PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

CrossMark click for updates

Phil. Trans. R. Soc. B (2012) **367**, 2584–2593
doi:10.1098/rstb.2012.0076

Research

Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues

Pau Creixell[1], Erwin M. Schoof[1], Chris Soon Heng Tan[2] and Rune Linding[1,*]

[1]*Cellular Signal Integration Group (C-SIG), Center for Biological Sequence Analysis (CBS), Department of Systems Biology, Technical University of Denmark (DTU), DK-2800 Lyngby, Denmark*
[2]*Center for Molecular Medicine of the Austrian Academy of Sciences (CeMM), Vienna, Austria*

As François Jacob pointed out over 30 years ago, evolution is a tinkering process, and, as such, relies on the genetic diversity produced by mutation subsequently shaped by Darwinian selection. However, there is one implicit assumption that is made when studying this tinkering process; it is typically assumed that all amino acid residues are equally likely to mutate or to result from a mutation. Here, by reconstructing ancestral sequences and computing mutational probabilities for all the amino acid residues, we refute this assumption and show extensive inequalities between different residues in terms of their mutational activity. Moreover, we highlight the importance of the genetic code and physico-chemical properties of the amino acid residues as likely causes of these inequalities and uncover serine as a mutational hot spot. Finally, we explore the consequences that these different mutational properties have on phosphorylation site evolution, showing that a higher degree of evolvability exists for phosphorylated threonine and, to a lesser extent, serine in comparison with tyrosine residues. As exemplified by the suppression of serine's mutational activity in phosphorylation sites, our results suggest that the cell can fine-tune the mutational activities of amino acid residues when they reside in functional protein regions.

Keywords: amino acid evolvability; mutation; phosphorylation site evolution

1. INTRODUCTION

Cells are constantly evolving in a race for adaptation to dynamic environmental challenges. As described by François Jacob over three decades ago [1], this process is more analogous to tinkering than to free design, in the sense that nature does not create a new protein function from a blank canvas nor with unlimited resources, but instead evolves through innovation with existing proteins (figure 1*a*,*b*). In line with this principle of functionalization by tinkering, most general models of protein evolution (e.g. duplication–divergence [2], neofunctionalization or subfunctionalization [3]) are based on gene duplication being the main source of new genes, proteins and consequently new cellular function.

In this study, we aim to extend the principle of tinkering in evolution, initially developed by Jacob [1], to include the effect the genetic code has on protein evolution. Our hypothesis is that evolution is not only constrained because it needs to tinker with existing proteins; it is also affected by the genetic code in the sense that genetic variation is not generated by substituting amino acid residues from the evolving protein at random, but instead the genetic code dictates that some amino acid substitutions will be more frequent than others (figure 1*c*).

2. THE INFLUENCE OF THE GENETIC CODE ON MUTATIONAL PATHS

In essence, substitutions between amino acid residues that are far away from each other in mutational space are less likely than between residues that are close to each other (figure 2). For instance, if we had to compute the probability of every amino acid residue to be the target of a mutation from methionine, we would have to consider the mutational distance and the physico-chemical similarity between the two residues. Isoleucine, leucine, phenylalanine, valine, threonine, lysine and arginine are, in terms of mutational distance, the closest residues to methionine, because they are all just one nucleotide mutation away from it (figure 2*a*). Alanine, valine, isoleucine and leucine are the closest residues in physico-chemical distance, because they are small hydrophobic residues similar to methionine (figure 2*b*). Combining these two distances (mutational and physico-chemical) that determine the genetic diversity generated and selection of protein variants, one can rationalize the amino acid substitution frequencies observed along evolution (figure 2*c*).

*Author for correspondence (linding@cbs.dtu.dk; www.lindinglab.org).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2012.0076 or via http://rstb.royalsocietypublishing.org.

One contribution of 13 to a Theme Issue 'The evolution of protein phosphorylation'.

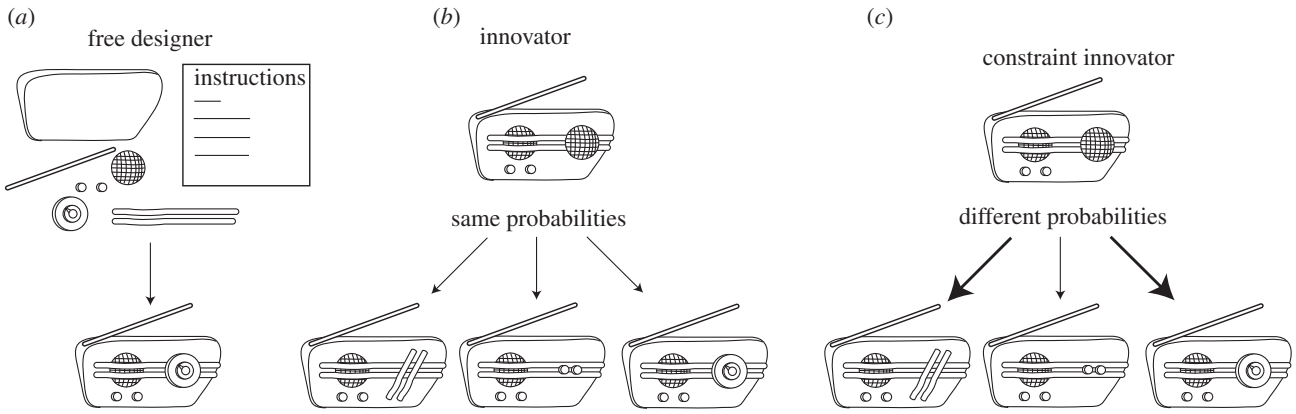


Figure 1. Creative methodologies and evolution. As an analogy to protein evolution in the hunt for new protein function, we have illustrated different strategies to design a radio. (*a*) As Jacob described several years ago, nature does not evolve by creating de novo protein function from a blank canvas resembling a free designer who can build a radio using some predefined instructions and any imaginable radio parts. (*b*) Instead, nature is more of an innovator who tinkers with existing proteins before finding new protein function by a process of mutation and selection. Following with our analogy, the tinkerer does not generate a radio from scratch, but it tinkers with existing devices by combining and substituting pieces, and the best design is selected for. (*c*) In this study, we extend this concept by highlighting the fact that the sources and targets of mutations cannot be chosen arbitrarily, but instead some amino acid substitutions will be more likely than others (different probabilities). Unlike in (*b*), where different substitution probabilities are not considered, tinkering with the loudspeaker in the radio is more likely to lead to some radio parts than others.

