

2011

The Jackprot Simulation Couples Mutation Rate with Natural Selection to Illustrate How Protein Evolution Is Not Random

Guillermo Paz-y-Miño C
University of Massachusetts Dartmouth

Avelina Espinosa
Roger Williams University, aespinosa@rwu.edu

Chunyan Y. Bai
Roger Williams University

Follow this and additional works at: http://docs.rwu.edu/fcas_fp

 Part of the [Biology Commons](#)

Recommended Citation

Paz-y-Miño-C, Guillermo and Avelina Espinosa, and Chunyan Y. Bai. 2011. "The Jackprot Simulation Couples Mutation Rate with Natural Selection to Illustrate How Protein Evolution Is Not Random." *Evolution: Education and Outreach* 4, (3): 502-514.

This Article is brought to you for free and open access by the Feinstein College of Arts and Sciences at DOCS@RWU. It has been accepted for inclusion in Feinstein College of Arts & Sciences Faculty Publications by an authorized administrator of DOCS@RWU. For more information, please contact mwu@rwu.edu.

The Jackprot Simulation Couples Mutation Rate with Natural Selection to Illustrate How Protein Evolution Is Not Random

Guillermo Paz-y-Miño C. · Avelina Espinosa ·
Chunyan Y. Bai

Published online: 24 March 2011
© Springer Science+Business Media, LLC 2011

Abstract Protein evolution is not a random process. Views which attribute randomness to molecular change, deleterious nature to single-gene mutations, insufficient geological time, or population size for molecular improvements to occur, or invoke “design creationism” to account for complexity in molecular structures and biological processes, are unfounded. Scientific evidence suggests that natural selection tinkers with molecular improvements by retaining adaptive peptide sequence. We used slot-machine probabilities and ion channels to show biological directionality on molecular change. Because ion channels reside in the lipid bilayer of cell membranes, their residue location must be in balance with the membrane’s hydrophobic/philic nature; a selective “pore” for ion passage is located within the hydrophobic region. We contrasted the random generation of DNA sequence for KcsA, a bacterial two-transmembrane-domain (2TM) potassium channel, from *Streptomyces lividans*, with an under-selection scenario, the “jackprot,” which predicted much faster evolution than by chance. We wrote a computer program in

Electronic supplementary material The online version of this article (doi:10.1007/s12052-011-0329-2) contains supplementary material, which is available to authorized users.

G. Paz-y-Miño C. (✉)
Department of Biology, University of Massachusetts Dartmouth,
285 Old Westport Road,
North Dartmouth, MA 02747–2300, USA
e-mail: gpazymino@umassd.edu

A. Espinosa
Department of Biology, Roger Williams University,
One Old Ferry Road,
Bristol, RI 02809, USA

C. Y. Bai
Department of Computer Science, Roger Williams University,
One Old Ferry Road,
Bristol, RI 02809, USA

JAVA APPLET version 1.0 and designed an online interface, *The Jackprot Simulation* <http://faculty.rwu.edu/cbai/JackprotSimulation.htm>, to model a numerical interaction between mutation rate and natural selection during a scenario of polypeptide evolution. Winning the “jackprot,” or highest-fitness complete-peptide sequence, required cumulative smaller “wins” (rewarded by selection) at the first, second, and third positions in each of the 161 KcsA codons (“jackdons” that led to “jackacids” that led to the “jackprot”). The “jackprot” is a didactic tool to demonstrate how mutation rate coupled with natural selection suffices to explain the evolution of specialized proteins, such as the complex six-transmembrane (6TM) domain potassium, sodium, or calcium channels. Ancestral DNA sequences coding for 2TM-like proteins underwent nucleotide “edition” and gene duplications to generate the 6TMs. Ion channels are essential to the physiology of neurons, ganglia, and brains, and were crucial to the evolutionary advent of consciousness. The Jackprot Simulation illustrates in a computer model that evolution is not and cannot be a random process as conceived by design creationists.

Keywords Design creationism · Ion channels · Jackacid · Jackdon · Molecular evolution

Introduction

The evolution of new protein motifs with specific amino acid sequences, via classical Darwinian trajectories, has been challenged (Behe 1998, 2001, 2002, 2007, 2009; Behe and Snoke 2004, 2005) by attributing randomness to molecular change, deleterious nature to intermediate mutations (rather than neutrality or selective advantage),

insufficient geological time or population size for molecular improvements to occur, and invoking “design” (= supernatural causation) for the materialization of complex molecular structures (Nelson 1996; Luskin and Gage 2008). This logic has been dismissed by researchers (Schneider 2000; Pennock 2001; Long et al. 2003; Young and Edis 2004; Lynch 2005; Forrest and Gross 2007; Petto and Godfrey 2007; Durrett and Schmidt 2008, 2009; Schneiderman and Allmon 2009; Paz-y-Miño C. and Espinosa 2010a) and journal editors (Hermodson 2005) based on fundamental evolutionary premises: (1) large variation in mutation rate between and within lineages, and/or protein sites, is susceptible to positive selection; (2) protein-site mutagenesis is associated with mutation and acceptance rates at multiple sites in a genome (= compensatory changes); (3) new protein functions after domain junction can experience faster evolution (e.g., fused genes); and (4) selection acts continuously and cumulatively (= “editing role”) on intermediate protein forms, increasing and maintaining molecular diversity, and expediting molecular evolution. Thus, single emergence of primordial genetic sequences or protein-adaptive change from “design creationism” is highly improbable.

Here, we use slot-machine probabilities—the “jackpot” model—and ion channel evolution to illustrate how mutation rate coupled with natural selection have expedited the diversification of ion channels, from simple two-transmembrane (2TM) proteins to complex, multi-domain (6TM) molecules highly tuned to respond to environmental stimuli and regulate ion passage through the cell membrane. Ion channels are essential to ionic homeostasis of all cells and crucial to the hyperpolarization and depolarization of

neurons (Kress and Mennerick 2009; Miller 2009); ultimately, irritability of individual neurons, communication among neuronal networks, ganglia activity, and brain power depend on ion-channel function.

