Medicinal Chemistry Research (2020) 29:1133–1146
https://doi.org/10.1007/s00044-020-02582-9

MEDICINAL
CHEMISTRY
RESEARCH

REVIEW ARTICLE

Rheostat positions: A new classification of protein positions relevant to pharmacogenomics

Aron W. Fenton[1] · Braelyn M. Page[1] · Arianna Spellman-Kruse[2] · Bruno Hagenbuch[3] · Liskin Swint-Kruse[1]

Received: 13 May 2020 / Accepted: 30 May 2020 / Published online: 7 June 2020

Abstract

To achieve the full potential of pharmacogenomics, one must accurately predict the functional outcomes that arise from amino acid substitutions in proteins. Classically, researchers have focused on understanding the consequences of individual substitutions. However, literature surveys have shown that most substitutions were created at evolutionarily conserved positions. Awareness of this bias leads to a shift in perspective, from considering the outcomes of individual substitutions to understanding the roles of individual protein positions. Conserved positions tend to act as "toggle" switches, with most substitutions abolishing function. However, nonconserved positions have been found equally capable of affecting protein function. Indeed, many nonconserved positions act like functional dimmer switches ("rheostat" positions): this is revealed when multiple substitutions are made at a single position. Each substitution has a different functional outcome; the set of substitutions spans a range of outcomes. Finally, some nonconserved positions appear neutral, capable of accommodating all amino acid types without modifying function. This paper reviews the currently-known properties of rheostat positions, with examples shown for pyruvate kinase, organic anion transporting polypeptide 1B1, the beta-lactamase inhibitory protein, and angiotensin-converting enzyme 2. Outcomes observed for rheostat positions have implications for the rational design of drug analogs and allosteric drugs. Furthermore, this new framework—comprising three types of protein positions—provides a new approach to interpreting disease and population-based databases of amino acid changes. In conclusion, although a full understanding of substitution outcomes at rheostat positions poses a challenge, utilization of this new frame of reference will further advance the application of pharmacogenomics.

Keywords Rheostat position · Saturating mutagenesis · Protein evolution · Specificity

Introduction

Many pharmacological agents exert their effects through binding to proteins. These binding events can (i) lead to direct inhibition of the protein's activity via competition with a natural, in vivo ligand or (ii) can have an agonist/antagonist effect that propagates through the protein to alter function at a distant site, as often occurs for receptors in signaling pathways. Both the binding by and signaling from pharmacological agents can be dramatically altered if the interacting protein has amino acid changes. This compels the field of pharmacogenomics. For example, population-wide polymorphisms that diminish efficacy can confound drug trials; even for drugs that pass clinical trials, substitutions that only occur in single individuals ("*n*-of-1") can cause toxicity that leads to adverse drug reactions.

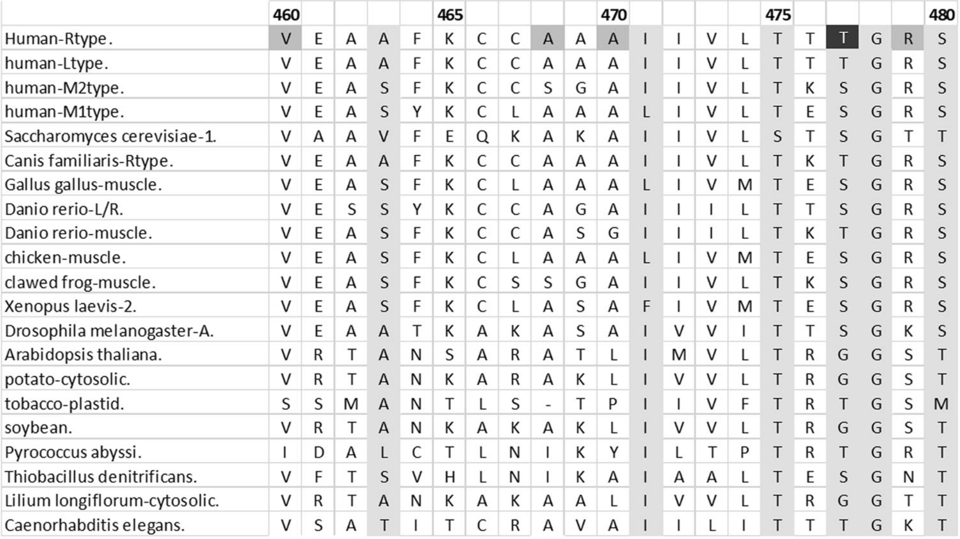
Unfortunately, genome sequencing has shown that it will be impossible to comprehensively illuminate the protein/drug relationship with laboratory experiments. Any two unrelated people can have >10,000 amino acid differences among their thousands of protein sequences (Ng et al. 2008; Lek et al. 2016); most children have a few de novo changes relative to their parents (Acuna-Hidalgo et al. 2016). Indeed, genome collections such as GNOMAD (Karczewski et al. 2020) have revealed that polymorphisms and *n*-of-1 substitutions (and everything in between) are common in many proteins. For most of these changes, effects on

✉ Liskin Swint-Kruse
lswint-kruse@kumc.edu

[1] Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, Kansas City, KS 66160, USA

[2] Department of Biochemistry, University of Nebraska—Lincoln, Lincoln, NE 68588, USA

[3] Department of Pharmacology, Toxicology, and Therapeutics, The University of Kansas Medical Center, Kansas City, KS 66160, USA



	460					465					470					475					480
Human-Rtype.	V	E	A	A	F	K	C	C	A	A	A	I	I	V	L	T	T	T	G	R	S
human-Ltype.	V	E	A	A	F	K	C	C	A	A	A	I	I	V	L	T	T	T	G	R	S
human-M2type.	V	E	A	S	F	K	C	C	S	G	A	I	I	V	L	T	K	S	G	R	S
human-M1type.	V	E	A	S	Y	K	C	L	A	A	A	L	I	V	L	T	E	S	G	R	S
Saccharomyces cerevisiae-1.	V	A	A	V	F	E	Q	K	A	K	A	I	I	V	L	S	T	S	G	T	T
Canis familiaris-Rtype.	V	E	A	A	F	K	C	C	A	A	A	I	I	V	L	T	K	T	G	R	S
Gallus gallus-muscle.	V	E	A	S	F	K	C	L	A	A	A	L	I	V	M	T	E	S	G	R	S
Danio rerio-L/R.	V	E	S	S	Y	K	C	C	A	G	A	I	I	I	L	T	T	S	G	R	S
Danio rerio-muscle.	V	E	A	S	F	K	C	C	A	S	G	I	I	I	L	T	K	T	G	R	S
chicken-muscle.	V	E	A	S	F	K	C	L	A	A	A	L	I	V	M	T	E	S	G	R	S
clawed frog-muscle.	V	E	A	S	F	K	C	S	S	G	A	I	I	V	L	T	K	S	G	R	S
Xenopus laevis-2.	V	E	A	S	F	K	C	L	A	S	A	F	I	V	M	T	E	S	G	R	S
Drosophila melanogaster-A.	V	E	A	A	T	K	A	K	A	S	A	I	V	V	I	T	T	S	G	K	S
Arabidopsis thaliana.	V	R	T	A	N	S	A	R	A	T	L	I	M	V	L	T	R	G	G	S	T
potato-cytosolic.	V	R	T	A	N	K	A	R	A	K	L	I	V	V	L	T	R	G	G	S	T
tobacco-plastid.	S	S	M	A	N	T	L	S	-	T	P	I	I	V	F	T	R	T	G	S	M
soybean.	V	R	T	A	N	K	A	K	A	K	L	I	V	V	L	T	R	G	G	S	T
Pyrococcus abyssi.	I	D	A	L	C	T	L	N	I	K	Y	I	L	T	P	T	R	T	G	R	T
Thiobacillus denitrificans.	V	F	T	S	V	H	L	N	I	K	A	I	A	A	L	T	E	S	G	N	T
Lilium longiflorum-cytosolic.	V	R	T	A	N	K	A	K	A	A	L	I	V	V	L	T	R	G	G	T	T
Caenorhabditis elegans.	V	S	A	T	I	T	C	R	A	V	A	I	I	L	I	T	T	T	G	K	T

Fig. 1 An example sequence alignment. Sequence alignments are represented with related protein (homologs) in horizontal rows, aligned so that equivalent positions fall into vertical columns. Amino acids are represented with the one letter code. When the same amino acid is present in many homologs, that position is considered to be conserved. Conservation is interpreted to indicated that other amino acids are not tolerated at that position, which in corollary indicates the importance of that particular side chain. This example shows part of the pyruvate kinase sequence alignment. The four human isozymes are the top rows and position numbering corresponds to RPYK; note that more sequences and more amino acid positions are in the alignment than shown. Light gray columns indicate positions with conserved ConSurf scores in the full alignment (Glaser et al. 2003; Landau et al. 2005; Ashkenazy et al. 2010). Black (conserved) and dark gray (not conserved) cells indicate positions of RPYK disease mutations (Pendergrass et al. 2006)

natural functions are not known, and the potential influences on new drug candidates is beyond our current ability to predict.