Next, we tested the validity and generality of this influence the genetic code has on mutational paths. In principle, one would expect the effect of the genetic code to decrease with time, because longer evolutionary distances would allow several mutations in the same amino acid residues to become more likely (figure 3*a*). As briefly suggested earlier (figure 2*c*), regardless of what amino acid substitution is more probable, purifying selection will act subsequently to disfavour substitutions that would lead to radical changes in the physico-chemical properties of the protein residue. Thus, unlike the effect of the genetic code, we expect the effect of the physico-chemical properties of the different amino acids to remain constant over time. To test the influence of the genetic code and physico-chemical properties on protein evolution, we reconstructed ancestral sequences at different evolutionary distances between humans and other vertebrates (figure 3*b* and see §7 for further details). Supporting our hypothesis, we indeed observed different targets of mutation at different evolutionary distances (figure 3*c*), with mutational targets closer in mutational space for shorter evolutionary distances (L1: human–orangutan) and less influenced by mutational distance for longer evolutionary distances (L7: human–frog).

3. MUTATIONAL PROPERTIES OF AMINO ACID RESIDUES

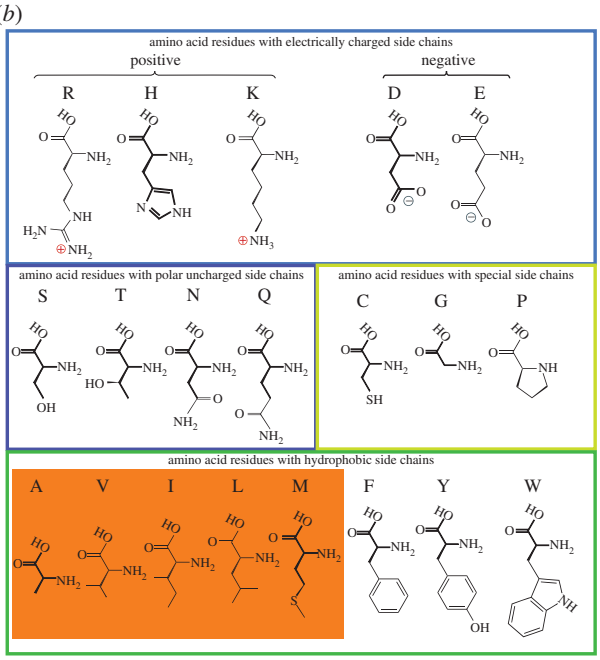
By expanding our analysis, we computed matrices to reflect the probability of every amino acid residue to mutate and become every other amino acid residue at different evolutionary distances (see the electronic supplementary material, table S1). To better describe the different mutational properties of amino acid residues represented in these matrices, we introduce two new terms, mutability and targetability. We define mutability as the probability of an amino acid residue to mutate, and targetability as the probability of an amino acid to be the result of a mutation. By extension, we have termed our matrices (which effectively contain mutability for each residue on their rows, targetability on their columns and conservation on their diagonal) mutability targetability (MUTA) matrices. The rationale behind developing our MUTA matrices is similar to the rationale behind matrices such as point accepted mutation (PAM) [4] or blocks of amino acid substitution matrix (BLOSUM) [5] but they differ fundamentally in their goal and, in consequence, also in the information they contain (figure 4). While matrices such as PAM or BLOSUM, default matrices used by popular tools such as BLAST (Basic Local Alignment Search Tool) [6], reflect the tendency of some amino acid residues to appear in a multiple sequence alignment of homologue proteins, MUTA matrices describe the probability of the different amino acid residues to mutate (mutability) and be targets of mutation (targetability). Given that MUTA matrices are derived not from conserved blocks but instead from a large range of sequences with different degrees of evolvability, they are likely to be more useful than previous matrices for evolutionary analysis (e.g. the characterization of phosphorylation sites or other protein sequences that do not necessarily reside in conserved protein regions).

To better visualize every amino acid residue's mutational properties, one can represent each amino acid residue as a data point on an x–y scatter plot, i.e. mutability–targetability plot (figure 5*a*,*b*). Following this strategy, we show mutability and targetability for every amino acid residue at different evolutionary distance (figure 5*c*); it is apparent that, contrary to common assumption, different amino acid residues have different mutational properties (i.e. mutability and targetability). Moreover, it is evident that there is a correlation between mutability and targetability whereby amino acids that tend to mutate more are

(*a*)

	T	C	A	G	
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	—	—	A
	Leu	Ser	—	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

(*b*)



(*c*)