Ion Channels as Exemplars of Protein Evolution

Ion channels are integral proteins in the plasma membrane of all cells and probably all organisms. A single or limited number of prokaryotic precursors gave origin to the large diversity of modern ion channels (Derst and Karschin 1998; Durell et al. 1999; Anderson and Greenberg 2001; Martinac et al. 2008). Their genetic evolution is very complex and includes numerous gene duplications (orthologous and paralogous in prokaryotes and eukaryotes), vast nucleotide change, and elaborate alternative splicing (Miller 2000; Anderson and Greenberg 2001; Sansom et al. 2002; Pichon et al. 2004; Hill et al. 2008). For didactic purposes, we summarize ion-channel diversification as follows: simplest forms of ion channels probably consisted of two transmembrane ($M1+M2=2TM$) hydrophobic domains with a pore-forming loop (P) in the middle (Fig. 1a); some modern potassium (K^+) channels are tetramers ($4\times 2TM$) of this type (e.g., KcsA K^+ of *Streptomyces lividans*, Fig. 1b and below). Additional transmembrane segments have evolved attached to the basic 2TM motif, generating 6TM proteins (one-subunit-6TM), which ancestral gene sequences have duplicated further into assemblages of two or four 6TM-linked subunits (two-subunits- or four-subunits-6TMs; Fig. 2). Assemblages of two 2TM-linked subunits ($2\times 2TM$; Fig. 2), or 6TM- and 2TM-linked subunits ($6TM+$

Fig. 1 Cell membrane topology of a 2TM potassium (K^+) channel: fundamental building blocks consist of multiples of transmembrane domains ($M1$ and $M2$), a pore-forming loop (P), and a signature (S) sequence of amino acids (TT V/I GYG or “ion selectivity filter”) highly conserved in K^+ channels across taxa; intra- and extracellular segments have specific names (i.e., pre-M1, Turret and extended regions, post-M2); K^+ ions are represented by small spheres. **b** Two subunits (cutaway view) of the tetrameric ($4\times 2TM$) structure of KcsA, one of the simpler prokaryotic K^+ channels from the soil bacterium *Streptomyces lividans*, are depicted

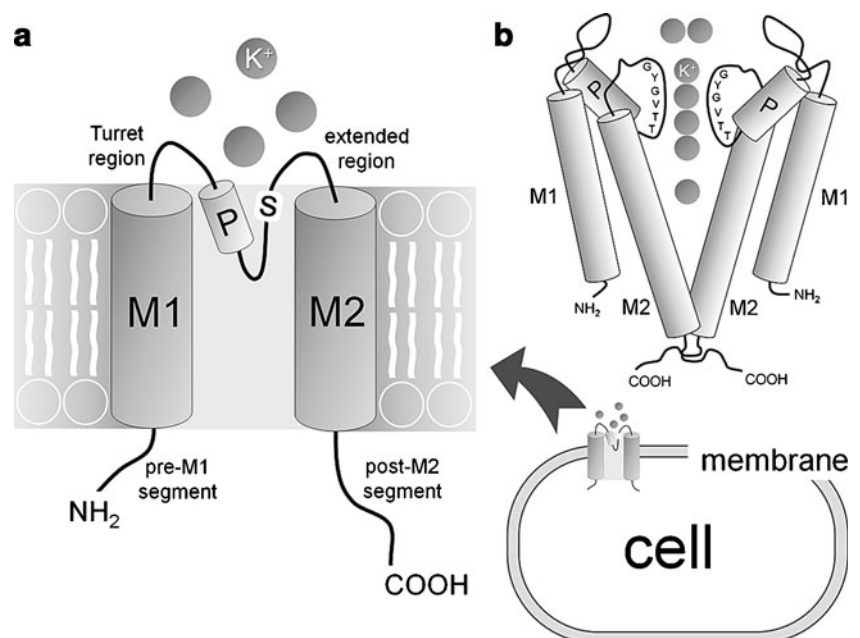
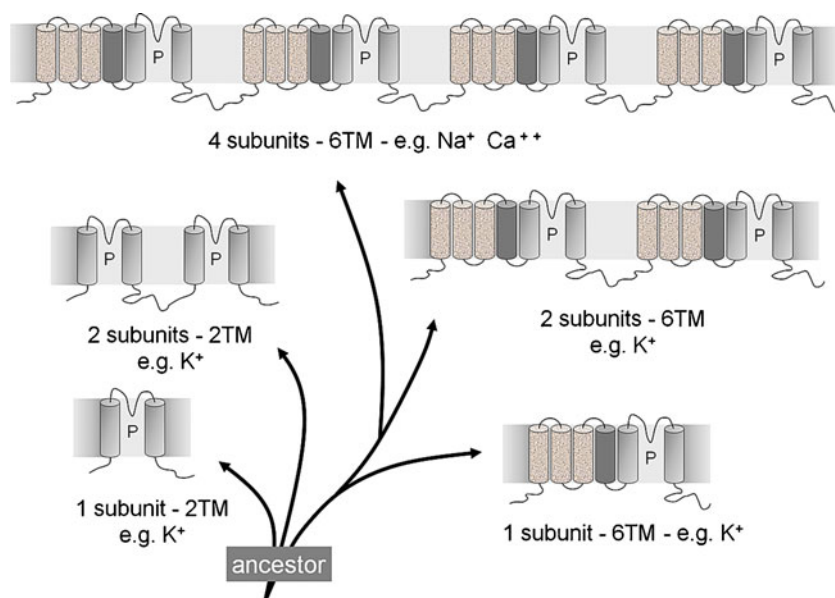


Fig. 2 Simplified pattern of ion channel evolution. Ancestral 2TM-like proteins probably gave origin to the complex families of ion channels known today. A precursor building block of two transmembrane domains, containing the pore-forming loop (*P*), is conserved in channels of many taxa. Gradual and cumulative nucleotide mutations, combined with gene duplications (orthologous and paralogous diversification), have given origin to the multiple-subunit-2TM and multiple-subunit-6TM channels



2TM; not depicted) have been described (Durell et al. 1999; Anderson and Greenberg 2001), suggesting great variability in the pattern of protein assemblage and gene fusion during ion channel evolution.

Cellular metabolism, including osmoregulation, secretory processes, signal transduction, and ion-homeostasis triggered the evolution of ion-transporting proteins in the plasma membrane (Derst and Karschin 1998), a function later “exapted” (= new adaptive role) to electrical excitability and signaling (via hyper/de/polarization) and communication (via networking) among neurons. Cells capable of detecting (sensing, e.g., mechanoreceptors), responding to and controlling the differential concentration of ions inside and outside the plasma membrane, by means of specialized proteins, probably evolved into primordial neurons (Galliot et al. 2009).