Thus, the identification of amino acid changes that are medically and/or pharmacologically relevant is very difficult. Some "obvious" positions are in protein binding sites, but amino acid changes far from binding sites are also commonly observed to have large functional consequences (e.g., (Wu et al. 2019; Swint-Kruse et al. 2003; Modi and Ozkan 2018)). Clearly, computational methods to predict outcomes of protein substitutions are vital. However, despite decades of study, available algorithms still lack reliable performance (Miller et al. 2017; Dong et al. 2015; Andreoletti et al. 2019; Zeng and Bromberg 2019).

To advance pharmacogenomics through computation, the necessary background knowledge originates from the field of protein chemistry. To probe which (and how) amino acid positions contribute to protein function, the primary tool of this field has been mutational studies. However, researchers in protein chemistry and biochemistry have faced the same problem that vexes pharmacogenomics: how to select the most relevant candidates for study. A "small" protein has 100 amino acid positions and most human proteins have hundreds of amino acids (Lipman et al. 2002), making the number of choices very large.

Thus, to narrow down the candidate positions to a tractable number, many researchers have turned to sequence alignments of protein homologs to guide experiments (e.g., Fig. 1). This tool readily identifies positions that are conserved throughout evolution, and thus are likely important. The identification of conserved sites has facilitated high-yield studies to determine the major functions of unknown proteins. A second use of sequence alignments has been to extrapolate the features known for one protein—such as an enzyme active site—to other homologs in the alignment.

The unintended consequence of this approach is that decades of protein mutagenesis studies have focused on conserved positions (Gray et al. 2012) and largely ignored the rest of protein sequence space. Nonetheless, positions that change during evolution (nonconserved) can also be important for function. Indeed, the outcomes of substitutions at nonconserved positions can be significant, as shown by the correlation of disease databases with sequence alignments (e.g., Fig. 1, dark gray (Pendergrass et al. 2006)). Due to the bias towards mutating conserved positions, much less is known about mutational outcomes at nonconserved positions. Thus, the computational algorithms developed to predict outcomes of mutations have not been trained or validated with datasets that include sufficient numbers of mutations at nonconserved positions.

Moreover, the bias toward mutating conserved positions has influenced common thought processes in the field of protein structure/function. Indeed, most biochemistry and biology textbooks introduce the amino acids by their physicochemical "similarities," and students are often taught that similar amino acids should have similar biochemical functions in a similar protein environment. For example, threonine and serine are presumed to be interchangeable, since they both have side chain hydroxyls. In contrast, a position that normally contains the hydrophobic amino acid leucine is not expected to tolerate the charged side chain of aspartate. The expectation that similar amino acids are tolerated for function and that dissimilar amino acids are catastrophic to function (or structure) appears to be supported by volumes of mutation/function studies of conserved positions. However, the relationship has seldom been tested at nonconserved positions.

Thus, our laboratories have been exploring the functional consequences of amino acid substitutions at nonconserved positions. The intent of this paper is to review the findings of our initial efforts.

Position classifications

In addition to bias toward the study of conserved positions, a second limitation of historical experimental studies is that the number of substitutions per position is usually restricted to a few amino acids, often just to alanine. However, additional information can be gleaned by considering the outcomes for multiple amino acids at each position. Doing so changes the researcher's perspective: instead of considering the role of a particular side chain at a given position ("residue"), one considers the overall role of the position (Zhang et al. 2011; Hodges et al. 2018).

From the perspective of a position and the assumptions derived from studies of conserved positions, researchers are led to expect a "toggle" outcome: Amino acids similar to the wild-type amino acid allow function, whereas all other destroy function (Fig. 2a). We find a useful analogy for this substitution behavior is a light switch toggle. In contrast, when choosing positions to substitute, researchers often rule-out nonconserved positions, assuming they make little contribution to function. In corollary, any substitution at a nonconserved position should not alter function; we classify such positions as "neutral" (Fig. 2b) (Martin et al. 2020).

However, in even our very earliest studies of nonconserved positions, substitution outcomes seldom matched either of these patterns. Many substitutions at nonconserved positions did indeed alter function, but the outcomes from alternative substitutions at individual positions varied over a wide, continuous range (e.g., Fig. 2c). Given the continuum,

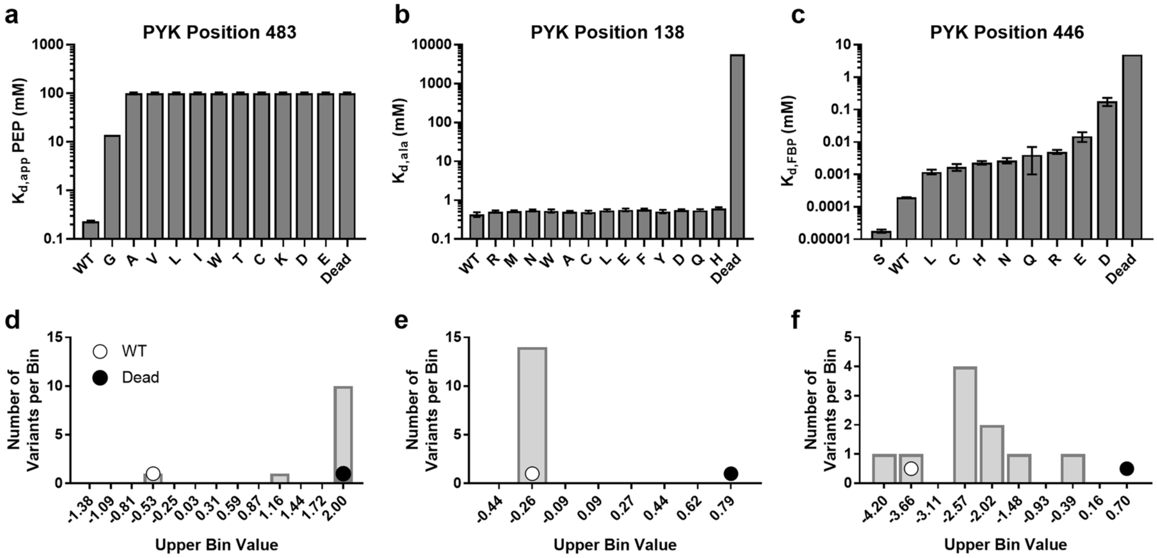


Fig. 2 Examples of toggle, neutral, and rheostat substitution outcomes for individual positions, using three of the functional parameters measured for human liver pyruvate kinase (LPYK). "$K_{d,app}$ PEP" is the apparent affinity for substrate phosphoenol pyruvate (corresponding to either (i) K_m if the Hill number is 1 or (ii) $K_{1/2}$ if the Hill number is < >1); "$K_{d,ala}$" and "$K_{d,FBP}$" are affinities for the two allosteric ligands. Amino acid substitutions are listed on the *x*-axis in panels (**a**–**c**). **a** At a toggle position, substitutions are either like wild-type or dead enzyme (i.e., no detectable function). **b** At a neutral position, most substitutions are like wild-type. **c** At a rheostat position, substitutions range from better than wild-type, to wild-type, to dead. To quantitatively summarize the overall substitution outcomes, data were binned for further analyses (Hodges et al. 2018). Histograms for each type of position in (**a**–**c**) are represented in (**d**–**f**). A white dot is used to indicate the bin that includes wild-type data and a black dot is used to indicate the bin corresponding to no detectable function. Note especially the binning pattern of a rheostat position (**f**), which has entries in multiple bins. Data are from (Wu et al. 2019; Hodges et al. 2018; Tang et al. 2017; Ishwar et al. 2015; Martin et al. 2020). The RheoScale scores determined from these histograms are as follows. **d** Position 483: neutral 0.00, rheostat 0.16, and toggle 0.91. The toggle score is above the significance threshold of 0.7. **e** Position 138: neutral 1.00, rheostat 0.03, and toggle 0.00. The neutral score is above the significance threshold of 0.7. **f** Position 446: neutral 0.00, rheostat 0.58, and toggle 0.00. The rheostat score is above the significance threshold of 0.5

we have termed positions with this type of outcome as “rheostat” positions, in analogy to a dimmer switch. It is interesting to note that, in several studies of rheostat positions, we frequently identify substitutions for which function is “better” than wild-type.