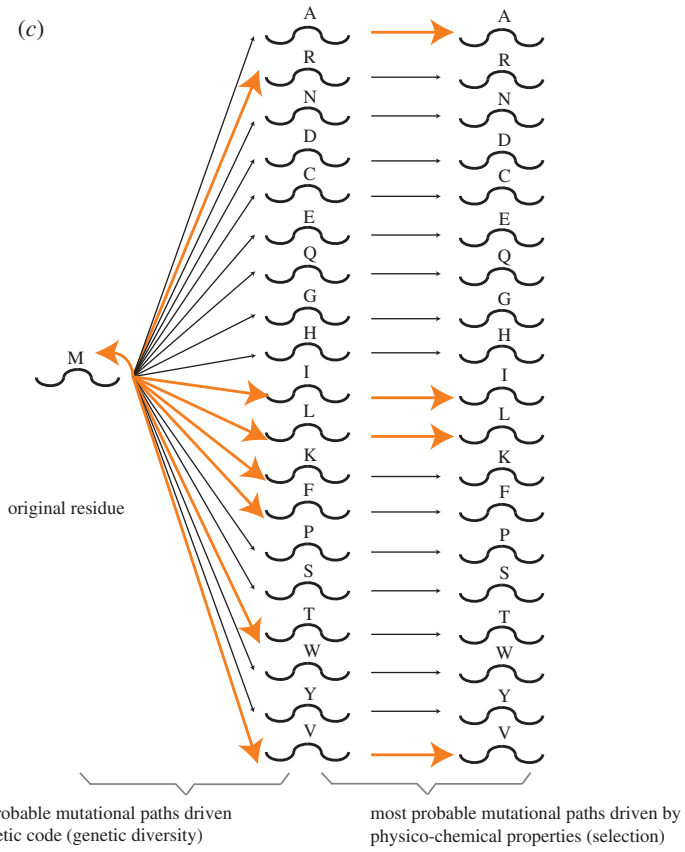


Figure 2. Exploring evolutionary mutational targets. (*a*) In this codon table, we have highlighted amino acid residues that are close to (one nucleotide mutation away from) methionine in mutational space. (*b*) In this table of physico-chemical properties of different amino acid residues, we have highlighted amino acid residues that are close (similar) to methionine in physico-chemical space (adapted from www.wikipedia.org). (*c*) Combining mutational and physico-chemical space allows rationalization of why some mutational paths (amino acid substitutions) are more frequent than others. Here, we have highlighted in orange the most preferred mutational paths owing to short mutational distance (first arrow) and short physico-chemical distance (second arrow). Residue conservation has been illustrated as a loop, and it should be considered as another possible mutational path with very short mutational and physico-chemical distance.

also more likely targets of mutations. This correlation indicates that the dynamic system of amino acid residue substitutions and frequencies lies in equilibrium in a stable steady state, where all the residues balance out residue loss and gain after mutation, which results in only small frequency fluctuations over time.

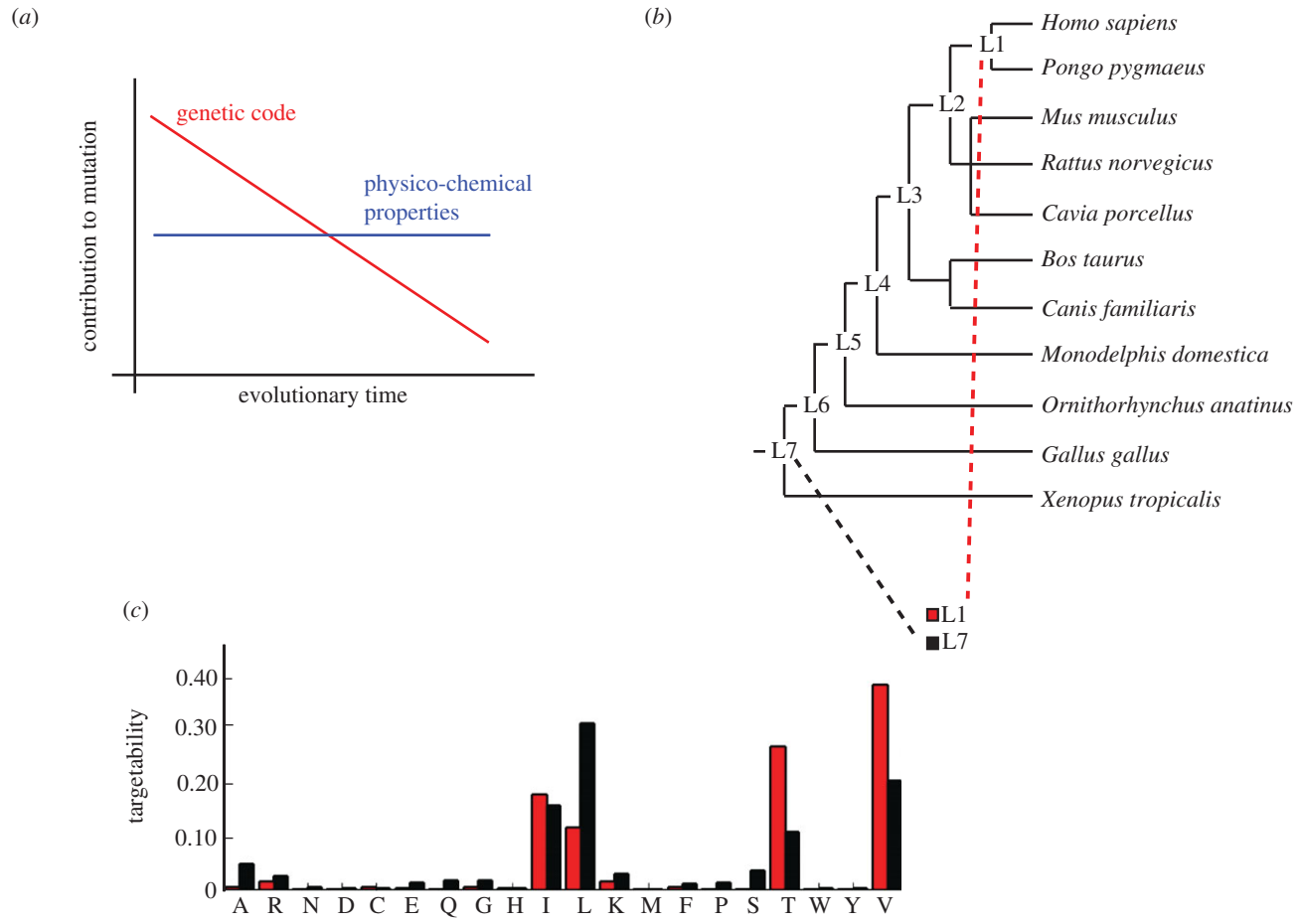


Figure 3. Exploring evolutionary mutational targets. (*a*) The relative contribution of the genetic code (by disfavouring amino acid residue substitutions that require several nucleotide mutations) and the physico-chemical properties (by disfavouring amino acid residue substitutions between dissimilar residues) to mutation will vary over evolutionary time. The restrictions imposed by the genetic code will have higher influence when comparing short-evolutionary distances, whereas the physico-chemical properties of amino acid residues will have a constant influence, because selection against radical changes in physico-chemical space will always be applied before a mutation becomes fixed. (*b*) Graphical representation of the phylogenetic tree whose ancestral sequences (L1, L2, L3, L4, L5, L6 and L7) we have reconstructed as described in §7. (*c*) Here, we confirm the principle described in (*a*), by comparing mutational targets of methionine between L1 and human and between L7 and human and showing that in shorter evolutionary distances (L1: red), methionine tends to mutate only to residues that are one nucleotide mutation away, while for longer times (L7: black), more targets are possible.

