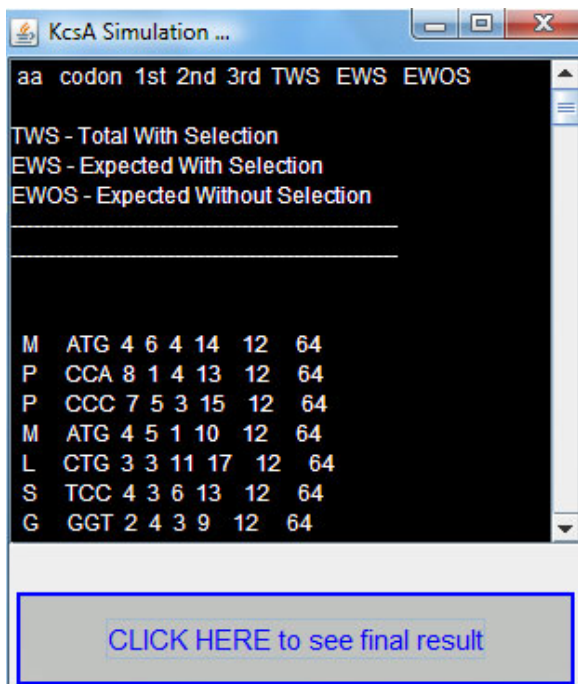
Run Example Sequence of KcsA, an Ion Channel Protein from Bacteria

KcsA - Run

A first window ("KcsA Nucleotide Sequence") will open containing a previously uploaded cDNA sequence (i.e. nucleotides from START to STOP codons without introns) of KcsA, as follows:



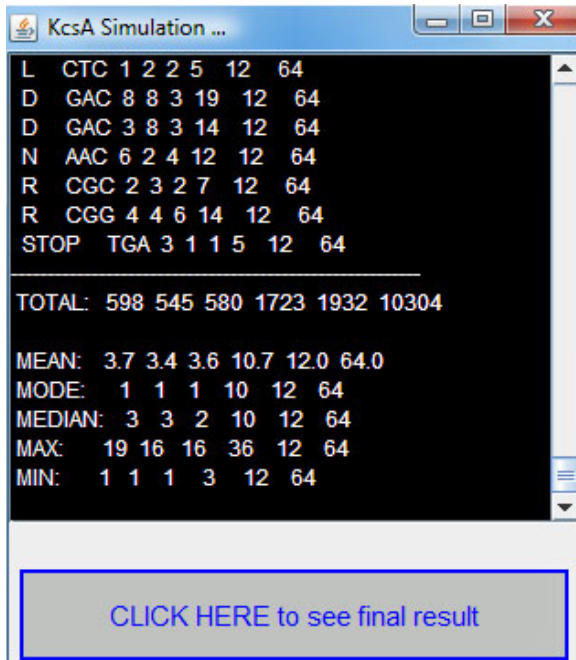
Click on the box "**See the simulation? CLICK HERE**" and a second window ("KcsA Simulation...") for the statistics will open. The new window (see below) shows a list of 161 amino acids aa which include the start codon, methionine M, and its nucleotides A, T, G. The first three numbers after "ATG" (i.e. 4, 6, 4) correspond to the number of "attempts" it took the program to generate each of the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> nucleotide positions (the analogy would be to draw one of four marked marbles from a hat –A, G, C or T– until the matching-to-the-sequence-nucleotide is generated or "ATG"): 4 attempts to generate A, 6 attempts to generate T, and 4 attempts to generate G.