We consider ion channels didactic exemplars of protein evolution, in the context of the “jackprot” model (below), for various reasons: (1) ion-channel genetics, genomics, proteomics, cell and tissue localization, electrophysiology, response to neurotoxins or medical-drugs, bioinformatics, structural modeling, X-ray crystallography, and involvement in prevalent diseases or “channelopathies” (e.g., genetically defective K⁺ channels: type II diabetes, cardiac arrhythmia and epilepsy; chloride Cl⁻ channels: cystic fibrosis; Ca⁺⁺ channels: Parkinson’s disease; and concerted activity of K⁺ and Cl⁻ channels: tumor metastasis) have been widely documented (Capener et al. 2002; Kunzelmann 2005; Rogers et al. 2006; Sontheimer 2008); (2) K⁺, Na⁺, and Ca⁺⁺ channels are textbook case studies in neurobiology and electrophysiology of neurons and muscle cells (Kress and Mennerick 2009); their role in action potentials, neuromuscular junctions and cardiac rhythm are familiar to wide audiences; (3) the evolutionary patterns of ion-

channel diversification, from simpler 2TM-like ancestors to more complex multiple-subunit 6TMs (Fig. 2), can be inferred from genomic analyses within (paralogous gene families) and between taxonomic lineages (orthologous gene families); (4) comparative DNA and amino-acid sequence analyses (e.g., *Homo* vs. *Rattus* vs. *Mus* vs. *Drosophila* vs. *Caenorhabditis* vs. *Paramecium* vs. *Escherichia* vs. *Arabidopsis*; Doyle et al. 1998; Shealy et al. 2003) reveal classical Darwinian patterns of ion-channel evolution via cumulative single-nucleotide mutations, gene duplications and fusions, and protein-domain junctions; and (5) neuronal networks, ganglia activity, and brain functions depend on ion-channels for sensitivity (i.e., touch/pressure/vibration, sound, light, chemosignals/odor, and electric fields) and electrical transmission of stimuli, motor (neuromuscular junction for voluntary or reflex movement) or excretory response (neuro-endocrine stimulation), behavior, and consciousness (Galliot et al. 2009; Kress and Mennerick 2009; Miller 2009). Thus, the ubiquitous inclusion of ion channels in significant empirical and practical aspects of the biology and health-related careers’ curricula makes them unique didactic tools for communicating evolutionary principles to all audiences, and promoting evolution literacy (innovation in science education has been prioritized by authors concerned with the misleading role of “design creationism” in public-outreach campaigns; Paz-y-Miño C. and Espinosa 2009a, b; Paz-y-Miño C. and Espinosa 2010a, b, c).

Slot-machine Probabilities and the “Jackprot”

The “jackprot” uses simplified slot-machine probability principles to demonstrate how mutation rate coupled with natural selection suffices to explain the origin and evolution of

highly specialized proteins, such as the single- (K^+) or multiple-subunit 6TM (Na^+ and Ca^{++}) channels. Winning the “jackprot,” or highest-fitness complete-peptide sequence, requires gradual and cumulative smaller “wins” (rewarded by selection) at the first, second and third nucleotide positions in each of the codons coding for a polypeptide (= “jackdons” that lead to “jackacids” that lead to the “jackprot”; Fig. 3). A slot-machine represents the cellular chemical apparatus, product itself of Darwinian evolution, required to generate, step by step, each of the three nucleotides coding for an amino acid. The probability of getting the correct triplet, for example, the start codon methionine or ATG, in a single attempt (or winning the “jackacid”), is equal to one in 64, or one divided by $4 \times 4 \times 4$ (i.e., the total number of possible nucleotides per position multiplied by itself three times). But because molecular evolution occurs gradually, a naturalistic assumption of the

“jackprot” model, each time any of the correct nucleotides is generated by the slot-machine, natural selection rewards it and keeps it (partial nucleotide win in a codon or “jackdon”). Therefore, the probability of arriving, nucleotide by nucleotide, at the ATG sequence is equal to one in 12, or one divided by $4+4+4$ (i.e., the summation of the individual probabilities for each nucleotide position), a much faster evolutionary process. Note that the sequential and additive arrival at the phenotypically meaningful sequence of A plus T plus G, represents, in reality, the accumulation of events fixed by natural selection during protein evolution, which entails clustered changes of multiple parts, and at diverse locations, within functional domains.

The genomic and amino acid sequences of the well-studied 2TM K^+ channel from the soil bacterium *S. lividans* (KcsA K^+ ; Schrempf et al. 1995; Doyle et al. 1998; Lu et al. 2001; Shealy et al. 2003; Williamson et al. 2003; Doyle

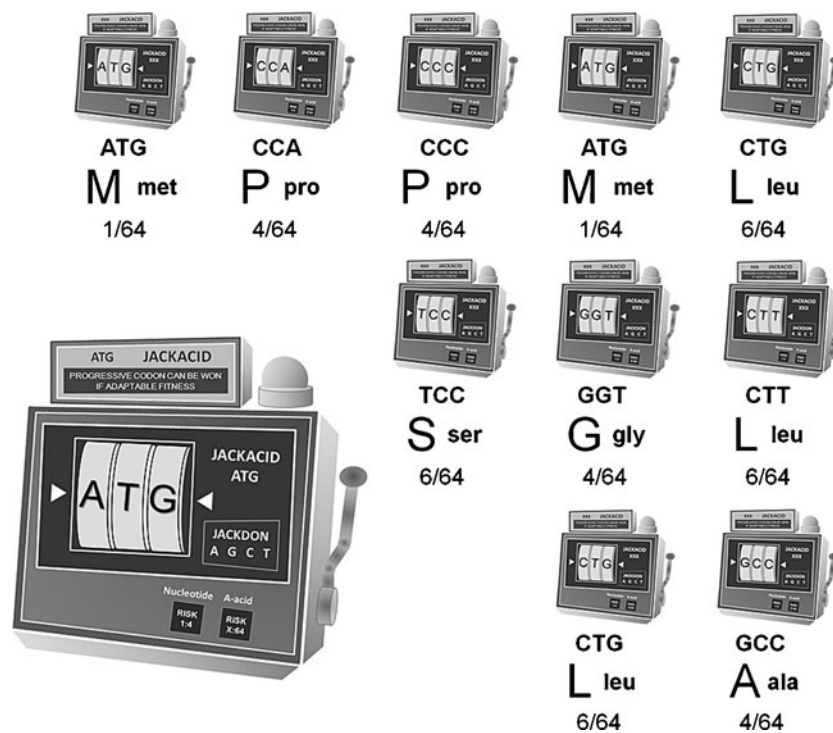


Fig. 3 The “jackprot” model of protein evolution. A slot machine (*lower left*) represents the cellular apparatus required to generate each of the three nucleotides coding for an amino acid, for example, the starting codon methionine ATG. The probability of generating ATG in a single attempt, without the influence of natural selection, is equal to one in 64 ($1/4 \times 4 \times 4$); however, each time a biologically meaningful nucleotide is generated by the slot-machine (mimicking mutation rate), natural selection would keep it as a building block of a codon and as a partial win, or “jackdon.” Thus, the probability of arriving under selection at the ATG sequence would be equal to one in 12 ($1/4+4+4$). Winning the “jackprot,” or highest-fitness complete-peptide sequence, for example, 160 amino acids plus one-stop codon in the sequence of KcsA, a K^+

channel from the bacterium *Streptomyces lividans*, would require gradual and cumulative smaller wins (“jackdons”) at each nucleotide position, which lead to larger rewards when a correct amino acid is generated by the slot machine (“jackacids”), and which subsequently lead to the “jackprot” or the complete 161 codons. The smaller slot machines represent each cell apparatus necessary to generate the first ten amino acids of KcsA (complete sequence available at GenBank Z37969; Swiss-Prot P0A334); the genomic sequence, letter coding/ acronym, and the number of codons coding for that specific amino acid within the genetic code are shown below each machine (e.g., ATG, M met, and one in 64)