In contrast, large discrepancies between mutability and targetability would lead to large fluctuations in frequency and, with time, to extinction or perpetuation (figure 5*b*). This correlation between mutability and targetability is therefore the only path to prevent amino acid residue extinction or perpetuation.

It is also apparent from our mutability–targetability plots that different residues use different evolutionary paths to hold their frequency stable. In one extreme, serine evolves very fast by mutating very often, while also being a more likely target of mutations, i.e. high mutability and high targetability. At the opposite extreme, tryptophan does not mutate frequently, but at the same time it is not a frequent target of mutations either, i.e. low mutability and targetability (figure 5*c*). Analogous to how different nucleotide or protein sequences can evolve at different speed, here we have uncovered that even individual amino acid residues can be fast- or slow-evolving (e.g. serine and tryptophan, respectively). Next, we will investigate the causes and consequences of the mutational properties of the different residues.

4. POSSIBLE CAUSES FOR DIFFERENT MUTATIONAL PROPERTIES OF AMINO ACID RESIDUES

The fact that serine is the fastest-evolving amino acid residue can perhaps give us some insights into why different amino acid residues would present different mutational properties. First, considering mutational space (figure 2*a*), it is apparent that serine is a unique residue in that it is the only amino acid whose six codons are distributed in two different groups, AGY and TCN, that are so far apart from each other (at least two nucleotide mutations away). As a consequence, serine will be more easily reached from another amino acid after mutation, i.e. it is very close in mutational space to most other amino acid residues (in most cases, only one nucleotide mutation away). In addition, from the perspective of physico-chemical distance (figure 2*b*), serine's moderate physico-chemical properties, without a bulky or charged side chain, make it less likely that the amino acid substitution will be rejected by selection (figure 2*c*), because it is close in physico-chemical space to most other amino acid residues.

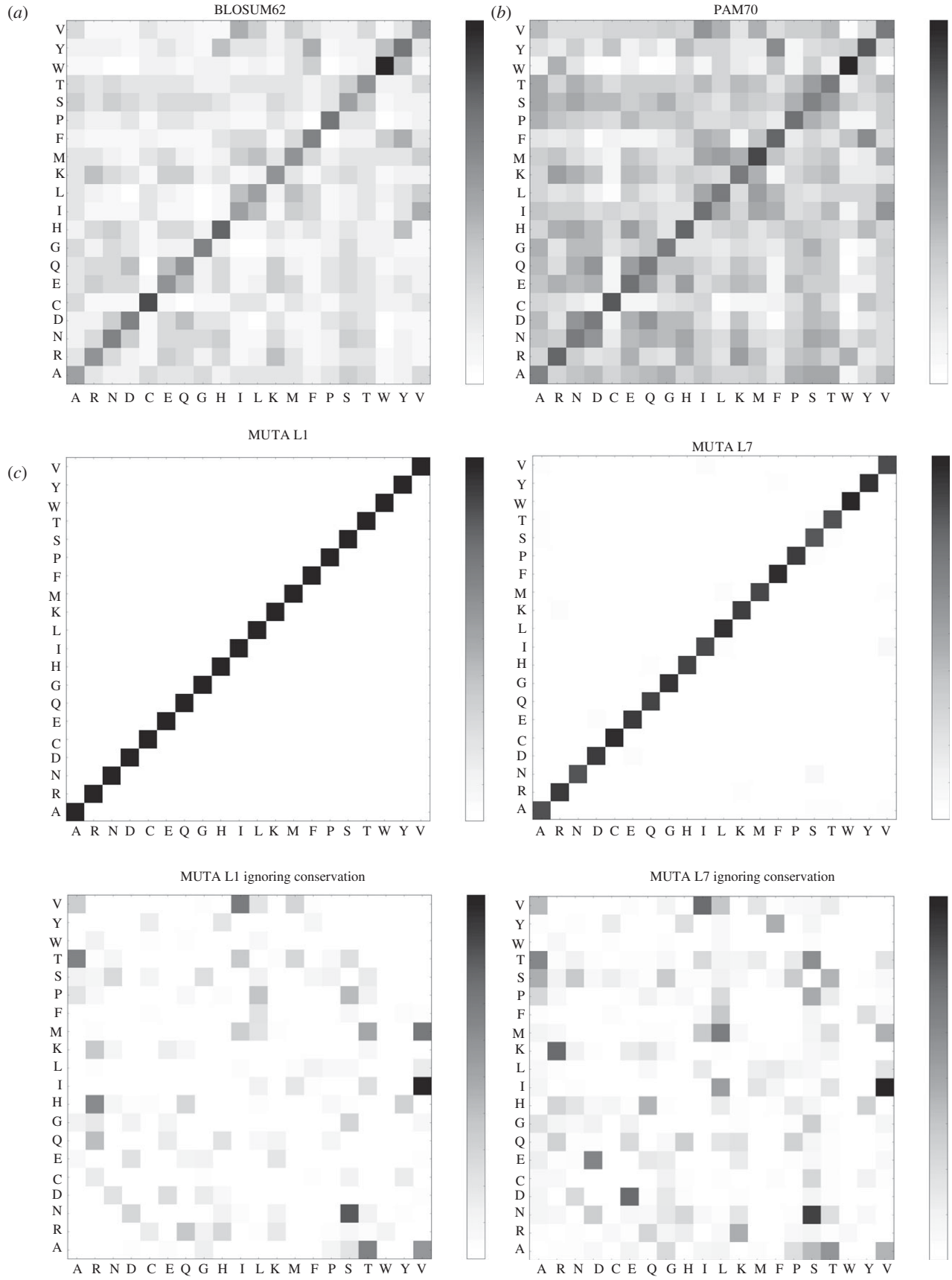


Figure 4. A comparison of amino acid substitution matrices. (*a*) Representation of a normalized version of the BLOSUM62 matrix. (*b*) Representation of a normalized version of the PAM70 matrix. (*c*) Representation of our L1 and L7 MUTA matrices, including versions without conservation (bottom) in order to better visualize the non-conservative amino acid substitutions.