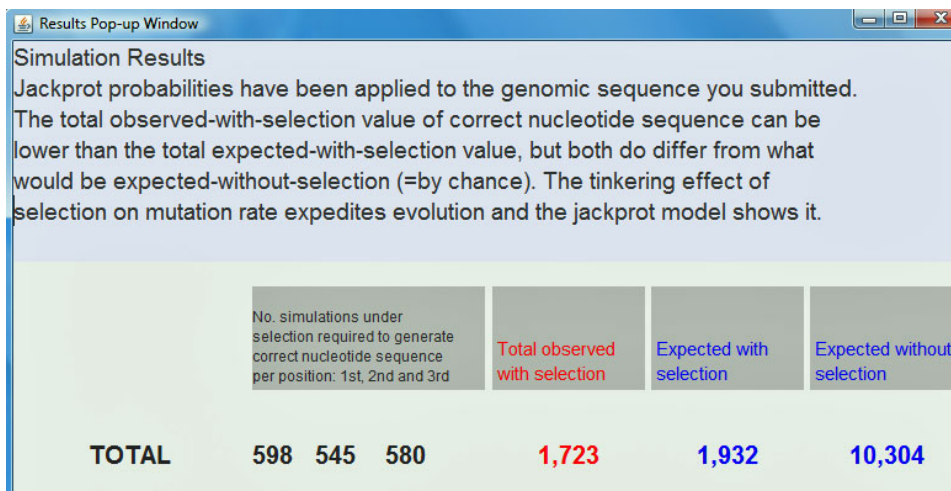
The next numbers in this "ATG row," which are 14, 12 and 64, correspond to: 14 or the total number of attempts with selection, meaning "natural selection at work" (TWS), observed to generate the codon; 12 or the statistically expected number of attempts to generate the codon with selection (EWS); and 64 or the statistically expected number of attempts to generate the codon without selection (EWOS).

Note that for these three columns (i.e. TWS, EWS and EWOS), only the values under the TWS column change since they correspond to the observed values generated under the "editing role" of natural selection, however, the expected values of EWS and the EWOS remain constant, which should be the case for any protein sequence.

By scrolling down this window ("KcsA Simulation..."), the TOTAL values corresponding to the descriptive statistics become visible (see below): the TOTAL values for the 1<sup>st</sup> position or 598, for the 2<sup>nd</sup> position or 545, and for the 3<sup>rd</sup> position or 580, as well as the values corresponding to total observed with selection (TWS = 1,723), expected with selection (EWS = 1,932), and expected without selection (EWOS = 10,304). The MEAN, MODE, MEDIAN, MAX (maximum) and MIN (minimum) values are also generated by the simulation:



Finally, click on the box "CLICK HERE to see the final result" and a third window ("Results Pop-up Window") with the summary statistics will open (see below): Total with selection (1,723), Expected with selection (1,932), and Expected without selection (10,304):



Note that this example of KcsA has been programmed to run indefinitely and produce the same results, thus guiding the instructor and/or student on how to use the simulation.

**How can I run my own sequence?** To enter your own cDNA sequence (i.e. nucleotides from START to STOP codons without introns) scroll down the main page and locate the **sequence input box** (lower right corner):

**Run Your Own Sequence**

> Enter your own cDNA sequence in the box below (text format only).  
 Example: atgtcaacacaacaactgactgtagat...  
 From any GenBank or nucleotide accession number click on the CDS link  
 (contains the coding sequence, DNA without introns; from start to stop codon)

Sequences can be obtained from NCBI (<http://www.ncbi.nlm.nih.gov/nucleotide/>)

Press Submit to run your own sequence.

Cut and paste any cDNA or nucleotide sequence obtained from NCBI (National Center for Biotechnology Information (NCBI <http://www.ncbi.nlm.nih.gov/nucleotide/>) or alternative sources and click **"Submit."** The Jackprot will generate the statistics just like in the KcsA example above. Note that for KcsA, the cDNA sequence was obtained from NCBI by following three steps: (1) identifying its accession number within NCBI, which is unique to every entry in the data base; this number was GenBank Z37969 and corresponds to **"*S.lividans skc1 gene for potassium channel protein.*"** (2) copying and pasting this number on the search box of the "Nucleotide" NCBI website, and (3) clicking on the "gene" or "CDS" links, which correspond to the coding region from START to STOP codons, without introns. The gene sequence is shown below and it can be cut and pasted entirely within the **sequence input box** of The Jackprot Simulation, including its row numbers (i.e. 1, 61, 121, 181...); the program shall run it:

```

1 atgccacca tgctgtccg tcttctggcc agattgtca aactgctgt cgggcgccac
61 ggcagtgcg tgcactggag ggcgcgggt gccgcgacgg tctcctggt gatcgtcctc
121 ctgcgggct cgtacttggc cgtcctggt gagcgcggcg caccgggccc gcagctgac
181 acgtatccc gggcgctgtg gtggtccgtg gagaccgcca cgaccgtcgg ctacggcgac
241 ctgtacccc tgactctgtg gggccggctc gtggccgtgg tggatgatgt cgcggggatc
301 acctcctcg gtctggtgac cgcgcgcgtg gccacctggt tcgtcggccg ggaacaagag
361 cgcgggggccc acttctgtcg ccaactccgag aaggccgccg aggaggccta cacgcgggacg
421 acccgggccc tgcacgagcg ttfcgaccgt ttggagcga tgcctgacga caaccgccg
481 tga
  