A second characteristic that we noted in early studies of rheostat positions was, when the functional outcomes were rank-ordered (e.g., Fig. 2b), the extent of functional change did not correlate with similar amino acid chemistries (Meinhardt et al. 2013). For example, phenylalanine and serine substitutions could result in similar functions, but their outcomes could differ from those of either tyrosine or threonine. This is not the first time that such discrepancies have been noted for the amino acid similarity rules (Pal et al. 2006; Jonson and Petersen 2001; Gilbert et al. 2012). Another correlation that failed for several rheostat positions is that the substitution rank order did not correlate with evolutionary frequency (Meinhardt et al. 2013). Again, this lack of trend had been noted in other studies (Pal et al. 2006; Hietpas et al. 2011). Thus, the textbook assumptions that are based on mutations of conserved positions fail to explain a large number of substitution outcomes at non-conserved positions.

Rheostatic outcomes affect various aspects of protein function

Using our own and other published datasets, we have now identified rheostat positions in a wide range of proteins, including (but not limited to) prokaryotic transcription factors (Meinhardt et al. 2013), human liver pyruvate kinase (LPYK; Fig. 2; (Wu et al. 2019; Hodges et al. 2018)), a drug uptake transport protein (OATP1B1; Fig. 3; (Ohnishi et al. 2014)), the Angiotensin-converting enzyme 2 (ACE2) to which the SARS-Cov-2 spike protein binds (Fig. 4; (Procko 2020)), and the β-lactamase inhibitory protein (“BLIP”; Fig. 5; (Adamski and Palzkill 2017)).

In addition, many of the “deep mutational scanning” data show evidence of rheostat positions (e.g., (Hodges et al. 2018; Roscoe et al. 2013)). In the experimental design of this technique (Fowler and Fields 2014; Roscoe et al. 2013), a region of the gene encoding a protein is subjected to saturating mutagenesis to create a library of variants. This library is then transformed (or transfected) into cells in culture, followed by biological competition in conditions that require the protein function of interest. After some number of generations, the library is re-sequenced with next generation sequencing to infer the frequency of each variant. In turn, clone frequency is used to infer the level of protein function for each variant, using the assumption that the two correlate. Some variations on this method use other readouts, such as cell sorting of fluorescent tags, prior to sequencing.

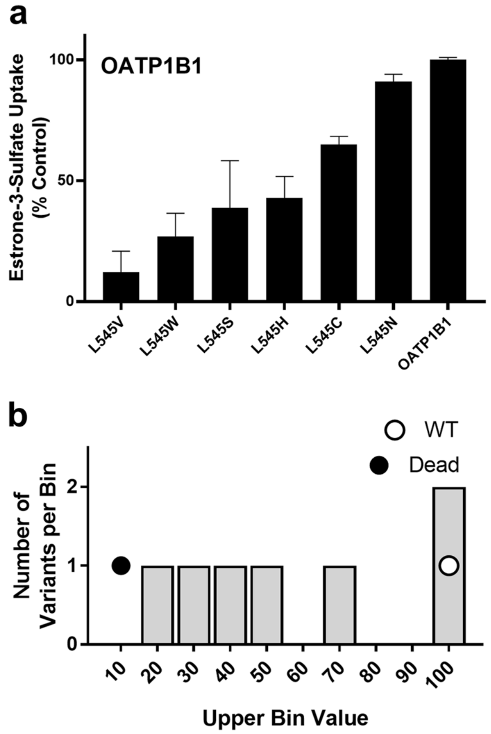


Fig. 3 Rheostat behavior in the drug uptake transport protein, OATP1B1. **a** Relative to wild-type OATP1B1, substitutions at position L545 show a range of diminished function depending on the amino acid substitution. Amino acid substitutions are listed on the *x*-axis. The measure of function, esterone-3-sulfate uptake, is on the *y*-axis. **b** Histogram of OATP1B1 functional data. The upper bin limit is shown along the *x*-axis. The number of amino acid substitutions that occupy each bin is shown along the *y*-axis. A white and black dot are used to denote the bin that contains the wild-type and dead values, respectively. Data are from Ohnishi et al. (2014). The RheoScale scores determined from these histograms are neutral 0.00, rheostat 0.63, and toggle 0.00. The rheostat score falls above the significance threshold of 0.5

The interpretation of deep mutational scanning results raises an interesting question: which functional parameter(s) are affected by substitutions at rheostat positions? Many deep mutational scanning studies detect alterations in phenotypes that comprise multiple aspects of the protein function (such as binding, catalysis, and/or allosteric coupling) as well as the amount of stable protein present in the cell. Thus, unless corresponding biochemical experiments are available, the origins of rheostatic substitution outcomes in deep mutational scanning data are not known.

In biochemical studies, we have detected rheostatic outcomes on K_d for binding (e.g., Fig. 2; (Wu et al. 2019; Hodges et al. 2018; Zhan et al. 2006; Meinhardt et al. 2013)), the magnitude of allosteric coupling (“*Q*”, (Wu et al. 2019; Hodges et al. 2018)), and substrate transport (Fig. 3a; (Ohnishi et al. 2014)). We fully expect that enzyme catalytic rates can also be modulated by substitutions at rheostat positions (Swint-Kruse 2016). Alternatively, some

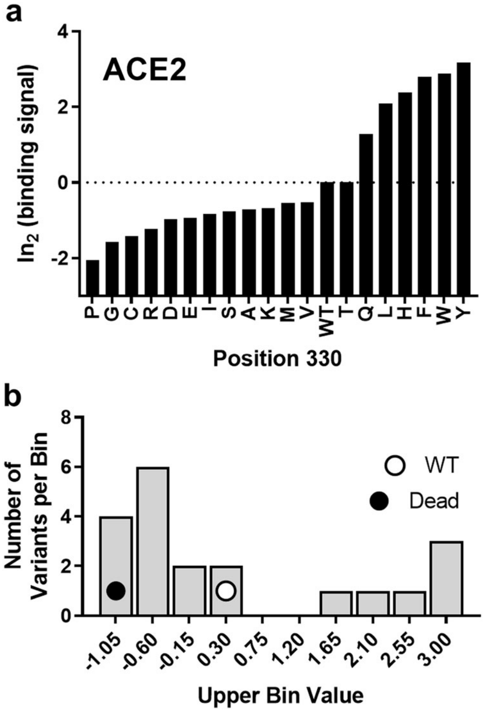


Fig. 4 Example rheostat position from human ACE2; data were taken from Procko (2020). **a** Functional data for position 330 of ACE2. Amino acid substitutions are listed on the *x*-axis. A measure of the binding of ACE2 and the SARS-CoV-2 spike protein is shown on the *y*-axis. Negative values indicate that binding is weaker than wild-type, whereas positive values indicate "better" function. **b** Histogram analyses of the functional data for ACE2 position 330. The upper bin value for each bin is shown along the *x*-axis. Bins that contain the "dead" and "wild-type" values are shown with a black and white dot, respectively. All data with scores above 3.00 were reset to this limit, following the example of Procko (2020). The nonfunctional "dead" protein value of −1.5 was calculated from (average + standard deviation) derived from all nonsense mutations (all replicates) plus the standard deviation. The Rheoscale scores calculated from this histogram were: neutral 0.05, rheostat 0.78, and toggle 0.21. The rheostat score is well above the significance threshold of 0.5

rheostat positions may impact protein stability: intermediate effects of substitutions, a hallmark of rheostat positions, on protein stability have long been known (e.g., (Karp et al. 2010; Klein et al. 2019; Chiti et al. 1999; Hargrove et al. 1994; Matsuura et al. 2018)). We also identified rheostat positions in results from deep mutational scanning experiments designed to detect altered stability in TIM barrels (Chan et al. 2017; Hodges et al. 2018). While we have to-date focused our own experimental studies on substitutions that alter protein function, it will be interesting to dissect the effects of rheostat position substitutions on function and stability in future studies, as discussed previously (Swint-Kruse 2016).

Notably, among our studies, we have identified a subset of rheostat positions that can simultaneously modulate two or more functional parameters (Wu et al. 2019; Zhan et al. 2006). LPYK has a catalytic site for substrate and two allosteric sites that bind different ligands (Fig. 6a). In high-throughput enzymatic studies, we simultaneously monitored five functional parameters—the apparent binding affinity for substrate ("$K_{d,app}$ PEP"; see the legend to Fig. 2), the binding affinity for each of the two allosteric ligands ("$K_{d,ala}$" and "$K_{d,FBP}$"), and allosteric coupling between binding substrate and each of the allosteric ligands ("Q_{ala}" and "Q_{FBP}"). We identified several rheostat positions in the allosteric binding sites for which substitutions altered two or three of these parameters. Furthermore, the modulations did not correlate with each other. That is, the amino acid rank order for allosteric ligand binding did not correlate with the amino acid rank order for allosteric coupling (Fig. 6b–d; (Wu et al. 2019)).