In comparison with serine, the other two amino acid residues coded by six codons (leucine and arginine) do not combine such mutational and physico-chemical proximity with other amino acid residues and, in consequence, are not as fast-evolving as serine. Despite the fact that leucine's physico-chemical properties are

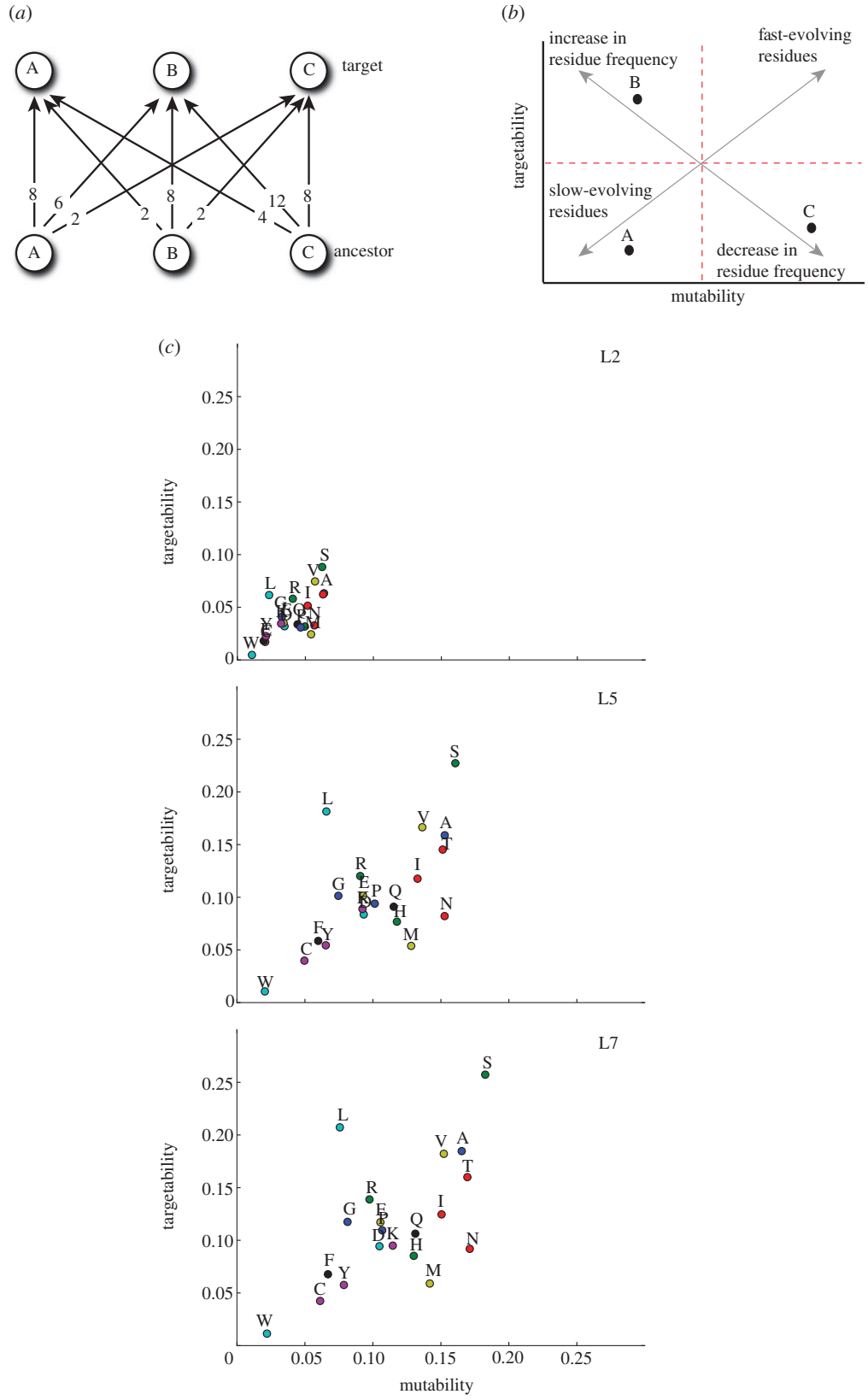


Figure 5. Mutability–targetability plots. (*a*) Toy model to represent three mutable objects and how they can evolve, with each letter representing one element at the ancestral (bottom) or target (top) sequence and each arrow representing the frequency of every possible mutational path. (*b*) Any mutable system, such as the one represented in (*a*), can be represented in a mutability–targetability plot, an *x*–*y* scatter plot where each element (e.g. A–C) is located in a precise coordinate depending on their mutational properties, i.e. how often it mutates (mutability) and how often it is the result of a mutation from another residue (targetability). Depending on their location, we can consider the mutable elements fast or slow evolving (high mutability and high targetability or low mutability and low targetability, respectively) or likely to increase or decrease in frequency (low mutability and high targetability or high mutability and low targetability, respectively). (*c*) Mutability–targetability plots computed for all the amino acid residues at different evolutionary distance (L2, L5 and L7). In order to avoid frequency-related biases, we normalized all mutation frequencies before computing mutabilities and targetabilities for each amino acid (for more information refer to §7).

(like in the case of serine) relatively moderate from a mutational perspective (figure 2*a*), unlike serine's, the two groups of codons that code for leucine, CTN and TTR, are relatively close to each another. As a direct consequence of this close mutational distance between the two groups of codons, the two extra codons that leucine is coded by (TTR) only provide leucine with direct mutational access to six extra codons, compared with eight extra codons that can be directly accessed from serine's two extra codons (AGY). In addition, half of the new codons that can be accessed from leucine's two extra codons are stop codons (TAA, TAG and TGA). Therefore, it can be concluded that, given their mutational proximity to themselves and to stop codons, the six leucine codons cannot contribute to making leucine a more mutable and targetable residue.

On the other hand, arginine which is coded, similar to leucine, by two groups of codons that are relatively close to each other in mutational space (CGN and AGR), would have the potential to have higher mutability and targetability but is probably affected by its extreme physico-chemical properties (charged and large residue), preventing many amino acid substitutions due to natural selection acting against them.

Overall, no other amino acid residue is encoded by as many codons so far apart from each other in mutational space which, combined with its weaker physico-chemical properties, make serine a fast-evolving, mutational hub.