```

Press Submit to run your own sequence.

```

1 atgccacca tgctgtccg tcttctggcc agattgtca aactgctgt cgggcgccac
61 ggcagtgcg tgcactggag ggcgcgggt gccgcgacgg tctcctggt gatcgtcctc
121 ctgcgggct cgtacttggc cgtcctggt gagcgcggcg caccgggccc gcagctgac
181 acgtatccc gggcgctgtg gtggtccgtg gagaccgcca cgaccgtcgg ctacggcgac
241 ctgtacccc tgactctgtg gggccggctc gtggccgtgg tggatgatgt cgcggggatc
301 acctcctcg gtctggtgac cgcgcgcgtg gccacctggt tcgtcggccg ggaacaagag
361 cgcgggggccc acttctgtcg ccaactccgag aaggccgccg aggaggccta cacgcgggacg
421 acccgggccc tgcacgagcg ttfcgaccgt ttggagcga tgcctgacga caaccgccg
481 tga
  
```

**Further Applications of The Jackprot Simulation.** Instructors and/or students could run the same simulation numerous times, e.g. a laboratory of 20 students could run the same sequence ten times per student and generate 200 different values for the *Total observed with selection* column. These numbers could be used for additional learning experiences, for example, for conducting and plotting an analysis of the mean distribution of values in Excel, which can help users understand how even under the strict and powerful effect of selection there shall be variability in the data. Below, we provide such analysis for 100 runs of the genomic sequence of KcsA (Tables S1, S2 and Figure S1). Alternatively, instructors and/or students could run nucleotide sequences of diverse proteins (small, medium or large nucleotide number) to compare and contrast the rates of *Total with selection* versus *Expected with selection* versus *Expected without selection* values among the different sequences. The Jackprot Simulation does allow for some creative work.

**Table S1.** Jackprot probabilities applied to 100 runs of the genomic sequence coding for 160 amino acids and a stop codon of KcsA. The total observed-under selection value of correct nucleotide sequence composition ranges from 1752 to 2139, with a mean value (1942.97) comparable to the expected under selection value (1932), and it does differ from what would be expected by chance without selection (10304). The tinkering effect of selection on mutation rate expedites evolution and The Jackprot Simulation shows it.

Simulation number	No. simulations under selection required to generate correct nucleotide sequence per position			Total observed with selection	Expected with selection	Expected without selection
	1st	2nd	3rd			
	1	598	679			
2	660	700	646	2006	1932	10304
3	649	600	678	1927	1932	10304
4	643	690	686	2019	1932	10304
5	646	602	556	1804	1932	10304
6	640	644	743	2027	1932	10304
7	607	571	676	1854	1932	10304
8	642	661	750	2053	1932	10304
9	697	622	659	1978	1932	10304
10	688	634	572	1894	1932	10304
11	621	676	688	1985	1932	10304
12	612	631	556	1799	1932	10304
13	700	652	654	2006	1932	10304
14	716	642	606	1964	1932	10304
15	692	594	635	1921	1932	10304
16	701	574	586	1861	1932	10304
17	550	583	662	1795	1932	10304
18	679	648	662	1989	1932	10304
19	711	605	629	1945	1932	10304
20	654	717	703	2074	1932	10304
21	653	707	637	1997	1932	10304
22	554	578	620	1752	1932	10304
23	601	592	574	1767	1932	10304
24	658	685	701	2044	1932	10304
25	678	724	657	2059	1932	10304
26	615	644	681	1940	1932	10304
27	626	627	730	1983	1932	10304
28	630	600	728	1958	1932	10304
29	632	602	639	1873	1932	10304
30	693	647	639	1979	1932	10304
31	639	617	657	1913	1932	10304
32	603	605	635	1843	1932	10304
33	652	702	683	2037	1932	10304
34	626	674	674	1974	1932	10304
35	693	587	623	1903	1932	10304
36	599	658	729	1986	1932	10304
37	732	669	635	2036	1932	10304
38	667	578	693	1938	1932	10304
39	596	623	567	1786	1932	10304