Similar "multi-rheostat" positions, with uncorrelated parameter changes, have been found by examining the whole-protein substitution study of the lactose repressor protein ("LacI"; (Suckow et al. 1996; Markiewicz et al. 1994). Both

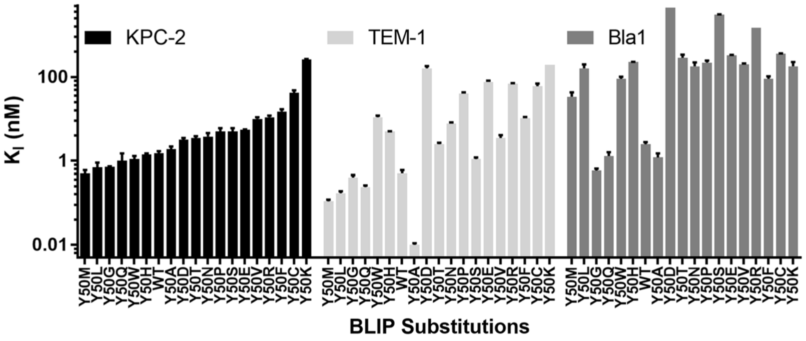


Fig. 5 BLIP position 50 is a rheostat position. In this study, BLIP position 50 was substituted with 18 amino acids (the 19th could not be purified) and binding was assessed for the three different beta-lactamases noted on the plot. The amino acid rank order for the left-most data is preserved in the middle and right datasets; the jagged patterns show that rank order changes for these two binding partners. Note that even the tightest and weakest substitutions differ among the datasets. Data were taken from Adamski and Palzkill (2017)

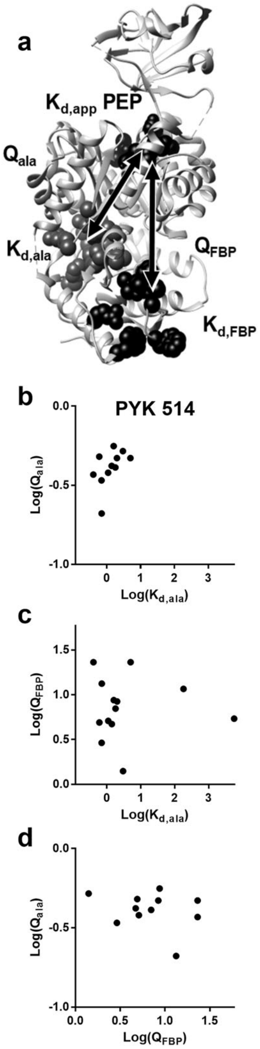


Fig. 6 a Protein monomer extracted from a structure of the LPYK homo-tetramer. This figure was rendered with UCSF chimera (Pettersen et al. 2004) using the pdb 4ip7 (Holyoak et al. 2013). Black spheres at the top of the structure highlight positions in the catalytic site; dark gray spheres indicate the allosteric site for alanine inhibitor binding; black spheres at the bottom of the structure indicate the allosteric site for fructose-1-6-bisphosphate activator binding. Binding affinities for the allosteric effectors are denoted with "$K_{\mathrm{d,ala}}$" and "$K_{\mathrm{d,FBP}}$", respectively. Arrows indicate the allosteric coupling ("Q") that occurs between the catalytic site and each of the allosteric sites. Both allosteric effectors alter the apparent affinity for substrate phosphoenol pyruvate binding ("$K_{\mathrm{d,app}}$ PEP"; see the legend to Fig. 2). **b**–**d** LPYK functional parameters altered by substitutions at rheostat positions do not correlate. For the example rheostat position 514, the values of three functional parameters with rheostat scores ≥0.5 were compared for each amino acid substitution (individual dots). (Note that two substitutions have estimates for $K_{\mathrm{d,ala}}$, but binding was too weak to estimate Q_{ala}; this leads to a different number of data points on (**b**) and (**c**)). No correlation was observed among the parameters for this or other LPYK rheostat positions, which indicates that each substitution has independent effects on the different functional parameters. Data were taken from Wu et al. (2019)

LPYK and LacI show coincidence between the locations of their multi-rheostat positions and regions of the protein involved in allosteric coupling, suggesting that multi-rheostat positions are enriched in and near allosteric sites. Given the uncorrelation and (as of yet) unpredictable outcomes on multiple functional parameters, these types of rheostat positions could be one thing that confounds predictions about substitutions of rheostat positions. For medicinal chemistry, this has implications for the rational design of allosteric drugs (e.g., (Guarnera and Berezovsky 2020; Daura 2019)).

Another function of particular interest to medicinal chemists is that of specificity. Specificity—which ligand is most preferred—is defined by the rank order of binding affinities for alternative ligands. Changes in specificity due to amino acid substitutions are defined by differences in the fold-change observed for each ligand, and/or a change in the rank order of preferred ligand (Creighton 1993; Tungtur et al. 2019); enzymologists often monitor changes in the ratio of $V_{\mathrm{max}}/K_{\mathrm{m}}$. Since binding affinity is at the core of specificity, we would expect and have observed specificity changes arising from amino acid substitutions at rheostat positions (Zhan et al. 2008; Tungtur et al. 2019).

An important, related observation is that the amino acid rank order for a rheostat position can also be substrate specific. For example, the Palzkill lab studied BLIP binding to three beta-lactamases. BLIP position 50 is clearly a rheostat position, as seen from the series of substitutions generated at this position. However, the rank order of the variant proteins was dependent on which of the three beta-lactamases was used as the binding partner (Fig. 5; (Adamski and Palzkill 2017)). This change in rank order provides another means to assess altered specificity. Such observations have significant implications for pharmacogenomics when protein targets interact with more than one drug or when designing drug analogs.

Quantifying the toggle, rheostat, and neutral character of a position

As we continued to identify rheostat positions, it quickly became apparent that not all positions neatly fell into the three classifications of toggle, neutral, and rheostat. For example, in LPYK, substitution effects on PEP affinities were seldom easily classified (e.g., Fig. 7a, b). Some positions showed substitution patterns that were close—but not strictly—neutral (e.g., Fig. 7a). Others showed substitution patterns that were in between rheostat and toggle (e.g., Fig. 7b). Thus, we have realized that each position's substitution pattern also falls on a continuum that is bounded by the idealized neutral, rheostat, and toggle outcomes.

To better interpret the "in between" patterns, we developed a histogram analysis to determine (i) the fraction of wild-type-like substitutions, which is reported as a "neutral score", (ii) the fraction of substitutions that abolish function, which is reported as a "toggle score", and (iii) the number

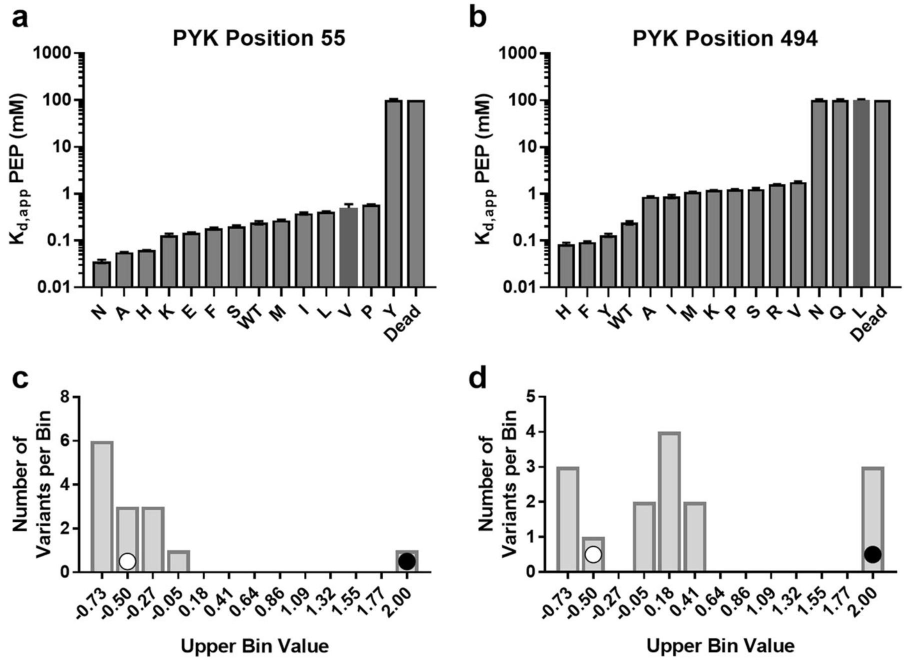


Fig. 7 Example of a near-neutral position and a part-toggle-part-rheostat position in LPYK. Amino acid substitutions are listed on the *x*-axis in panels **a**, **b**. **a** Position 55 shows a substitution pattern that is between neutral and rheostatic. Most of the substitutions have activities close to that of wild-type, but some substitutions result in "dead" protein variants. **b** Position 494 shows a substitution pattern that is partially that of a rheostat position and partially that of a toggle position. Some of the amino acid substitutions improve the function to different degrees whereas other substitutions result in wild-type-like or "dead" values. **c** Binned data further demonstrates the near-neutral data for position 55, with the wild-type like data shown in the bin with the white dot. Substitutions that result in no detectable function are denoted by the black dot as a "dead" bin. The RheoScale scores determined from this histogram are as follows: neutral 0.08, rheostat 0.28, and toggle 0.08. None of these scores fall above the relevant significance thresholds; this position is best described as "weakly rheostatic". **d** Binned data further demonstrates the range of behaviors exhibited by substitutions at position 494. The RheoScale scores determined from this histogram are as follows: neutral 0.00, rheostat 0.41, and toggle 0.21. None of these scores fall above the relevant significance thresholds; this position is best described as "weakly rheostatic". Substitutions show both (i) toggle behavior, acting closely to the wild-type behavior (white circle) or showing no detectable function (black circle), and (ii) rheostat behavior, falling into a range of bins between wild-type and dead. Data were taken from (Tang et al. 2017; Ishwar et al. 2015; Wu et al. 2019; Hodges et al. 2018)