In conclusion, despite the fact that other causes such as bioenergetic costs or tendency to reside in fast-evolving protein regions are also plausible explanations for the mutational properties of the differences residues, we argue that these differences are founded on the mutational and physico-chemical distance from each amino acid residue to every other one of them.

5. IMPLICATIONS FOR PHOSPHORYLATION SITE EVOLUTION

Having described the general mutational properties of the different amino acid residues, we wanted to investigate to what extent cells can modulate these general properties for specific residues. Given the large mutational differences between serine (the fastest-evolving amino acid residue), threonine (a relatively high-evolving residue) and tyrosine (a rather slow-evolving residue), we investigated the consequences that different mutability and targetability may have for protein phosphorylation and evolution of phosphorylation sites.

If these general mutational properties were maintained in phosphorylation sites, one would expect to see fast removal of non-functional phosphorylation sites and fast introduction of a high number of new phosphorylation sites for fast-evolving residues (with high mutability and targetability) like serine or threonine. On the contrary, one would expect to see higher conservation for slow-evolving residues such as tyrosine. We have illustrated these different scenarios for serine, threonine and tyrosine (figure 6*a*).

To test this hypothesis, we computed the sequence conservation of human phosphorylated serines, threonines and tyrosines, and used the sequence conservation of these residues regardless of phosphorylated state as baseline for comparison (figure 6*b*). In addition, to discard the possibility that our results are driven by difference in the likelihood of residues to reside on disordered regions of proteins, we included disorder predictions in our results. In general, our results show the expected trend with phosphorylated residues being more conserved than non-phosphorylated residues, highlighting the likelihood that these are functional sites [7]. Moreover, in line with the general trend for the three residues observed earlier (figure 5*c*), phosphorylated serines and threonines are also much more conserved than phosphorylated tyrosines. Nevertheless, our results (figure 6*b*) also highlight some important subtleties that differ from our previous observations that serine is the most mutationally active residue (figure 5*c*); for instance, we observed a higher degree of conservation for phosphorylated serines than for phosphorylated threonines. Since the overall trend is maintained when taking into account disorder predictions, we can conclude that these observations are not driven by different disorder propensity of the different amino acid residues.

Moreover, we computed the fraction of amino acid residues that were excluded from our analysis because they reside in alignment gaps (see the electronic supplementary material, figure S1), which allowed us to refute the possibility that our observations could be explained by major differences in the propensity of different residues to reside in alignment gaps. In theory, a higher gap propensity of tyrosine (and to a lesser extend threonine) with respect to serine could be a trivial explanation for the different degrees of conservation we observe, because we would have excluded them from our analysis, but the gap propensities we computed do not support this hypothesis.

These results suggest that the cell is indeed capable of modulating the general mutational properties of amino acid residues under special circumstances. Moreover, the higher conservation of phosphorylated serines in comparison with phosphorylated threonines (observation which is in agreement with previous published work [8]) suggests that serine phosphorylation sites have more ancient functional properties, whereas threonine phosphorylation sites have more recent ones. Finally, it is perhaps surprising that the phosphotyrosine system, the signalling system that has appeared most recently in evolution [9], also presents the highest conservation. However, this apparent contradiction can be resolved if this system did not evolve gradually, but instead it evolved by a sudden burst (as supported in the literature [9–11]) and has subsequently remained more conserved (at least at the sequence level) than the phosphoserine or phosphothreonine systems.

6. DISCUSSION AND CONCLUSIONS

In this study, we have uncovered natural forces driving different mutability (probability that a given amino acid residue will be mutated) for different amino acid residues as well as different targetability (probability that a given amino acid residue will be the result of a mutation). These inequalities have made apparent different evolutionary paths for different amino acid residues, with some being slow-evolving and relying for their existence on high conservation (low