2004) helps us exemplify how the “jackprot” works. KcsA is one of the simpler K^+ channels (Fig. 1b): 483 nucleotides code for its 160 amino acids plus a stop codon (GenBank Z37969; Swiss-Prot P0A334). KcsA probably retains many features of earlier 2TM ancestors; the amino acid sequence of the M1 and M2 domains resemble the transmembrane segments immediately connected to the pore region in K^+ channels of prokaryotes, invertebrates, vertebrates, and plants (Doyle et al. 1998; Williamson et al. 2003). The pore signature sequence is nearly identical (TT V/I GYG) to that of bacteria, protists, fruit flies, nematodes, mice, rats, and humans (Doyle et al. 1998; Lu et al. 2001), suggesting a common origin of all these channels.

Although randomness can be a statistical component of mutation rate in ion channel evolution, synergistic biological restrictions (e.g., structural compatibility of purine:pyrimidine pairing in DNA; differential codon representation per amino acid; residue site specificity for plasma membrane hydrophobicity and hydrophilicity; pore- and signature-sequence location in the non-polar region of the plasma membrane; and codon bias intrinsic to taxonomic lineages, below) impose directionality on molecular assemblage, and KcsA exemplifies it. Natural selection has tinkered with molecular improvements in ancestors of KcsA by favoring and retaining adaptive peptide sequence for optimal function.

Why Is Evolution Not a Random Process?

We address this question as follows: (1) the probability of arriving randomly at the correct arrangement of 483 nucleotides in the genetic code for KcsA is equal to the allocation of any of the four nucleotides (A, G, C, and T) each multiplied by four per nucleotide position, or 4×4 four hundred and eighty-three times (4^{483}); the probability of generating by chance the correct codon sequence for the 160 amino acids of KcsA, plus one-stop codon, is equal to 64 (the number of codons in the genetic code) multiplied by 64 one hundred and sixty-one times (64^{161}). This could occur once every 46 million years, assuming a mutation rate of one nucleotide every 95,000 years and 2,085 generations per year (*S. lividans* reproduces every 4.2 hours; Palacin et al. 2003); this didactic estimate is based on an average mutation rate of 5.0 base pairs every 10^{10} nucleotides per generation (Drake 1991; Drake et al. 1998; Lynch 2006; Bentley et al. 2008; note that our estimate disregards the tetrameric configuration of KcsA, whose structural and functional assemblage must have required additional time-consuming evolution). But mutations are complex, occur in clusters, occur at different rates within and between genes (= “hot spots” in the genome); and networks

of genes can coevolve (e.g., interacting ion-channel genes), thus increasing and maintaining informational complexity, decreasing uncertainty, and expediting evolution (for detailed discussions on computational methods and theoretical implications see Schneider 2000; Lynch 2005, 2006; Durrett and Schmidt 2008; Stern and Orgogozo 2009). Interestingly, the first ancestors of *Streptomyces* species appeared as recently as 450 million years ago (Chater 2006), and *S. lividans*' clade (*violaceoruber/coelicolor*) apparently separated from its sister clade, *avermitilis*, 220 million years ago (Hatano et al. 1994; Kawamoto and Ochi 1998; Duangmal et al. 2005; Chater and Chandra 2006; Ventura et al. 2007). *S. lividans* is probably a very recent taxon, much younger than 220 million years old, and likely more recent than the 46 million years needed to generate at random one of its plasma membrane proteins, KcsA (below). (2) Because nucleotide transitions (A/G to G/A or C/T to T/C) are more probable than transversions (purine to/from pyrimidine), due to structural and polar affinity between complementary bases, the sole random arrival at the correct arrangement of the 483 nucleotides of KcsA would be reduced to one in two, rather than one in four (above), nucleotides per complementary position of DNA sequence, or 2^{483} (a much faster process than 4^{483}). Note also that redundancy in codon coding (nine amino acids are coded by two codons each, five by four, three by six, one by three, and two by one) determines differential probability of amino acid site allocation; for example, amino acids coded by two codons each (phe, tyr, his, gln, asn, lys, asp, glu, and cys) have a two in 64 probability of being allocated in a peptide sequence; in contrast, amino acids coded by six codons each (leu, ser, and arg) have a six in 64 probability of participating in the protein. This implies that amino acids coded by six codons each would be three times more frequent in KcsA than those coded by two codons each. But this is not the case (Table 1), although amino acids coded by six codons each occur at an average frequency of 9.7% (wide range [$r=3.7-14.9$]), no different than the 9.3% expected by chance, the amino acids coded by two codons each are four times less frequent than those coded by six codons each, they occur at an average frequency of 2.2% [$r=0.0-5.5$], rather than 3.1% expected by chance. Note that cys (coded by two codons) does not even occur in KcsA, although, according to chance, it should be present at a frequency of 3.1%. Further discrepancy between observed and expected frequency of occurrence applies to the rest of the amino acids of KcsA: those coded by four codons each (gly, thr, ala, val, and pro) occur at an average frequency of 8.6% [$r=3.1-13.6$], rather than the 6.2% expected by chance; only ile is coded by three codons and occurs at a frequency of 1.8%, rather than 4.6% expected by chance, while met and trp are coded by one codon each and occur

Table 1 KcsA amino acid (aa) composition ($N=160$ aa, plus one-stop codon)

aa	Acronym	No. of codons in a genetic code ^a	Observed no. of aa (percent) in KcsA	Expected ^b no. of aa (percent) in KcsA
G	gly	4	14 (8.69)	10 (6.25)
S	ser	6	6 (3.73)	15 (9.37)
T	thr	4	12 (7.45)	10 (6.25)
C	cys	2	0 (0.00)	5 (3.12)
Y	tyr	2	5 (3.10)	5 (3.12)
N	asn	2	1 (0.62)	5 (3.12)
Q	gln	2	2 (1.24)	5 (3.12)
K	lys	2	2 (1.24)	5 (3.12)
R	arg	6	17 (10.56)	15 (9.37)
H	his	2	5 (3.10)	5 (3.12)
D	asp	2	4 (2.48)	5 (3.12)
E	Glu	2	9 (5.59)	5 (3.12)
A	ala	4	22 (13.66)	10 (6.25)
V	val	4	16 (9.94)	10 (6.25)
L	leu	6	24 (14.91)	15 (9.37)
I	ile	3	3 (1.86)	7 (4.68)
P	pro	4	5 (3.10)	10 (6.25)
M	met	1	4 (2.48)	3 (1.56)
F	phe	2	4 (2.48)	5 (3.12)
W	trp	1	5 (3.10)	3 (1.56)
Stop		3	1 (0.62)	7 (4.68)
Total		64	161 (≈ 100)	≈ 161 (≈ 100)