and range of different outcomes sampled by all substitutions, which is used to calculate a "rheostat score" (Wu et al. 2019; Hodges et al. 2018; Martin et al. 2020). Example histograms are shown in Figs. 2d–f, 5b, 7c, d. At an ideal rheostat position, all bins should be occupied by at least one substitution, allowing access to the full range of functional outcomes and generating a score of 1.0. For lower scores (reflecting the nonideal outcomes described above), empirical thresholds have been devised to determine which positions have dominant neutral (Martin et al. 2020), rheostat (Wu et al. 2019; Hodges et al. 2018), and toggle substitution behaviors (Wu et al. 2019). Simulations with these thresholds have shown that 10–12 substitutions per position are sufficient to classify the rheostat, toggle, or neutral character of many protein positions (Hodges et al. 2018); although it would be optimal to have all 19 substitutions, this is often cost and time prohibitive.

The application of the neutral score warrants additional discussion here. On a practical level, one must have a good estimation for the error associated with the scores, and particularly for the wild-type score, in order to properly ascertain which substitutions are neutral (Martin et al. 2020). If all (or most) substitutions for a given position are similar to wild-type for a given functional parameter, the position is described as neutral for that parameter. However, if a protein has multiple functional parameters (such as LPYK, above), a position must be neutral in all parameters in order to be neutral for the overall protein function. As extensively discussed in Martin et al. (2020), it is impossible to assess all possible functional parameters, particularly when unknown protein–protein interactions might occur in vivo. Nevertheless, the identification of biochemically neutral and near-neutral positions provides a critical comparison set for understanding the contributions of rheostat positions.

Because they aggregate data from multiple substitutions at one position, the neutral, rheostat, and toggle scores are also useful for mapping functional outcomes onto structures and for comparing substitution outcomes with results from sequence analyses. Deep mutational scanning studies have also grappled with how to aggregate data from multiple substitutions. In such studies, one approach has been to calculate a "conservation score" for each position by averaging the functional outcomes of all its substitutions. However, as shown in the ACE2 example of Procko (2020) (Fig. 8), conservation scores likely correlate with toggle scores. In contrast, rheostat scores showed low correlation with conservation scores. Indeed, some positions with strong rheostat scores had modest/low conservation scores and thus might be overlooked even though several of their substitutions had large effects on function. Rheostat score calculations can quickly identify rheostat positions in the large datasets generated by deep mutational scanning experiments. In turn, this will (i) identify a new group of functionally critical positions in the proteins of interest that might otherwise be hidden within these large datasets and (ii) provide the new examples required for studies that determine new substitution rules for rheostat positions.

Rheostat positions might be identified by comparing disease and genome/exome databases

We have also considered whether the data collected to build various databases might be useful for discriminating rheostat, toggle, and neutral positions. As an example, non-spherocytic hemolytic anemia is caused by a range of point mutations in the pyruvate kinase isozyme that is expressed in erythrocytes (RPYK). A database of 215 disease-causing mutations has been assembled by several groups (Pendergrass et al. 2006; Secrest et al. 2020; Canu et al. 2016). We assume that these mutations either greatly reduced or abolished enzyme function. In contrast, recent efforts to curate natural protein variants in the human population via genome/exome sequencing may include amino acid changes that are *not* connected to disease. The GNOMAD database (Karczewski et al. 2020) reports an additional 270 substitutions in RPYK. Since the latter are absent from the disease database, many of these GNOMAD substitutions are expected to have little effect on phenotype.

Any attempt to use patient and population databases to understand the outcomes of amino acid substitutions requires first considering the distinct definitions of biological change and biochemical change. The detectable limit of biochemical change is defined by the detectable limit of change for each biochemical assay. Biological change is defined by a biochemical change that is large enough to exert a phenotype. However, the thresholds for biological change are condition dependent (e.g., (Soskine and Tawfik 2010)) because altered conditions can arise from other changes in the genome or various environmental exposures. For example, many glucose-6-phosphate dehydrogenase substitutions are perfectly benign unless the individual takes a certain class of drugs or eats fava beans (Luzzatto and Arese 2018; Bubp et al. 2015). Thus, biochemical experiments remain of critical importance for interpreting substitution outcomes observed in biological databases.

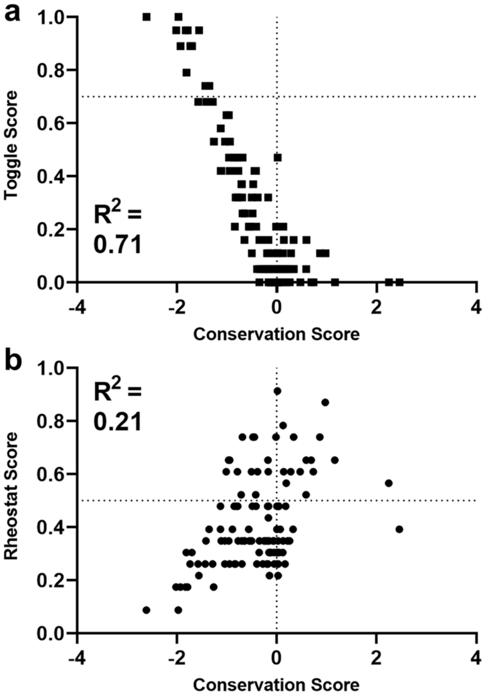


Fig. 8 Toggle and conservation scores agree whereas rheostat scores yield a different view of substitution data. Experimental data were obtained from Procko (2020) and comprised saturating mutagenesis for 117 positions in and near the binding site for the spike SARS-CoV-2 protein (for a total of 2223 variants plus wild-type). Conservation scores are shown along the *x*-axis. According to Procko, "Conservation scores are calculated from the average of the log2 enrichment ratios for all amino acid substitutions at each residue position." Positive conservation scores show that most substitutions at a position lead to enriched binding of the SARS-CoV-2 spike protein, whereas negative values indicate that most substitutions at a position diminished binding. **a** Comparison between toggle scores and conservation scores. Toggle scores were calculated from the substitution data for each of the 117 positions, using the RheoScale calculator and the "nCov-S-High sorts" log2 enrichment ratios of replicate 1. Further details to the calculation are in the legend to Fig. 4. The dashed line at $y = 0.7$ indicates the empirical significance threshold for toggle scores that was previously determined (Wu et al. 2019). **b** Comparison between rheostat scores and conservation scores. Rheostat scores were calculated simultaneously with toggle scores. The dashed line at $y = 0.5$ indicates the empirical significance threshold for rheostat scores that was previously determined (Hodges et al. 2018). As seen by the low Pearson correlation coefficient, rheostat and conservation scores reveal different aspects of the aggregate substitution data for each position. This can be understood by considering two hypothetical examples, one with the set of functional values [−2, −1, 0, 1, 2] and one with the set of functional values [−1, −1, 0, 1, 1]. Both positions have a conservation score of 0, but the first position has a stronger rheostat score than the second position. In the ACE2 data, positions with common conservation scores have a wide variety of rheostat scores

In the RPYK example, several relevant biological conditions could impact interpretation of patient and population databases. First, some substitutions may be so catastrophic that they cannot support life and thus are never detected in a living person; these would be absent from both databases. Second, some intermediate changes in RPYK biochemical function may not be enough to cause disease and would fall into the population database rather than the patient database, although in a biochemical study they would be classified as having a significant change. Third, RPYK is a tetramer. When the protein is expressed from two wild-type alleles, RPYK forms a homo-tetramer. However, if one allele contains a mutation, the protein can assemble into a hetero-tetramer which, in turn, can enhance or diminish the effect of the changed amino acid. Finally, genetic differences outside of the *pklr* gene that codes for RPYK could impact the phenotypes of individuals.