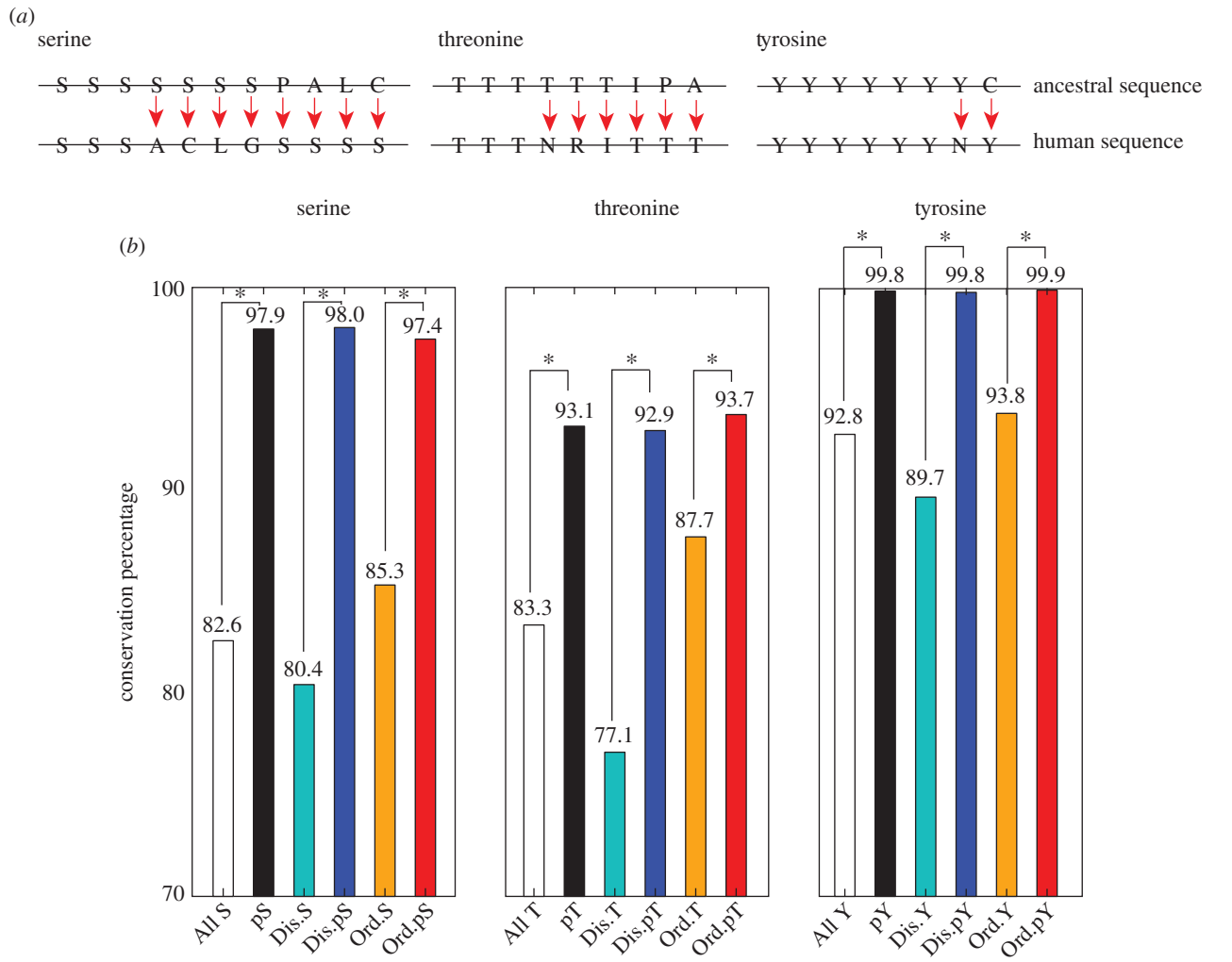


Figure 6. Phosphorylation site evolution. (*a*) As a direct consequence of our previous results, our hypothesis was that phosphorylation at serine, with the highest mutability and targetability rates, would evolve faster than phosphorylation at threonine, which has slightly lower but still relatively high mutability and targetability rates, and much faster than tyrosine, with very low mutability and targetability rates. (*b*) For each phosphorylatable amino acid residue, we have computed the fraction of phosphorylated residue (in black) that is conserved (same amino acid) in the ancestor L7. For comparison, we have computed conservation fraction for the residue, regardless of phosphorylated state (white) and conditional on whether the residue resides in a disordered (cyan both phosphorylated and unphosphorylated residues and blue only phosphorylated residues) or ordered protein region (orange both phosphorylated and unphosphorylated residues and red phosphorylated). Statistical significance ($*p < 0.05$) was assessed by Fisher's exact test.

mutability and targetability), such as tryptophan, and others being fast-evolving and relying for their existence on a high number of mutations leading to it (high mutability and targetability), such as the most mutationally active residue, serine.

In addition, we have computed matrices at different evolutionary distances (and therefore with different degrees of contribution from the genetic code), which may be important for assessing mutations for diseases such as cancer, where somatic mutations are accumulated through a fast evolutionary process. In essence, using our MUTA matrices, computed on short evolutionary distances and highly constrained by the genetic code, one should be able to compute the likelihood of different amino acid residue substitutions occurring in diseases associated with alterations to the genome.

Given the influence that both the genetic code and the physico-chemical properties of amino acid residues have on mutability and targetability, it would perhaps be natural to explore whether some causal relationships exist between them. While several hypotheses for the evolution of the genetic code exist [12], perhaps the most accepted view is that the organization of the genetic code can be explained by a combination of the occupation of codon space by the new amino acid residues as soon as they appeared from their predecessors, and optimization, to some extent, driven by physico-chemical properties of amino acid residues. We therefore argue that the observed mutability and targetability of amino acid residues is, at a higher degree, a consequence rather than a cause of the organization of the genetic code.

Finally, by contrasting the general trends in conservation of the phosphorylatable residues in human (S, T and Y) to the conservation of phosphorylation sites (pS, pT and pY), we have uncovered a higher degree of conservation for phosphorylation sites on serine than expected. This would suggest that the cell can modulate the mutational properties of amino acid residues in special circumstances as, in the case of serine,

given their ancient functional importance, it prevents these residues from evolving as fast as they would under normal circumstances. In line with this perception, it has been reported recently that aspartic and glutamic acid tend to become phosphorylatable residues (serine and threonine) during evolution, in a mechanism that has been suggested as a transition from a static to a dynamic regulation of protein folding [13]. Because these two groups of residues are far from each other in mutational space (two mutations away), we can conclude that, similarly as we found for the phosphotyrosine signalling system in our previous work [8,14], this observation is likely to be driven by positive selection.

It will be important to unravel the plausible mechanisms that have led to different mutability and targetability rates (e.g. amino acid preference to residue in fast-evolving or disordered regions), whether different species with different genetic codes or codon preferences have different residues' mutational properties, and to what extent these properties determine the frequency of every amino acid residue. Moreover, it will also be important to implement tools that can use these metrics to assess the importance of mutations in cell signalling systems associated with cancer progression. We argue this will eventually lead to a better foundation for network-based medicine.

7. MATERIAL AND METHODS

(a) *Alignments and computation of ancestral sequences*

Sequences of known and inferred proteins of 11 vertebrate species, including *Homo sapiens*, with at least 6X genome coverage were retrieved from the Ensembl online database (release 55) at http://jul2009.archive.ensembl.org/info/data/ftp/. These 11 metazoan species are *H. sapiens* (human), *Pongo pygmaeus* (orangutan), *Cavia porcellus* (guinea pig), *Rattus norvegicus* (rat), *Mus musculus* (mouse), *Monodelphis domestica* (opossum), *Canis familiaris* (dog), *Bos taurus* (cow), *Ornithorhynchus anatinus* (platypus), *Gallus gallus* (chicken) and *Xenopus tropicalis* (frog). The InParanoid algorithm (v. 2.0) [15] was used to infer orthologous sequences of human proteins across the ten other vertebrate species using the retrieved proteomes. The BLOSUM80 scoring matrix is used with other default parameters in InParanoid. In all cases, only the longest translation of each known/inferred genes was fed into InParanoid for orthologue prediction. The sequence of each known human phosphoprotein was then grouped with its inferred orthologous protein sequences for multiple sequence alignment using the mafft algorithm (v. 6.240, E-INS-i option with default parameters) [16]. Ancestral sequences were inferred from each multiple sequence alignment using the Codeml program in PAML phylogenetic software suite [17]. The phylogenetic relationship depicted in figure 2*b* [18] was input to Codeml with CodonFreq = 2 and using WAG substitution matrix [19].