The observed composition of amino acids in KcsA differs from what would be expected by chance (Chi-square=52.91; $df=19$; $p \leq 0.001$; the stop codon was excluded), which suggests that directional selection has favored and retained adaptive peptide sequence in KcsA for optimal function. Protein sequence available at NCBI (GenBank Z37969; Swiss-Prot P0A334)

^a Corresponds to the number of codons coding for a specific amino acid

^b Expected values were estimated from the percentile differential allocation of codons coding for each amino acid or stop signal in the genetic code

at frequencies of 2.4% and 3.1%, respectively, rather than the 1.5% expected by chance (Chi-square=52.91; $df=19$; $p \leq 0.001$). (3) Although the frequency of polar (P) plus electrically charged (EC) amino acids ($N=12$) in the genetic code does not differ from that of their non-polar (NP) counterparts ($N=8$; binomial two-tailed test, n.s.), peptide site specificity for plasma membrane hydrophobicity and hydrophilicity follow a non-random pattern in KcsA (Table 2): P+EC versus NP amino acids are unequally distributed across the lipid bilayer (Chi-square=21.20; $df=5$; $p \leq 0.001$); NP residues are significantly more frequent than P+EC amino acids in the hydrophobic regions M1 (binomial two-tailed test; $p=0.014$) and M2 (binomial two-tailed test; $p=0.04$), while P+EC residues in the post-M2 segment are significantly more frequent than NP residues inside the cytoplasmic environment (binomial two-tailed test; $p=0.003$), evidence of strong selective pressure for residue location; the phenomenon is striking considering the overall abundance of polarity in the 160 amino acids of KcsA (77 P+EC vs. 83 NP; binomial two-tailed test, n.s.). (4) Non-random pattern of third-codon sequence and overall nucleotide content are also evident in KcsA: 89% GC versus 11% AT in the third codon position, and 68% GC- versus 32% AT-overall content

(data generated from genomic sequence; NCBI-GenBank Z37969), rather than the 1:1 ratio, in each case, expected by chance (values coincide with high GC frequency for third nucleotide position and high GC-overall content described for *S. lividans*; Wright and Bibb 1992; Fuglsang 2005; Wu et al. 2005); selection at translation has favored the codon bias composition of KcsA, intrinsic to its lineage (Wright and Bibb 1992).

Winning the “Jackpot”

We ran a simulation to generate, under selection, the genomic sequence coding for the 160 amino acids plus one-stop codon of KcsA (Table 3). By blindly drawing from a hat one of four marked marbles (A, G, C, and T), each representing a nucleotide, we generated observed-under-selection values (= number of draws it took until the matching-to-the-sequence-nucleotide was drawn). The total observed-under-selection value of correct nucleotide sequence composition (= 1,799) was even lower than and statistically different from the total expected-under-selection value (= 1,932, or 4+4+4 per codon, above; Chi-square=339.08; $df=160$; $p \leq 0.05$), and it did differ from what would

Table 2 KcsA amino acid site specificity for plasma membrane hydrophobicity and hydrophilicity

Membrane region ^a aa sequence	Amino acid polarity				Total P+EC vs. NP	Statistical difference binomial-test <i>p</i> value
	No. polar (P)	Electrically charged (EC)		No. non- polar (NP)		
		No. acidic	No. basic			
Pre-M1 segment MPPMLSGLLARLVKLLGRHGSALH	5	0	5	15	10/15	n.s.
M1 WRAAGAATVLLVIVLLAGSYLAVLA	5	0	1	19	6/19	0.014
Turret region ERGAPGAQLI	3	1	1	5	5/5	n.s.
Pore-forming loop TYPRALWWSVETA	4	1	1	7	6/7	n.s.
Signature sequence TTVGYG	5	0	0	1	5/1	N/A
Extended region DLYPVTL	2	1	0	4	3/4	N/A
M2 WGRLVAVVVMVAGITSFGLVTAALA	6	0	1	18	7/18	0.04
Post-M2 segment TWFVGREQERRGHFVRHSEKAAEE AYTRTRTRALHERFDRLERMLDDNRR	10	10	15	14	35/14	0.003
Total	40	13	24	83	77/83	n.s.

Polar (P) plus electrically charged (EC) versus non-polar (NP) amino acids are unequally distributed across the plasma membrane (Chi-square=21.20; $df=5$; $p \leq 0.001$; P+EC vs. NP for the signature sequence and extended region were excluded from the Chi-square analysis because their expected values in the contingency table were less than five). Note how NP residues are significantly more frequent than P+EC amino acids in the hydrophobic regions M1 and M2, while P+EC residues in the post-M2 segment are significantly more frequent inside the cytoplasmic environment than NP amino acids. Residue regions correspond to Fig. 1

^a Membrane region after Shealy et al. 2003; protein sequence available at NCBI (GenBank Z37969; Swiss-Prot P0A334)

be expected without selection (Chi-square=7,081.95; $df=160$; $p \leq 0.001$). The correct nucleotides for the first, second, and third positions were generated in an average of 3.3, 4.0, and 4.7 number of steps, respectively, and the correct codons in 11.1 average steps. The effect of selection was such that the “jackprot” generated the 161 codons in one sixth (1,799 vs. 10,304) the number of steps expected by chance! This implies that a protein similar to KcsA could evolve in just eight million years, instead of 46 million years, as computed above.

The Jackprot Simulation: Computer Programs and Online Interface

We wrote a computer program in JAVA APPLET version 1.0 and designed an online interface, *The Jackprot Simulation* <http://faculty.rwu.edu/cbai/JackprotSimulation.htm>, to model a numerical interaction between mutation rate and natural selection during a scenario of polypeptide evolution. Instructors and/or students can access the simulation online and run exemplar statistics identical to those in Table 3, and also cut and paste any cDNA or nucleotide sequence obtained from the National Center for Biotechnology Information (NCBI <http://www.ncbi.nlm.nih.gov/nucleotide/>) or alternative

sources; *The Jackprot Simulation* will generate statistics analogous to those in Table 3. The online interface is friendly and self explanatory, and we provide a comprehensive description of how to use it in S1 in the Electronic Supplementary Materials. The computer programs *JackprotSimulation.java* and *JackprotSupport.java* are also available in S2 and S3 in the Electronic Supplementary Materials.