With these definitions and caveats in mind, we compared the patient and population database information for RPYK. We identified twelve positions (Table 1) with at least one substitution in the disease-causing database and two substitutions in the GNOMAD dataset (excluding any listed in the disease database). By definition, rheostat positions should have substitutions that fall into both datasets; thus, these positions could be enriched for rheostat positions. Positions that contain no disease-causing mutations, yet have multiple substitutions in the GNOMAD database, are candidates for near-neutral and neutral positions. In RPYK, 42 positions were identified to have at least two mutations in the GNOMAD database and no known disease-causing substitutions (Table 2). Finally, positions for which only disease-causing mutations are known may very well be candidate toggle (or near-toggle) positions; 19 such positions were identified by this criterion in RPYK (Table 3).

Table 1 Pyruvate kinase positions with substitutions in both disease and GONMAD databases

LPYK position number[a]	Disease mutation[b]	GNOMAD mutation[c]	Ala scan max fold change[d]
L42	P	F, H	3.12
P51	H	A, L	2.73
R55	P	G, H	100[e]
I188	T	F, V	1.47
R306	P, Q, W	G, T	100
D308	N, Q	G, H, Y	100
D359	N	C, G	100
G375	R	A, E	N/A[f]
V429	M	A, L	1.91
V459	W	G, L, Q	7.64
R487	S	C, H	2
M537	V	I, T	100

[a]Positions numbers in LPYK scheme are reported here for consistencies among our studies; original reports and GNOMAD used RPYK numbers

[b]Any overlap between the two datasets has been removed from the GNOMAD column

[c]Positions that include only a single variant in GNOMAD are not included; at least two variants were required for inclusion

[d]Data taken from (Tang and Fenton 2017)

[e]A fold-change of 100 was assigned for a catastrophic loss of any one of the five functional parameters, including substitutions that lost all catalytic activity

[f]Not available. Wild-type LPYK was either alanine or glycine at this position

Next, we compared the patient and population RPYK databases to a biochemical database comprising a whole-protein, alanine scan of LPYK (Tang and Fenton 2017). LPYK is a second gene product from the *pklr* gene, and LPYK and RPYK only differ by the loss of 31 amino acids on the N-terminus of LPYK. In the alanine-scanning study, the five LPYK functional parameters (Fig. 6) were monitored for the alanine substitution at almost every position that was not alanine or glycine in wild-type LPYK. The fold-change in each of the five LPYK functions, relative to wild-type, was determined; the maximal observed fold-change was compared with the sets of positions extracted from the disease and GNOMAD databases. (For ~10% of the alanine substituted positions, all enzymatic activity was lost and parameters could not be measured; these "dead" variants are reported as having "100" fold change.) Our reasoning was the maximal fold-change was likely to cause the largest biological effect, although this assumption should be more deeply considered in future studies. (For example, a fourfold effect in substrate binding might have a smaller biological outcome than a fourfold effect in allosteric regulation.) Next, we made the following approximations:

(i) A rheostat position, by definition, has the possibility of presenting a range of functional outcomes. Thus, we would expect the alanine scan data in Table 1 (Disease + GNOMAD) to span a wide range of values.

(ii) In contrast, because substitutions at neutral positions are defined by having functions similar to wild-type, the alanine substitution data in Table 2 (GNOMAD only) should show small values of maximal fold change if these positions are enriched for neutral positions. The application of this assumption has several caveats: first, one must define a threshold for "small change". For initial estimations herein, we defined "small" as less than fourfold. Second, some positions in Table 2 may be weakly rheostatic (substitutions led to biochemical changes that were not large enough to cause biological disease), and their alanine values could be intermediate. Third,

Table 2 Pyruvate kinase positions with GNOMAD-only substitutions

LPYK position[a]	Disease mutation	GNOMAD mutation[b]	Ala scan max fold change[c]	Match expectation[d]
M34		T, V	10.00	
R68		C, H	4.73	
R72		A, C, G, S	3.00	Yes
E94		A, Q	2.00	Yes
N102		I, K	1.52	Yes
S116		C, T	1.15	Yes
G138		E, R, W	N/A[e]	
V144		A, M	1.95	Yes
V156		A, M	3.45	Yes
N167		K, N	2.45	Yes
N175		Q, S	2.73	Yes
V180		L, M	4.36	
P181		L, S	2.34	Yes
G212		S, V	N/A	
G213		D, S	N/A	
R218		Q, W	2.18	Yes
A226		S, T, V	N/A	
D229		E, H, N	8.18	
L230		F, W	12.73	
R242		C, H	1.57	Yes
V245		M, W	2.45	Yes
V266		I, L	1.29	Yes
P272		A, L	2.18	Yes
G276		R, S, V	N/A	
K282		Q, R	100[f]	
R291		M, S	2.45	Yes
D293		E, G	2.11	Yes
L296		P, Q, V	9.09	
E312		K, Q	9.09	
R351		Q, W	5.27	
P352		A, S	100	
I371		T, V	100	
N381		K, T	1.39	Yes
A394		D, S, T, V	N/A	
R404		G, P, Q, W	2.09	Yes
R411		C, H	7.27	
R412		Q, W	3.09	Yes
A414		E, V	N/A	
V462		D, I	2.53	Yes
P489		L, T	1.10	Yes
R516		C, H	2.09	Yes
G532		C, S	N/A	
				% Matched: 52

[a] Positions numbers in LPYK scheme are reported here for consistencies among our studies; original reports and GNOMAD used RPYK numbers

[b] Positions that include only a single variant in GNOMAD are not included; at least two variants were required for inclusion

[c] Data taken from (Tang and Fenton 2017)

[d] For the alanine substitution, maximum fold-change for any of five biochemical parameters was less than four

[e] Not available. Wild-type LPYK was either alanine or glycine at this position

[f] A fold-change of 100 was assigned for a catastrophic loss of any one of the five functional parameters, including substitutions that lost all catalytic activity

Table 3 Pyruvate kinase positions for which all substitutions are in the disease dataset

LPYK position[a]	Disease mutations	GNOMAD mutations[b]	Ala scan max fold change[c]	Match expectation[d]
S99	P, T, Y		4.73	
R104	D, W		2.55	Yes
A106	T, V		N/A[e]	
R132	C, L		100[f]	Yes
L241	P, V		100	Yes
F256	L, V		100	Yes
D262	N, V		100	Yes
A264	I, V		N/A	
V289	L, M		10.82	Yes
G310	A, D		N/A	
G327	E, R		N/A	
R354	G, W		8.46	
N362	D, K, S		100	Yes
A363	D, S, V		N/A	
E376	G, K		100	Yes
K379	D, E		2.93	Yes
E396	A, D, N		100	Yes
A464	T, V		N/A	
G480	E, R		N/A	
				% Matched: 53

[a] Positions numbers in LPYK scheme are reported here for consistencies among our studies; original reports and GNOMAD used RPYK numbers

[b] No additional GNOMAD variants were identified for these positions

[c] Data taken from (Tang and Fenton 2017)

[d] For the alanine substitution, maximum fold-change for any of five biochemical parameters was either (i) greater than ten-fold (very deleterious) or (ii) less than fourfold (similar to wild-type)

[e] Not available. Wild-type LPYK was either alanine or glycine at this position

[f] A fold-change of 100 was assigned for a catastrophic loss of some functionality, including substitutions that lost all catalytic activity

disease mutations may not yet have been observed for some of these positions.

(iii) Finally, if the "disease only" positions (Table 3) are enriched for toggle positions, then the alanine fold-change should either be very large or similar to wild-type (if it is one of the few tolerated amino acids). Again, we recognize caveats: In this case, we need to define "very large". As a first approximation, we used ≥10-fold. Second, the loss of function required to cause disease could be smaller than the "dead" function used to define a biochemical toggle position and thus alanine substitutions are not obligated to be biochemically catastrophic.

Despite heavy reliance on many caveats, the comparison of alanine scan results for the three sets of positions is intriguing. As predicted, the alanine results in Table 1 do span a range. In Table 2, 52% of positions again matched expectation for a dataset enriched in neutral positions. Furthermore, in Table 3 53% of positions matched expectations for a dataset enriched in toggle positions. The latter two successes are greater than the 33% chance that would be expected for randomly making the correct assignment if one assumes that "perfect" rheostat/neutral/toggle behavior can be assigned to each position. Unfortunately, Tables 1 and 3 contain too few positions to determine whether the differences among Tables 1, 2, and 3 are statistically significant. Since RPYK has one of the largest sets of database and experimental data available, this illustrates another challenge of using databases to robustly identify the locations of rheostat, neutral, and toggle positions.

Nevertheless, these results support the hypothesis that databases can be used to generate sets of positions enriched for rheostat, neutral, and toggle positions for future study. Furthermore, in some cases, there may already be evidence of intermediate functional effects arising from known polymorphisms. For example, the *SLCO1B1* gene encoding the OATP1B1 drug transporter is highly polymorphic. The most frequent, single nucleotide polymorphisms are c.388A>G (p.N130D, 48% frequency according to GNOMAD), c.521T>C (p.V174A, 13%), c.463C>A (p.P155T, 11%), c.1929A>C (p.L643F, 4.6%), and c.733A>G (p.I245V, 0.64%). Some of these SNPs are known to alter OATP1B1 protein expression and/or function.