(b) *From coevolution matrices to mutability and targetability rates*

For each pair of ancestral-human sequences, we computed a 20×20 coevolution matrix describing the evolution tendency of each amino acid, with the ancestral amino acid in the row position and human-aligned residue in the column position. In order to avoid inaccuracies caused by alignment positions with lower quality, we filtered out alignment positions in or next to gaps (see the electronic supplementary material, figure S1 for more information on the fraction of residues excluded). We produced mutability and targetability rates by normalizing the coevolution matrices by row, i.e. effectively balancing out differences in amino acid residue frequencies. The mutability rate for each residue is then measured as the sum of all mutation frequencies, i.e. row sum minus conservation. On the other hand, the targetability rate is measured as the sum of mutation frequencies of all amino acid residues leading to a given amino acid residue, i.e. column sum minus conservation.

(c) *Phosphorylation site evolution*

We have traced the evolution of human phosphorylation sites on serine, threonine and tyrosine by measuring the fraction of each that is conserved versus the fraction that has appeared recently in evolution. More specifically, we compiled a list of human phosphorylation sites obtained from the PhosphoSitePlus [20] and phosphoELM databases [21] and computed what fraction of those are conserved in our inferred ancestral sequences and thus compared how the three signalling systems have evolved. In order to predict disorder propensity for all the proteins analysed, we ran Disopred v. 2.0 [22].

We thank the editor Tony Hunter for critical input on this manuscript. R.L. is a Lundbeck Foundation Fellow. R.L. is further supported by a Sapere Aude Starting Grant from The Danish Council for Independent Research and a Career Development Award from the Human Frontier Science Program.

REFERENCES

1 Jacob, F. 1977 Evolution and tinkering. *Science* **196**, 1161–1166. (doi:10.1126/science.860134)
2 Pastor-Satorras, R., Smith, E. & Solé, R. V. 2003 Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* **222**, 199–210. (doi:10.1016/S0022-5193(03)00028-6)
3 Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. & Postlethwait, J. 1999 Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545.
4 Dayhoff, M. O., Schwartz, R. & Orcutt, B. C. 1978 A model of evolutionary change in proteins. In *Atlas of protein sequence and structure*, vol. 5, suppl. 3 (ed. M. O. Dayhoff), pp. 345–352. Washington, DC: National Biomedical Research Foundation.
5 Henikoff, S. & Henikoff, J. G. 1992 Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA* **89**, 109 15–109 19. (doi:10.1073/pnas.89.22.10915)
6 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
7 Linding, R. 2010 (R)evolution of complex regulatory systems. *Sci Signal.* **3**, eg4. (doi:10.1126/scisignal.3127eg4)
8 Tan, C. S. H., Pasculescu, A., Lim, W. A., Pawson, T., Bader, G. D. & Linding, R. 2009 Positive selection of

tyrosine loss in metazoan evolution. *Science* **325**, 1686–1688. (doi:10.1126/science.1174301)
9 Lim, W. A. & Pawson, T. 2010 Phosphotyrosine signaling: evolving a new cellular communication system. *Cell* **142**, 661–667. (doi:10.1016/j.cell.2010.08.023)
10 Pincus, D., Letunic, I., Bork, P. & Lim, W. A. 2008 Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages. *Proc. Natl Acad. Sci. USA* **105**, 9680–9684. (doi:10.1073/pnas.0803161105)
11 Manning, G., Young, S. L., Miller, W. T. & Zhai, Y. 2008 The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc. Natl Acad. Sci. USA* **105**, 9674–9679. (doi:10.1073/pnas.0801314105)
12 Koonin, E. V. & Novozhilov, A. S. 2009 Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* **61**, 99–111. (doi:10.1002/iub.146)
13 Pearlman, S., Serber Jr, Z. & Ferrell, J. E. 2011 A mechanism for the evolution of phosphorylation sites. *Cell* **147**, 934–946. (doi:10.1016/j.cell.2011.08.052)
14 Tan, C. S. H., Schoof, E. M., Creixell, P., Pasculescu, A., Lim, W. A., Pawson, T., Bader, G. D. & Linding, R. 2011 Response to comment on positive selection of tyrosine loss in metazoan evolution. *Science* **332**, 917. (doi:10.1126/science.1188535)
15 Remm, M., Storm, C. E. & Sonnhammer, E. L. 2001 Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052. (doi:10.1006/jmbi.2000.5197)
16 Katoh, K. & Toh, H. 2008 Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinformatics* **9**, 286–298. (doi:10.1093/bib/bbn013)
17 Yang, Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
18 Hedges, S. B. 2002 The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**, 838–849. (doi:10.1038/nrg929)
19 Whelan, S. & Goldman, N. 2001 A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699. (doi:10.1093/oxfordjournals.molbev.a003851)
20 Hornbeck, P. V., Kornhauser, J. M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V. & Sullivan, M. 2012 PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **40**, D261–D270. (doi:10.1093/nar/gkr1122)
21 Dinkel, H., Chica, C., Via, A., Gould, C. M., Jensen, L. J., Gibson, T. J. & Diella, F. 2010 Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.* **39**, D261–D267. (doi:10.1093/nar/gkq1104)
22 Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F. & Jones, D. T. 2004 The DISOPRED server for the prediction of protein disorder. *Bioinformatics* **20**, 2138–2139. (doi:10.1093/bioinformatics/bth195)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2012) **367**, 3058
doi:10.1098/rstb.2012.0384

Correction

Phil. Trans. R. Soc. B **367**, 2584–2593 (19 September 2012) (doi:10.1098/rstb.2012.0076)

Research article: Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues

Pau Creixell, Erwin M. Schoof, Chris Soon Heng Tan and Rune Linding

Part (*a*) of figure 2 incorrectly highlighted some amino acid residues that are more than one mutation away from methionine. In line with this, part (*c*) erroneously portrayed phenylalanine as being close to methionine in mutational space, and in the main text (§2, THE INFLUENCE OF THE GENETIC CODE ON MUTATIONAL PATHS) it was incorrectly stated that phenylalanine is one of the closest residues to methionine. The corrected figure can be found below. This error does not affect any of our results or conclusions.