40	686	719	678	2083	1932	10304
41	653	650	647	1950	1932	10304
42	663	632	650	1945	1932	10304
43	623	617	588	1828	1932	10304
44	646	582	690	1918	1932	10304
45	663	622	646	1931	1932	10304
46	705	625	734	2064	1932	10304
47	647	582	639	1868	1932	10304
48	604	688	672	1964	1932	10304
49	636	612	660	1908	1932	10304
50	541	614	716	1871	1932	10304
51	664	613	596	1873	1932	10304
52	740	656	743	2139	1932	10304
53	613	644	693	1950	1932	10304
54	661	567	689	1917	1932	10304
55	677	695	736	2108	1932	10304
56	631	625	609	1865	1932	10304
57	723	656	575	1954	1932	10304
58	653	650	679	1982	1932	10304
59	723	617	593	1933	1932	10304
60	627	633	647	1907	1932	10304
61	612	658	599	1869	1932	10304
62	576	647	601	1824	1932	10304
63	583	604	593	1780	1932	10304
64	682	601	643	1926	1932	10304
65	603	638	684	1925	1932	10304
66	610	641	649	1900	1932	10304
67	625	601	651	1877	1932	10304
68	639	604	595	1838	1932	10304
69	634	658	675	1967	1932	10304
70	706	685	711	2102	1932	10304
71	673	658	675	2006	1932	10304
72	722	629	741	2092	1932	10304
73	633	657	624	1914	1932	10304
74	695	607	651	1953	1932	10304
75	693	676	636	2005	1932	10304
76	564	669	668	1901	1932	10304
77	660	693	564	1917	1932	10304
78	678	690	636	2004	1932	10304
79	648	703	699	2050	1932	10304
80	606	622	698	1926	1932	10304
81	591	687	721	1999	1932	10304
82	633	692	603	1928	1932	10304
83	660	683	633	1976	1932	10304
84	671	634	623	1928	1932	10304
85	663	599	623	1885	1932	10304
86	616	643	632	1891	1932	10304
87	740	618	707	2065	1932	10304
88	702	634	712	2048	1932	10304

89	580	579	651	1810	1932	10304
90	680	589	668	1937	1932	10304
91	659	654	692	2005	1932	10304
92	658	601	643	1902	1932	10304
93	622	684	624	1930	1932	10304
94	586	673	687	1946	1932	10304
95	629	675	645	1949	1932	10304
96	647	613	676	1936	1932	10304
97	715	648	650	2013	1932	10304
98	604	680	634	1918	1932	10304
99	581	641	635	1857	1932	10304
100	637	733	643	2013	1932	10304
Min	541	567	556	1752	1932	10304
Max	740	733	750	2139	1932	10304
Mean	647.48	640.44	655.05	1942.97	1932	10304
Median	647	641	651	1939	1932	10304
Mode	660	658	635	2006	1932	10304
St Deviation	43.95486206	39.60622337	45.9622979	81.21420175	0	0
Variance	1932.029899	1568.652929	2112.532828	6595.746566	0	0

**The Jackprot Simulation: Paz-y-Miño C., Espinosa, Bai © 2011**  
<http://faculty.rwu.edu/cbai/JackprotSimulation.htm>  
<http://pazymino1evolutionliteracy.blogs.umassd.edu/2011/01/02/the-jackprot-simulation/>

**Table S2.** Analysis of the mean of the total observed-under selection values (from Table S1) corresponding to 100 runs of The Jackprot Simulation where the genomic sequence coding for 160 amino acids and a stop codon of KcsA were uploaded. Data rearranged from minimum to maximum values; number of simulations within each Z score are highlighted in color.

Simulation number	Total observed with selection	Number of simulations within each Z score	Z score
22	1752	3	Z -3
23	1767		
63	1780		
39	1786	12	Z -2
17	1795		
12	1799		
5	1804		
89	1810		
62	1824		
43	1828		
68	1838		
32	1843		
7	1854		
99	1857		
16	1861		
56	1865	36	Z -1
47	1868		
61	1869		
50	1871		
29	1873		
51	1873		
67	1877		
85	1885		
86	1891		
10	1894		
66	1900		
76	1901		
92	1902		
35	1903		
60	1907		
49	1908		
31	1913		
73	1914		
54	1917		
77	1917		
44	1918		
98	1918		
15	1921		
65	1925		
64	1926		
80	1926		