Conclusions

The “jackprot” helps us understand how natural selection introduces speed into molecular evolution by interacting with mutation rate and retaining complex molecular structures and assemblages of high fitness value. Ion channels are ideal examples to illustrate how biological constraints have driven channel diversification from simpler 2TM-like ancestors to the complex single-K⁺ or multi-subunit-6TM Na⁺ and Ca⁺⁺ proteins. Because of their ubiquitous distribution across taxa, relevance in biology, neurobiology and health-career curricula, and significance in modern behavioral and cognitive studies, ion channels are a sophisticated yet friendly didactic tool for communicating evolutionary principles to all audiences. Alternative perspectives to Darwinian evolution, which

Table 3 Jackprot probabilities applied to the genomic sequence coding for 160 amino acids (aa) and a stop codon of KcsA

aa	Acronym	Codon	No. of simulations under selection required to generate correct nucleotide sequence per position			Total observed with selection ^a	Expected with selection ^b	Expected without selection ^c
			1st	2nd	3rd			
M	met	atg	1	2	4	7	12	64
P	pro	cca	4	6	2	12	12	64
P	pro	ccc	3	5	4	12	12	64
M	met	atg	6	3	9	18	12	64
L	leu	ctg	5	3	5	13	12	64
S	ser	tcc	2	5	2	9	12	64
G	gly	ggt	2	1	2	5	12	64
L	leu	ctt	3	4	3	10	12	64
L	leu	ctg	1	4	1	6	12	64
A	ala	gcc	1	2	2	5	12	64
R	arg	aga	3	1	4	8	12	64
L	leu	ttg	6	1	1	8	12	64
V	val	gtc	3	3	3	9	12	64
K	lys	aaa	2	2	2	6	12	64
L	leu	ctg	1	12	5	18	12	64
L	leu	ctg	2	4	5	11	12	64
L	leu	ctc	3	1	2	6	12	64
G	gly	ggg	10	7	2	19	12	64
R	arg	cgc	3	4	3	10	12	64
H	hys	cac	3	2	5	10	12	64
G	gly	ggc	3	8	9	20	12	64
S	ser	agt	4	3	1	8	12	64
A	ala	gcg	1	1	6	8	12	64
L	leu	ctg	2	3	4	9	12	64
H	hys	cac	1	5	6	12	12	64
W	trp	tgg	4	4	3	11	12	64
R	arg	agg	6	5	2	13	12	64
A	ala	gcc	4	7	1	12	12	64
A	ala	gcg	1	5	8	14	12	64
G	gly	ggt	1	3	1	5	12	64
A	ala	gcc	9	2	2	13	12	64
A	ala	gcg	3	1	2	6	12	64
T	thr	acg	1	1	3	5	12	64
V	val	gtc	10	4	6	20	12	64
L	leu	ctc	2	2	2	6	12	64
L	leu	ctg	1	2	5	8	12	64
V	val	gtg	3	5	2	10	12	64
I	ile	atc	1	1	2	4	12	64
V	val	gtc	2	8	5	15	12	64
L	leu	ctc	5	1	2	8	12	64
L	leu	ctc	1	17	1	19	12	64
A	ala	gcg	2	9	2	13	12	64
G	gly	ggc	1	3	2	6	12	64
S	ser	tcg	1	7	2	10	12	64
Y	tyr	tac	5	1	1	7	12	64

Table 3 (continued)

aa	Acronym	Codon	No. of simulations under selection required to generate correct nucleotide sequence per position			Total observed with selection ^a	Expected with selection ^b	Expected without selection ^c
			1st	2nd	3rd			
L	leu	ttg	1	9	6	16	12	64
A	ala	gcc	1	4	1	6	12	64
V	val	gtc	1	2	2	5	12	64
L	leu	ctg	5	2	1	8	12	64
A	ala	gct	4	4	1	9	12	64
E	glu	gag	1	5	8	14	12	64
R	arg	cgc	2	2	8	12	12	64
G	gly	ggc	1	2	2	5	12	64
A	ala	gca	7	2	2	11	12	64
P	pro	ccg	3	5	5	13	12	64
G	gly	ggc	9	5	1	15	12	64
A	ala	gcg	5	3	7	15	12	64
Q	gln	cag	2	18	2	22	12	64
L	leu	ctg	1	2	3	6	12	64
I	ile	atc	2	2	1	5	12	64
T	thr	acg	2	2	4	8	12	64
Y	tyr	tat	1	3	1	5	12	64
P	pro	ccg	2	5	3	10	12	64
R	arg	cgg	4	1	1	6	12	64
A	ala	gcg	3	4	6	13	12	64
L	leu	ctg	1	5	1	7	12	64
W	trp	tgg	9	8	5	22	12	64
W	trp	tgg	3	4	1	8	12	64
S	ser	tcc	2	2	2	6	12	64
V	val	gtg	2	3	5	10	12	64
E	glu	gag	6	1	3	10	12	64
T	thr	acc	8	4	2	14	12	64
A	ala	gcg	4	2	1	7	12	64
T	thr	acg	1	7	4	12	12	64
T	thr	acc	4	2	10	16	12	64
V	val	gtc	3	2	5	10	12	64
G	gly	ggc	2	2	4	8	12	64
Y	tyr	tac	1	4	2	7	12	64
G	gly	ggc	2	4	5	11	12	64
D	asp	gac	2	3	4	9	12	64
L	leu	ctg	3	1	3	7	12	64
Y	tyr	tac	1	4	1	6	12	64
P	pro	ccc	2	2	7	11	12	64
V	val	gtg	3	3	4	10	12	64
T	thr	act	4	2	14	20	12	64
L	leu	ctg	9	4	1	14	12	64
W	trp	tgg	2	2	1	5	12	64
G	gly	ggc	3	1	1	5	12	64
R	arg	cgg	11	1	6	18	12	64
L	leu	ctc	3	5	6	14	12	64

Table 3 (continued)

aa	Acronym	Codon	No. of simulations under selection required to generate correct nucleotide sequence per position			Total observed with selection ^a	Expected with selection ^b	Expected without selection ^c
			1st	2nd	3rd			
V	val	gtg	8	3	14	25	12	64
A	ala	gcc	3	13	5	21	12	64
V	val	gtg	5	4	6	15	12	64
V	val	gtg	1	4	1	6	12	64
V	val	gtg	2	7	3	12	12	64
M	met	atg	3	2	6	11	12	64
V	val	gtc	3	9	2	14	12	64
A	ala	gcc	7	6	4	17	12	64
G	gly	ggg	3	3	2	8	12	64
I	ile	atc	1	3	1	5	12	64
T	thr	acc	6	2	2	10	12	64
S	ser	tcc	7	3	10	20	12	64
F	phe	ttc	2	3	1	6	12	64
G	gly	ggt	10	3	4	17	12	64
L	leu	ctg	2	8	7	17	12	64
V	val	gtg	4	2	3	9	12	64
T	thr	acc	1	4	1	6	12	64
A	ala	gcc	2	6	2	10	12	64
A	ala	gcg	1	1	9	11	12	64
L	leu	ctg	2	8	1	11	12	64
A	ala	gcc	2	8	1	11	12	64
T	thr	acc	8	5	6	19	12	64
W	trp	tgg	1	2	1	4	12	64
F	phe	ttc	5	4	4	13	12	64
V	val	gtc	1	5	9	15	12	64
G	gly	ggc	6	2	1	9	12	64
R	arg	cgg	2	2	1	5	12	64
E	glu	gaa	6	2	4	12	12	64
Q	gln	caa	4	3	4	11	12	64
E	glu	gag	2	7	3	12	12	64
R	arg	cgc	4	7	7	18	12	64
R	arg	cgg	4	5	3	12	12	64
G	gly	ggc	2	1	6	9	12	64
H	his	cac	1	2	11	14	12	64
F	phe	ttc	5	7	11	23	12	64
V	val	gtg	6	2	11	19	12	64
R	arg	cgc	2	1	2	5	12	64
H	his	cac	3	1	4	8	12	64
S	ser	tcc	4	7	1	12	12	64
E	blu	gag	9	1	6	16	12	64
K	lys	aag	4	8	2	14	12	64
A	ala	gcc	9	1	3	13	12	64
A	ala	gcc	1	7	3	11	12	64
E	glu	gag	4	3	13	20	12	64
E	glu	gag	9	3	4	16	12	64

