Received: 17 January 2017 | Revised: 16 March 2017 | Accepted: 24 March 2017

DOI: 10.1002/humu.23222

SPECIAL ARTICLE

WILEY HGVS HUMAN GENOME VARIATION SOCIETY

Benchmarking predictions of allostery in liver pyruvate kinase in CAGI4

Qifang Xu[1] | Qingling Tang[2] | Panagiotis Katsonis[3] | Olivier Lichtarge[3] | David Jones[4] | Samuele Bovo[5] | Giulia Babbi[5] | Pier L. Martelli[5] | Rita Casadio[5] | Gyu Rie Lee[6] | Chaok Seok[6] | Aron W. Fenton[2] | Roland L. Dunbrack Jr[1]

[1]Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania

[2]Department of Biochemistry and Molecular Biology, The University of Kansas Medical Center, Kansas City, Kansas

[3]Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, Texas

[4]Department of Computer Science, University College London, London, United Kingdom

[5]Biocomputing Group, CIG/Interdepartmental Center «Luigi Galvani» for Integrated Studies of Bioinformatics, Biophysics and Biocomplexity, University of Bologna, Bologna, Italy

[6]Department of Chemistry, Seoul National University, Seoul, Republic of Korea

Correspondence
Aron W. Fenton, The University of Kansas Medical Center, Biochemistry and Molecular Biology, MS 3030, 3901 Rainbow Boulevard, Kansas City, Kansas 66160.
Email: afenton@kumc.edu
Roland L. Dunbrack, Jr. Institute for Cancer Research, Fox Chase Cancer Center, 333 Cottman Ave., Philadelphia, PA 19111.
Email: roland.dunbrack@fccc.edu

Contract grant sponsors:NIH (R01 GM084453, R13 HG006650, U41 HG007346, R13 HG006650).

For the CAGI Special Issue

Abstract

The Critical Assessment of Genome Interpretation (CAGI) is a global community experiment to objectively assess computational methods for predicting phenotypic impacts of genomic variation. One of the 2015–2016 competitions focused on predicting the influence of mutations on the allosteric regulation of human liver pyruvate kinase. More than 30 different researchers accessed the challenge data. However, only four groups accepted the challenge. Features used for predictions ranged from evolutionary constraints, mutant site locations relative to active and effector binding sites, and computational docking outputs. Despite the range of expertise and strategies used by predictors, the best predictions were marginally greater than random for modified allostery resulting from mutations. In contrast, several groups successfully predicted which mutations severely reduced enzymatic activity. Nonetheless, poor predictions of allostery stands in stark contrast to the impression left by more than 700 PubMed entries identified using the identifiers "computational + allosteric." This contrast highlights a specialized need for new computational tools and utilization of benchmarks that focus on allosteric regulation.

KEYWORDS

allosteric effect, CAGI experiment, liver pyruvate kinase, missense mutation

1 | INTRODUCTION

Blind challenge experiments, such as CASP (Moult et al., 2016) and CAPRI (Lensink et al., 2017), have provided independent assessment of computational prediction methods in structural biology. They have spurred the development of new methods and the integration of multiple methods in prediction pipelines. The Critical Assessment of Genome Interpretation (CAGI) experiment seeks to achieve the same goals by providing prediction challenges in a number of different areas. In this report, we describe a challenge involving the effect of mutations on the allosteric coupling of effectors and substrate binding to human liver pyruvate kinase (L-PYK). The focus of this competition was to predict the influence of mutations on the allosteric regulation of L-PYK by a negative regulator, alanine, and a positive effector, fructose-1,6-bisphosphate (Fru-1,6-BP). Numerous methods for predicting the effect of mutations on allosteric effector binding have been published in recent years (Collier & Ortiz, 2013; Feher et al., 2014).

The definition of allostery applicable to studies of L-PYK is the affinity of the enzyme for its substrate, phosphoenolpyruvate (PEP), in the absence versus presence of an allosteric effector, recognizing that the effector binds to a site distinct from the active site (Carlson & Fenton, 2016; Fenton, 2008, 2012; Fenton & Alontaga, 2009; Fenton &

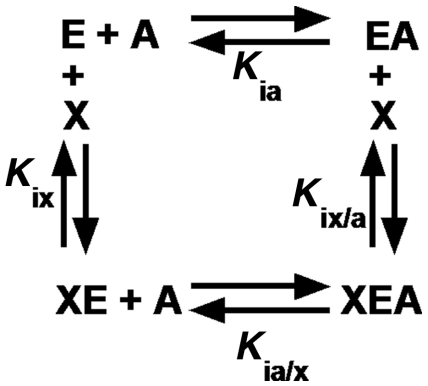

FIGURE 1 Reaction scheme for an allosteric energy cycle in which an enzyme (E) can bind one substrate (A) and one allosteric effector (X). K_{ia} is the equilibrium dissociation constant of the substrate binding to the enzyme in the absence of effector. $K_{ia/x}$ is the equilibrium dissociation constant of the substrate binding to the enzyme in the presence of saturating concentrations of effector. K_{ix} is the equilibrium dissociation constant of the effector when substrate is absent, whereas $K_{ix/a}$ is the equilibrium dissociation constant of effector in the presence of saturating concentrations of substrate