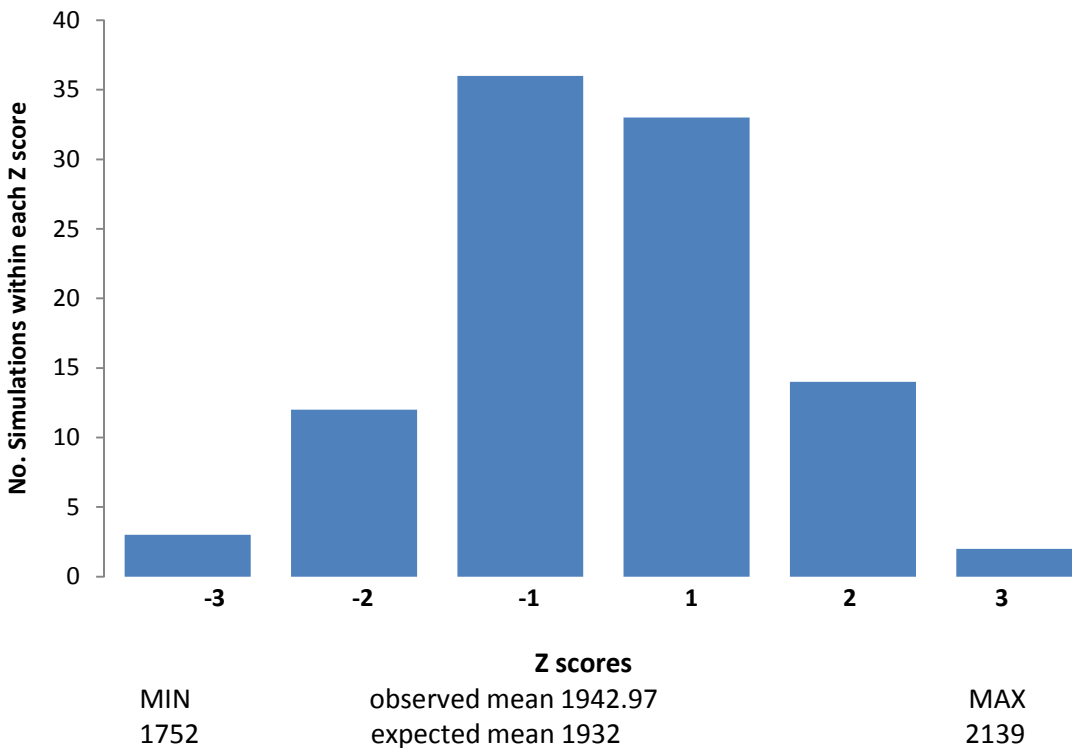
3	1927		
82	1928		
84	1928		
93	1930		
45	1931		
59	1933		
96	1936		
90	1937		
38	1938		
26	1940	mean	
19	1945		
42	1945		
94	1946		
95	1949		
41	1950		
53	1950		
74	1953		
57	1954		
1	1958		
28	1958		
14	1964		
48	1964		
69	1967		
34	1974		
83	1976		
9	1978		
30	1979	33	Z +1
58	1982		
27	1983		
11	1985		
36	1986		
18	1989		
21	1997		
81	1999		
78	2004		
75	2005		
91	2005		
2	2006		
13	2006		
71	2006		
97	2013		
100	2013		
4	2019		
6	2027		
37	2036		
33	2037		
24	2044		
88	2048		
79	2050		
8	2053	14	Z +2

25	2059		
46	2064		
87	2065		
20	2074		
40	2083		
72	2092		
70	2102		
55	2108	2	Z +3
52	2139		
Min	1752		
Max	2139		
Mean	1942.97		
St Deviation	81.21420175		

**The Jackprot Simulation: Paz-y-Miño C., Espinosa, Bai © 2011**  
<http://faculty.rwu.edu/cbai/JackprotSimulation.htm>  
<http://pazymino1evolutionliteracy.blogs.umassd.edu/2011/01/02/the-jackprot-simulation/>

Z score	-3	-2	-1	1	2	3
No. of simulations within each Z score	3	12	36	33	14	2

**Figure S1.** Distribution of the total observed-under selection values corresponding to 100 runs of The Jackprot Simulation, a computer program where the genomic sequence coding for 160 amino acids and a stop codon of KcsA were uploaded.



**The Jackprot Simulation: Paz-y-Miño C., Espinosa, Bai © 2011**  
<http://faculty.rwu.edu/cbai/JackprotSimulation.htm>  
<http://pazymino1evolutionliteracy.blogs.umassd.edu/2011/01/02/the-jackprot-simulation/>