Table 3 (continued)

aa	Acronym	Codon	No. of simulations under selection required to generate correct nucleotide sequence per position			Total observed with selection ^a	Expected with selection ^b	Expected without selection ^c
			1st	2nd	3rd			
A	ala	gcg	1	1	1	3	12	64
Y	tyr	tac	2	1	5	8	12	64
T	thr	acg	11	5	2	18	12	64
R	arg	cgg	1	1	5	7	12	64
T	thr	acg	2	5	5	12	12	64
T	thr	acc	2	3	3	8	12	64
R	arg	cgg	1	5	2	8	12	64
A	ala	gcg	7	2	4	13	12	64
L	leu	ctg	4	5	8	17	12	64
H	his	cac	3	3	6	12	12	64
E	glu	gag	3	2	6	11	12	64
R	arg	cgt	10	8	9	27	12	64
F	phe	ttc	4	13	2	19	12	64
D	asp	gac	4	1	3	8	12	64
R	arg	cgt	1	8	1	10	12	64
L	leu	ttg	1	2	3	6	12	64
E	glu	gag	1	2	6	9	12	64
R	arg	cga	3	5	2	10	12	64
M	met	atg	6	1	3	10	12	64
L	leu	ctc	1	4	2	7	12	64
D	asp	gac	2	8	3	13	12	64
D	asp	gac	2	4	2	8	12	64
N	asn	aac	1	1	5	7	12	64
R	arg	cgc	1	22	3	26	12	64
R	arg	cgg	3	2	1	6	12	64
Stop		tga	3	2	4	9	12	64
Total			543	646	610	1,799	1,932	10,304
Mean			3.37	4.01	3.78	11.17	12	64
Mode			1	2	2	8	12	64
Median			3	3	3	10	12	64
Max			11	22	14	27	12	64
Min			1	1	1	3	12	64

The total observed-under-selection value of correct nucleotide sequence composition (= 1,799) is lower than and statistically different from the total expected-under-selection value (= 1,932; Chi-square=339.08; $df=160$; $p\leq 0.05$), and it does differ from what would be expected by chance without selection (Chi-square=7081.95; $df=160$; $p\leq 0.001$). The tinkering effect of selection on mutation rate expedites evolution and the jackpot model shows it. Protein sequence available at NCBI (GenBank Z37969; Swiss-Prot P0A334)

^a Data generated from under-selection simulations in which one of four marked marbles (A, G, C, and T), each representing a nucleotide, was drawn from a hat until the matching-to-the-sequence-nucleotide was generated, e.g., atg=one, two, and four times/first, second, and third positions, respectively

^b Corresponds to one in 12, or one divided by 4+4+4, i.e., the summation of the individual probabilities for each nucleotide position

^c Corresponds to one in 64, or one divided by 4×4×4, i.e., the total number of possible nucleotides per position multiplied by itself three times

attribute randomness to molecular change, deleterious nature to single-gene mutations, insufficient geological time or population size for molecular improvements to occur, or

invoke “design creationism” to account for complexity in molecular structures and biological processes, are empirically unfounded and conceptually wrong.

Acknowledgments We thank Dr. Aguan Wei for comments on the manuscript and Dr. Lawrence Salkoff and the Salkoff Lab for motivating us to explore the complexity and evolutionary significance of ion channels. G. Paz-y-Miño C. is supported by the Office of Faculty Development at UMass Dartmouth (Innovation in Teaching Awards AY0910 and Undergraduate Research Grants F09) and A. Espinosa by NIH-NCRR grant no. 2 P20RR16457-04. One reviewer provided important comments that helped us improve the manuscript.