For example, OATP1B1 N130D has reduced, increased or unchanged function, depending on the transported substrate (Niemi et al. 2011). This altered specificity would be consistent with the behavior of a rheostat position similar to that at BLIP position 50 (Fig. 5). The protein product of another polymorphism, OATP1B1 P155T appears to have unchanged function, whereas OATP1B1 V174A has decreased membrane expression (Niemi et al. 2011). Pharmacologically, OATP1B1 V174A has been linked to statin-induced adverse effects; almost 20% of patients with the CC genotype developed myopathy at high doses of simvastatin within the first 5 years of treatment (Link et al. 2008). If the locations of these polymorphisms could be identified as either rheostat, toggle, or neutral positions, it would help understand whether additional substitutions at these locations would lead to adverse drug reactions: outcome predictions for substitutions at rheostat positions are currently extremely unreliable, but substitution predictions for toggle positions are much more reliable (Miller et al. 2017) and substitutions at neutral positions can be disregarded.

Do toggle, rheostat, and neutral positions correlate with evolutionary conservation patterns?

Based on our original study (Meinhardt et al. 2013), it is tempting to associate rheostat positions with nonconserved positions (those that change during evolution) and toggle positions with conserved positions (those that do not change during evolution). However, these may not be obligate associations. Furthermore, since nonconserved positions are also associated with neutral positions, an important question that follows for each type of position is "How nonconserved is nonconserved?"

In fact, nonconservation does not necessarily mean that the position experiences random change during evolution. First, there are established methods to calculate the degree of conservation (Shannon 1948). In addition, nonconserved positions[1] in a protein sequence can be classified by other criteria. Two ways of grouping positions are: (i) the extent to which one position co-evolves with another position (e.g., (Parente and Swint-Kruse 2013) and references therein) or with multiple positions (e.g., (Lee et al. 2012; Parente et al. 2015)); and (ii) the extent to which substitutions follow the division of subfamilies within a phylogenetic tree ("phylogenetic" positions) (e.g., (Ye et al. 2008; Mihalek et al. 2006; Gu et al. 2013)). Like the degree of conservation, these two classifications of positions are evaluated using sequence alignments (Fig. 1). The presence of these evolutionary patterns implies an evolutionary constraint that (i) is related to the structural or functional requirements of the protein and thus (ii) indicates a position that is sensitive to amino acid substitutions.

The current challenge to assessing whether rheostat positions have a particular bioinformatic signature is developing experimental datasets large enough for analysis. Addressing this question requires the availability of large (preferably whole-protein) substitution studies, in combination with sequence analyses. To date, our most clear-cut observation is that the locations of rheostat positions are not readily identified by co-evolutionary analyses. Furthermore, several new questions can already be asked. For example, is a position that is rheostatic in one homolog obligatorily rheostatic in other homologs? Indeed, a study of engineered transcription factors showed that the classification of a position can differ among paralogs (Meinhardt et al. 2013; Hodges et al. 2018). This indicates a limit to how much information can be gleaned from sequence analyses.

In contrast to the challenges associated with identifying rheostat positions via sequence alignments, the locations of

[1] Note that analyses to describe the evolutionary features of a protein's *positions* (e.g., position 55 is a highly conserved position) differ from the class of computer algorithms designed to predict the outcome of particular amino acid substitutions (e.g., V55D is a catastrophic substitution). Substitution algorithms almost universally incorporate sequence alignments, and sometimes position analyses, along with other types of information; see (Miller et al. 2017) for an overview of various techniques. Both substitution and position analyses require careful attention to which sequences are used to build the protein family. For further discussion about the issues of family size and score thresholding, we refer the reader to a Perspective devoted to these issues (Swint-Kruse 2016).

neutral positions may correlate well with those positions that have low sequence conservation (i.e., high sequence entropy) and that lack all types of evolutionary patterns. In one such example, we used several sequence analyses to generate a composite “least patterned” score and used it to successfully identify neutral and near-neutral positions in LPYK (Martin et al. 2020). The LPYK findings match those determined by the SNAP algorithm (Hecht et al. 2015), which has performed well at discriminating individual, neutral substitutions (e.g., (Wang and Bromberg 2019)). The Bromberg lab is now working to develop a predictor for the locations of rheostat, toggle, and neutral positions (Miller et al. 2019). Our findings for LPYK (Martin et al. 2020) indicate that one way to further improve the success at predicting neutral positions might be to include more types of sequence analyses in the predictor.

Structural characteristics of rheostat positions do not (yet) show clear features

Finally, rheostat positions must have some structural characteristic(s) that lead to the noncanonical substitution outcomes. By definitions, positions that modulate (rather than abolish) function must tolerate a wide range of substitutions without grossly altering the protein structure. To date, we have observed rheostat positions in globular soluble proteins (e.g., LPYK of Fig. 2) and in integral membrane proteins (e.g., OATP1B1 of Fig. 3); efforts are ongoing to determine whether rheostat positions exist in intrinsically disordered proteins. We have observed rheostat positions to have both near- and long-range effects on various functions, but no obvious correlation with proximity to binding/active sites (e.g., (Meinhardt et al. 2013; Wu et al. 2019)). As noted above, “multi-rheostat” positions may be enriched in allosteric regions of the protein. An intriguing possibility is that rheostat positions modulate protein dynamics, as hypothesized by Meinhardt et al. (2013). Finally, as has long been expected, the locations of neutral and near-neutral positions in LPYK were enriched on their surfaces; however, a large fraction of their surface positions were not neutral (Martin et al. 2020).

Conclusion

Rheostat positions will impact the field of pharmacogenomics in a myriad of ways. Rheostat positions are present in a wide variety of proteins and they do not follow canonical substitution rules. Substitutions at rheostat positions in allosteric sites will have (as yet) highly unpredictable outcomes on the activities of allosteric drugs and on drug specificity. Efforts to correlate signature patterns with the locations of rheostat positions are ongoing: they may be more likely to have intermediate evolutionary conservation levels and they might be identified by comparing disease and genome/exome databases. Structurally, rheostat positions may be located at key dynamic nodes. More studies are needed on rheostat positions in order to formulate new substitution rules and prediction algorithms, in order to advance the field of pharmacogenomics.

Acknowledgements This research was supported by the National Institutes of Health GM115340, GM118589, and GM077336, and by a grant from the W.M. Keck Foundation. The foundational studies that led to the discovery of rheostat positions were supported by Center of Biomedical Research Excellence in Protein Structure and Function (COBRE-PSF), with a grant from the National Institutes of Health P20 RR17708. The COBRE-PSF was led by Dr. Robert Hanzlik, for whom the authors express deep appreciation for his leadership and service to the University of Kansas and to other Kansas research institutions. We thank Melissa Ruggiero (KUMC) for discussions about the overlap between polymorphisms and rheostat positions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Acuna-Hidalgo R, Veltman JA, Hoischen A (2016) New insights into the generation and role of de novo mutations in health and disease. Genome Biol 17:241

Adamski CJ, Palzkill T (2017) Systematic substitutions at BLIP position 50 result in changes in binding specificity for class A β-lactamases. BMC Biochem 18:2

Andreoletti G, Pal LR, Moult J, Brenner SE (2019) Reports from the fifth edition of CAGI: the critical assessment of genome interpretation. Hum Mutat 40:1197–1201

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucl Acids Res 38(Suppl): W529–W533

Bubp J, Jen M, Matuszewski K (2015) Caring for glucose-6-phosphate dehydrogenase (G6PD)-deficient patients: implications for pharmacy. P T 40:572–574

Canu G, De Bonis M, Minucci A, Capoluongo E (2016) Red blood cell PK deficiency: an update of PK-LR gene mutation database. Blood Cells Mol Dis 57:100–109

Chan YH, Venev SV, Zeldovich KB, Matthews CR (2017) Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints. Nat Commun 8:14614

Chiti F, Taddei N, White PM, Bucciantini M, Magherini F, Stefani M, Dobson CM (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat Struct Biol 6:1005–1009

Creighton T (1993) Proteins: structures and molecular properties, 2nd edn. W. H. Freeman and Company, New York

Daura X (2019) Advances in the computational identification of allosteric sites and pathways in proteins. Adv Exp Med Biol 1163:141–169

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X (2015) Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 24:2125–2137

Fowler DM, Fields S (2014) Deep mutational scanning: a new style of protein science. Nat Methods 11:801–807

Gilbert GE, Novakovic VA, Kaufman RJ, Miao H, Pipe SW (2012) Conservative mutations in the C2 domains of factor VIII and factor V alter phospholipid binding and cofactor activity. Blood 120:1923–1932

Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19:163–164

Gray VE, Kukurba KR, Kumar S (2012) Performance of computational tools in evaluating the functional impact of laboratory-induced amino acid mutations. Bioinformatics 28:2093–2096

Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z, Zeng Y (2013) An update of DIVERGE software for functional divergence analysis of protein family. Mol Biol Evol 30:1713–1719

Guarnera E, Berezovsky IN (2020) Allosteric drugs and mutations: chances, challenges, and necessity. Curr Opin Struct Biol 62:149–157

Hargrove MS, Krzywda S, Wilkinson AJ, Dou Y, Ikeda-Saito M, Olson JS (1994) Stability of myoglobin: a model for the folding of heme proteins. Biochemistry 33:11767–11775

Hecht M, Bromberg Y, Rost B (2015) Better prediction of functional effects for sequence variants. BMC Genom 16(Suppl 8):S1

Hietpas RT, Jensen JD, Bolon DNA (2011) Experimental illumination of a fitness landscape. Proc Natl Acad Sci USA 108:7896–7901

Hodges AM, Fenton AW, Dougherty LL, Overholt AC, Swint-Kruse L (2018) RheoScale: A tool to aggregate and quantify experimentally determined substitution outcomes for multiple variants at individual protein positions. Hum Mutat 39:1814–1826

Holyoak T, Zhang B, Deng J, Tang Q, Prasannan CB, Fenton AW (2013) Energetic coupling between an oxidizable cysteine and the phosphorylatable N-terminus of human liver pyruvate kinase. Biochemistry 52:466–476

Ishwar A, Tang Q, Fenton AW (2015) Distinguishing the interactions in the fructose 1,6-bisphosphate binding site of human liver pyruvate kinase that contribute to allostery. Biochemistry 54: 1516–1524

Jonson PH, Petersen SB (2001) A critical view on conservative mutations. Protein Eng 14:397–402

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv. https://doi.org/10.1101/531210

Karp DA, Stahley MR, Garcia-Moreno B (2010) Conformational consequences of ionization of Lys, Asp, and Glu buried at position 66 in staphylococcal nuclease. Biochemistry 49:4138–4146

Klein SA, Majumdar A, Barrick D (2019) A second backbone: the contribution of a buried asparagine ladder to the global and local stability of a leucine-rich repeat protein. Biochemistry 58: 3480–3493

Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucl Acids Res 33:W299–W302

Lee Y, Mick J, Furdui C, Beamer LJ (2012) A coevolutionary residue network at the site of a functionally important conformational change in a phosphohexomutase enzyme family. PLoS ONE 7: e38114

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation C (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291

Link E, Parish S, Armitage J, Bowman L, Heath S, Matsuda F, Gut I, Lathrop M, Collins R (2008) SLCO1B1 variants and statin-induced myopathy–a genomewide study. N Engl J Med 359:789–799

Lipman DJ, Souvorov A, Koonin EV, Panchenko AR, Tatusova TA (2002) The relationship of protein conservation and sequence length. BMC Evol Biol 2:20

Luzzatto L, Arese P (2018) Favism and glucose-6-phosphate dehydrogenase deficiency. N Engl J Med 378:60–71

Markiewicz P, Kleina LG, Cruz C, Ehret S, Miller JH (1994) Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. J Mol Biol 240:421–433

Martin TA, Wu T, Tang Q, Dougherty LL, Parente DJ, Swint-Kruse L, Fenton AW (2020) Identification of biochemically neutral positions in liver pyruvate kinase. Proteins, in press. https://doi.org/10.1002/prot.25953

Matsuura Y, Takehira M, Makhatadze GI, Joti Y, Naitow H, Kunishima N, Yutani K (2018) Strategy for Stabilization of CutA1 Proteins Due to Ion-Ion Interactions at Temperatures of over 100 degrees C. Biochemistry 57:2649–2656

Meinhardt S, Manley Jr. MW, Parente DJ, Swint-Kruse L (2013) Rheostats and toggle switches for modulating protein function. PLoS ONE 8:e83502

Mihalek I, Res I, Lichtarge O (2006) Evolutionary trace report_maker: a new type of service for comparative analysis of proteins. Bioinformatics 22:1656–1657

Miller M, Bromberg Y, Swint-Kruse L (2017) Computational predictors fail to identify amino acid substitution effects at rheostat positions. Sci Rep. 7:41329

Miller M, Vitale D, Kahn PC, Rost B, Bromberg Y (2019) funtrp: identifying protein positions for variation driven functional tuning. Nucl Acids Res 47:e142

Modi T, Ozkan SB (2018) Mutations utilize dynamic allostery to confer resistance in TEM-1 beta-lactamase. Int J Mol Sci 19:3808

Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, Axelrod N, Busam DA, Strausberg RL, Venter JC (2008) Genetic variation in an individual human exome. PLoS Genet 4:e1000160

Niemi M, Pasanen MK, Neuvonen PJ (2011) Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. Pharmacol Rev 63:157–181

Ohnishi S, Hays A, Hagenbuch B (2014) Cysteine scanning mutagenesis of transmembrane domain 10 in organic anion transporting polypeptide 1B1. Biochemistry 53:2261–2270

Pál G, Kouadio J-LK, Artis DR, Kossiakoff AA, Sidhu SS (2006) Comprehensive and quantitative mapping of energy landscapes for protein–protein interactions by rapid combinatorial scanning. J Biol Chem 281:22378–22385

Parente DJ, Ray JC, Swint-Kruse L (2015) Amino acid positions subject to multiple coevolutionary constraints can be robustly identified by their eigenvector network centrality scores. Proteins 83:2293–2306

Parente DJ, Swint-Kruse L (2013) Multiple co-evolutionary networks are supported by the common tertiary scaffold of the LacI/GalR proteins. PLoS ONE 8:e84398

Pendergrass DC, Williams R, Blair JB, Fenton AW (2006) Mining for allosteric information: natural mutations and positional sequence conservation in pyruvate kinase. IUBMB Life 58:31–38

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612

Procko E (2020) The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. bioRxiv. https://doi.org/10.1101/2020.03.16.994236

Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA (2013) Analyses of the effects of all ubiquitin point mutants on yeast growth rate. J Mol Biol 425:1363–1377

Secrest MH, Storm M, Carrington C, Casso D, Gilroy K, Pladson L, Boscoe AN (2020) Prevalence of pyruvate kinase deficiency: a systematic literature review. Eur J Haematol, https://doi.org/10.1111/ejh.13424. Online ahead of print

Shannon C (1948) The mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656

Soskine M, Tawfik DS (2010) Mutational effects and the evolution of new protein functions. Nat Rev Genet 11:572–582

Suckow J, Markiewicz P, Kleina LG, Miller J, Kisters-Woike B, Müller-Hill B (1996) Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure. J Mol Biol 261:509–523

Swint-Kruse L (2016) Using evolution to guide protein engineering: the devil IS in the details. Biophys J 111:10–18

Swint-Kruse L, Zhan H, Fairbanks BM, Maheshwari A, Matthews KS (2003) Perturbation from a distance: mutations that alter LacI function through long-range effects. Biochemistry 42:14004–14016

Tang Q, Alontaga AY, Holyoak T, Fenton AW (2017) Exploring the limits of the usefulness of mutagenesis in studies of allosteric mechanisms. Hum Mutat 38:1144–1154

Tang Q, Fenton AW (2017) Whole-protein alanine-scanning mutagenesis of allostery: a large percentage of a protein can contribute to mechanism. Hum Mutat 38:1132–1143

Tungtur S, Schwingen KM, Riepe JJ, Weeramange CJ, Swint-Kruse L (2019) Homolog comparisons further reconcile in vitro and in vivo correlations of protein activities by revealing over-looked physiological factors. Protein Sci 28:1806–1818

Wang Y, Bromberg Y (2019) Identifying mutation-driven changes in gene functionality that lead to venous thromboembolism. Hum Mutat 40:1321–1329

Wu T, Swint-Kruse L, Fenton AW (2019) Functional tunability from a distance: rheostat positions influence allosteric coupling between two distant binding sites. Sci Rep. 9:16957

Ye K, Vriend G, AP IJ (2008) Tracing evolutionary pressure. Bioinformatics 24:908–915

Zeng Z, Bromberg Y (2019) Predicting functional effects of synonymous variants: a systematic review and perspectives. Front Genet 10:914

Zhan H, Swint-Kruse L, Matthews KS (2006) Extrinsic interactions dominate helical propensity in coupled binding and folding of the lactose repressor protein hinge helix. Biochemistry 45:5896–5906

Zhan H, Taraban M, Trewhella J, Swint-Kruse L (2008) Subdividing repressor function: DNA binding affinity, selectivity, and allostery can be altered by amino acid substitution of nonconserved residues in a LacI/GalR homologue. Biochemistry 47:8058–8069

Zhang Z, Norris J, Schwartz C, Alexov E (2011) In silico and in vitro investigations of the mutability of disease-causing missense mutation sites in spermine synthase. PLoS ONE 6:e20373