Hutchinson, 2009; Fenton et al., 2010; Ishwar et al., 2015). This definition describes allostery by four enzyme forms that constitute the corners of a thermodynamic energy cycle (Fig. 1), and it provides a mechanism to quantify allosteric function in the form of the allosteric coupling constant (Q_{ax}) (Fenton, 2012; Reinhart, 1983, 1988, 2004; Weber, 1972):

$$Q_{ax} = \frac{K_{ia}}{K_{ia/x}} = \frac{K_{ix}}{K_{ix/a}}$$

K_{ia} and $K_{ia/x}$ are equilibrium dissociation constants for binding the substrate (A) in the absence or presence, respectively, of an allosteric effector, X, as defined in Figure 1. $Q_{ax} = 1$ indicates that the system is not allosteric. When $Q_{ax} > 1$, there is positive allosteric coupling between the binding of X to a protein and the binding of A to the same protein at distinct sites. When $Q_{ax} < 1$, there is a negative or inhibitory coupling between the X and A sites.

The predictors were provided two sets of mutations for predictions of enzyme activity and allosteric effects in L-PYK. Q_{ax} was determined for each active mutant protein by determining PEP affinity (via titrations of activity over a concentration range of PEP) over a concentration range of effector. Experiment 1 consisted of 113 mutations at nine sites in or near to the binding of the negative allosteric regulator, alanine. Participants were asked to provide a probability that each mutant enzyme was active (i.e., not the level of activity) and the value of Q_{ax} for alanine for each mutant. Experiment 2 consisted of mutations to alanine at 430 sites throughout the protein. Participants were then asked to predict the enzyme activity and Q_{ax} values for the effectors alanine and Fru-1,6-BP. Since alanine is a negative regulator, all values of $Q_{ax\text{-}Ala}$ are between 0 and 1, whereas the value of Q_{ax} for Fru-1,6-BP is unbounded. Predictors were provided with the maximum value ($Q_{ax\text{-}Fru\text{-}1,6\text{-}BP} = 320$) found in the alanine-scanning experiment.

2 | METHODS AND MATERIALS

2.1 | Experimental data generation

Wild-type and mutant human L-PYK were expressed in the *E. coli* FF50 strain, which lacks endogenous *pyk* genes, and partially purified using ammonium sulfate fractionation followed by dialysis, as previously described (Fenton & Alontaga, 2009; Ishwar et al., 2015). L-PYK catalyzes the following reaction:

$$\text{Phosphoenolpyruvate} + \text{ADP} \rightarrow \text{Pyruvate} + \text{ATP}$$

Activity measurements were performed at 30°C using a lactate dehydrogenase assay to detect the production of pyruvate by L-PYK. Lactate dehydrogenase catalyzes the following reversible reaction:

$$\text{Pyruvate} + \text{NADH} \rightleftharpoons \text{NAD}^{*} + \text{Lactate}$$

As the L-PYK reaction proceeds, producing pyruvate, the concentration of NADH decreases, which can be detected by monitoring absorbance at 340 nm (A_{340}). Reaction conditions contained 50 mM HEPES or bicine, 10 mM $MgCl_2$, 2 mM (K)ADP, 0.1 mM EDTA, 0.18 mM NADH, and 19.6 U/ml lactate dehydrogenase. PEP and effector concentrations were varied. The rate of the decrease in A_{340} due to NADH utilization was recorded at each concentration of PEP and these initial velocity rates as a function of PEP concentration were used to evaluate the apparent affinity for PEP ($K_{app\text{-}PEP}$) at any one effector concentration. K_{ix} and Q_{ax} for each mutant and the wild type were obtained by fitting the observed $K_{app\text{-}PEP}$ to the equation:

$$K_{app-PEP} = K_a \left(\frac{K_{ix} + [X]}{K_{ix} + Q_{ax}[X]} \right)$$

where $K_a = K_{app\text{-}PEP}$ when the concentration of effector $[X] = 0$.

The dataset represents two experiments, which are characterizations of mutant human L-PYK proteins expressed in *E. coli*, named experiment 1 and experiment 2. Experiment 1 consisted of site-directed mutations at residue positions with a side chain contacting with alanine or very near the bound alanine. A total of 113 substitutions were introduced at nine different sites, of which 23 mutant proteins were completely inactive (no measurable enzyme activity). $Q_{ax\text{-}Ala}$ was determined for the 90 mutant proteins with activity. In experiment 2, 430 residues were mutated into alanine across the entire protein, of which 44 did not have detectable enzyme activity. Allosteric coupling Q_{ax} for inhibition by alanine and activation by Fru-1,6-BP were separately determined.