References

- Anderson PAV, Greenberg RM. Phylogeny of ion channels: clues to structure and function. *Comp Biochem Physiol*. 2001;129:17–28.
- Behe MJ. Molecular machines: experimental support for the design inference. *Cos Pur*. 1998;1:27–35.
- Behe MJ. Reply to my critics: a response to reviews of Darwin's black box: the biochemical challenge to evolution. *Biol Philos*. 2001;16:685–709.
- Behe MJ. The challenge of irreducible complexity. *Nat Hist*. 2002;111:74.
- Behe MJ. The edge of evolution. New York, NY: Free Press; 2007.
- Behe MJ. Waiting longer for two mutations. *Genetics*. 2009;181:819–20.
- Behe MJ, Snoke DW. Simulating evolution by gene duplication of protein features that require multiple amino acid residues. *Prot Sci*. 2004;13:2651–64.
- Behe MJ, Snoke DW. A response to Michael Lynch. *Prot Sci*. 2005;14:2226–7.
- Bentley SD, Corton C, Brown SE, Barron A, Clark L, Doggett J, et al. Genome of the Actinomucete plant pathogen *Clavibacter michiganensis* subsp. *sepedonicus* suggests recent niche adaptation. *J Bacteriol*. 2008;190:2150–60.
- Capener CE, Kim HJ, Arinaminpathy Y, Sansom MSP. Ion channels: structural bioinformatics and modeling. *Hum Mol Genet*. 2002;11:2425–33.
- Chater KF. *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Phil Trans R Soc B*. 2006;361:761–68.
- Chater KF, Chandra G. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev*. 2006;30:651–72.
- Derst C, Karschin A. Evolutionary link between prokaryotic and eukaryotic K⁺ channels. *J Exp Biol*. 1998;201:2791–99.
- Doyle DA. Structural themes in ion channels. *Eur Biophys J*. 2004;33:175–9.
- Doyle DA, Morais-Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, et al. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science*. 1998;280:69–77.
- Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci*. 1991;88:7160–4.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics*. 1998;148:1667–86.
- Duangmal K, Ward AC, Goodfellow M. Selective isolation of members of the *Streptomyces violaceoruber* clade from soil. *FEMS Microbiol Lett*. 2005;245:321–27.
- Durell SR, Hao Y, Nakamura T, Bakker EP, Guy R. Evolutionary relationships between K⁺ channels and symporters. *Biophys J*. 1999;77:775–88.
- Durrett R, Schmidt D. Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics*. 2008;180:1501–9.
- Durrett R, Schmidt D. Reply to Michael Behe. *Genetics*. 2009;181:821–2.
- Forrest BC, Gross PR. Biochemistry by design. *Trends Biochem Sci*. 2007;32:301–10.
- Fuglsang A. Intragenic position of UUA codons in streptomycetes. *Microbiology*. 2005;151:3150–2.
- Galliot B, Quiquand M, Ghila L, de Rosa R, Miljkovic-Licina M, Chera S. Origins of neurogenesis, a cnidarian view. *Dev Biol*. 2009;332:2–24.
- Hatano K, Tamura T, Nishii T. Taxonomic status of *Streptomyces coelicolor* A3(2) and *Streptomyces lividans* 66. *Actinomycetologica*. 1994;8:47–50.
- Hermodson M. Editorial and position papers. *Prot Sci*. 2005;14:2215–6.
- Hill AS, Nishino A, Nakajo K, Zhang G, Fineman JR, Selzer ME, et al. Ion channel clustering at the axon initial segment and node of Ranvier evolved sequentially in early chordates. *PLoS Genet*. 2008;4:e1000317. doi:10.1371/journal.pgen.1000317.
- Kawamoto S, Ochi K. Comparative ribosomal protein (L11 and L30) sequence analyses of several *Streptomyces* spp. commonly used in genetic studies. *Int J System Bacteriol*. 1998;48:597–600.
- Kress GJ, Mennerick S. Action potential initiation and propagation: upstream influences on neurotransmission. *Neuroscience*. 2009;158:211–22.
- Kunzelmann K. Ion channels and cancer. *J Membr Biol*. 2005;205:159–73.
- Long M, Betrán E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet*. 2003;4:865–75.
- Lu Z, Klem AM, Ramu Y. Ion conduction pore is conserved among potassium channels. *Nature*. 2001;413:809–13.
- Luskin C, Gage LP. A reply to Francis Collins's Darwinian arguments for common ancestry of apes and humans. In: House HW, editor. *Intelligent Design 101*. Grand Rapids: Kregel Publications; 2008. p. 215–35.
- Lynch M. Simple evolutionary pathways to complex proteins. *Prot Sci*. 2005;14:2217–25.
- Lynch M. The origins of eukaryotic gene structure. *Mol Biol Evol*. 2006;23:450–68.
- Martinac B, Saimi Y, Kung C. Ion channels in microbes. *Phys Rev*. 2008;88:1449–90.
- Miller C. An overview of the potassium channel family. *Gen Biol*. 2000;1:4.1–5.
- Miller G. On the origin of the nervous system. *Science*. 2009;325: 24–6.
- Nelson PA. The role of theology in current evolutionary reasoning. *Biol Philos*. 1996;11:493–517.
- Palacin A, de la Fuente R, Valle I, Rivas LA, Mellado RP. *Streptomyces lividans* contains a minimum functional signal recognition particle that is involved in protein secretion. *Microbiology*. 2003;149:2435–42.
- Paz-y-Miño C G, Espinosa A. Assessment of biology majors versus non-majors views on evolution, creationism and intelligent design. *Evo Edu Outreach*. 2009a;2:75–83.
- Paz-y-Miño C G, Espinosa A. Acceptance of evolution increases with student academic level: a comparison between a secular and a religious college. *Evo Edu Outreach*. 2009b;4:655–75.
- Paz-y-Miño C G, Espinosa A. Using horizontal gene transfer and common descent to depict evolution and contrast it with common design. *J Eukaryot Microbiol*. 2010a;57:11–8.
- Paz-y-Miño C G, Espinosa A. New England faculty and college students differ in their views about evolution, creationism, intelligent design, and religiosity. *Evo Edu Outreach*. 2010b. doi:10.1007/s12052-010-0298-x.
- Paz-y-Miño C G, Espinosa A. On the theory of evolution versus the concept of evolution: three observations. *Evo Edu Outreach*. 2010c. doi:10.1007/s12052-010-0299-9.
- Pennock RT, editor. *Intelligent Design Creationism and its Critics*. Cambridge, MA: Massachusetts Institute of Technology; 2001.
- Petto AJ, Godfrey LR, editors. *Scientists confront intelligent design and creationism*. New York, NY: WW Norton and Company; 2007.
- Pichon Y, Prime L, Benquet P, Tiaho F. Some aspects of the physiological role of ion channels in the nervous system. *Eur Biophys J*. 2004;33:211–26.

- Rogers M, Tang L, Madge DJ, Stevens EB. The role of sodium channels in neuropathic pain. *Semin Cell Dev Biol.* 2006;17:571–81.
- Sansom MSP, Shrivastava IH, Bright JN, Tate J, Capener CE, Biggin PC. Potassium channels: structures, models, simulations. *Biochim Biophys Acta.* 2002;1565:294–307.
- Schneider TD. Evolution of biological information. *Nuc Aci Res.* 2000;28:2794–9.
- Schneiderman JS, Allmon WD, editors. For the rock record: geologists on intelligent design. Berkeley, CA: University of California Press; 2009.
- Schrempf H, Schmidt O, Kümmerlen R, Hinnah S, Müller D, Betzler M, et al. A prokaryotic potassium ion channel with two predicted transmembrane segments from *Streptomyces lividans*. *EMBO J.* 1995;14:5170–8.
- Shealy RT, Murphy AD, Ramarathnam R, Jakobsson E, Subaramaniam S. Sequence-function analysis of the K⁺-selective family of ion channels using a comprehensive alignment and the KcsA channel structure. *Biophys J.* 2003;84:2929–42.
- Sontheimer H. An unexpected role for ion channels in brain tumor metastasis. *Exp Biol Med.* 2008;233:779–91.
- Stern DL, Orgogozo V. Is genetic evolution predictable? *Science.* 2009;323:746–51.
- Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, et al. Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiol Biol Rev.* 2007;71:495–548.
- Williamson IM, Alvis SJ, East JM, Lee AG. The potassium channel KcsA and its interaction with the lipid bilayer. *Cell Mol Life Sci.* 2003;60:1581–90.
- Wright F, Bibb MJ. Codon usage in the G+C-rich *Streptomyces* genome. *Gene.* 1992;113:55–65.
- Wu G, Culley DE, Zhang W. Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. *Microbiology.* 2005;151:2175–87.
- Young M, Edis T, editors. Why intelligent design fails. New Jersey: Rutgers University Press; 2004.