2.2 | Performance assessment of L-PYK enzyme activity

From the binary experimental enzyme activity data (1 = positive = active; 0 = negative = inactive), we calculated the number of true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) for all participating groups in experiment 1 and experiment 2. From these, we calculated the true-positive rate (TPR),

TABLE 1 Groups participating in L-PYK enzyme activity and allostery prediction challenges

Group number	Affiliation	Authors
53	Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, TX	Panagiotis Katsonis, Olivier Lichtarge
54	Department of Computer Science, University College London, Gower Street, London WC1E 6BT, United Kingdom	David Jones
55	Biocomputing Group, CIG/Interdepartmental Center «Luigi Galvani» for Integrated Studies of Bioinformatics, Biophysics and Biocomplexity, University of Bologna, Bologna, Italy	Samuele Bovo, Giulia Babbi, Pier Luigi Martelli, Rita Casadio
56	Department of Chemistry, Seoul National University, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea	Gyu Rie Lee, Chaok Seok

true-negative rate (TNR), positive predictive value (PPV), and negative predictive value (NPV):

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$$

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}}$$

We also calculated four measures that assess overall accuracy: total accuracy (ACC), balanced accuracy (BACC), Matthews correlation coefficient (MCC) (Matthews, 1975), and F1 score. F1 score is the harmonic mean of precision (PPV) and sensitivity (TPR).

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$$

$$\mathrm{BACC} = \frac{1}{2}(\mathrm{TPR}+\mathrm{TNR})$$

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

$$\mathrm{F1} = 2\frac{\mathrm{TPR}\times\mathrm{PPV}}{\mathrm{TPR}+\mathrm{PPV}}$$

Since some predictors provided real values (between 0 and 1), these were converted into binary predictions as described below in the *Results* section.

2.3 | Evaluation of predictions of $Q_{ax\text{-}Ala}$ and $Q_{ax\text{-}Fru\text{-}1,6\text{-}BP}$

Spearman's rho (ρ), or Spearman's rank correlation coefficient, measures the monotonic correlation between prediction and experimental data. $\rho = 1$ means the predictions and experimental data points have identical rankings. For data set (p_i, e_i), prediction data points are converted into ranks Rp_i, and experimental data points are converted into ranks Re_i. Then, ρ is calculated from the formula:

$$\rho = \frac{\mathrm{cov}(\mathrm{Rp},\mathrm{Re})}{\sigma_{\mathrm{Rp}}\sigma_{\mathrm{Re}}}, \quad -1 \le \rho \le 1$$

Kendall's tau (τ), or Kendall rank correlation coefficient, like Spearman's rho, measures the rank correlation between two variables. For data set (p, e), any pair of (p_i, e_i) and (p_j, e_j), where $i \neq j$, are said to be concordant if both $p_i > p_j$ and $e_i > e_j$, or if both $p_i < p_j$ and $e_i < e_j$. They are discordant, if both $p_i > p_j$ and $e_i < e_j$, or if $p_i < p_j$ and $e_i > e_j$. If $p_i = p_j$ or $e_i = e_j$, the pair is neither concordant nor discordant. We use C for the set of concordant pairs, and D for the set of discordant pairs. τ is defined as the difference between the number of concordant pairs (|C|) and the number of discordant pairs (|D|), divided by the total number of pair combinations ($n \times (n-1)/2$). The formula is given as following:

$$\tau = \frac{|C| - |D|}{n(n-1)/2}$$

All statistical calculations and kernel density estimates of the data were performed in R (R Core Team, 2015).

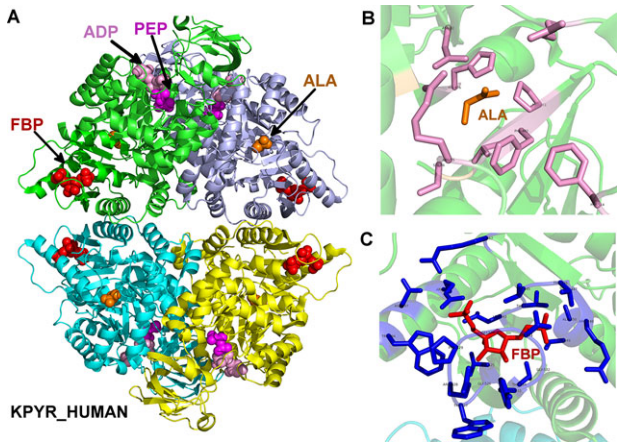

FIGURE 2 Structure of human pyruvate kinase, as well as the binding sites of inhibitor alanine and activator fructose-1,6-bisphosphate. A: A modeled structure of L-PYK tetramer with substrates PEP and ADP, allosteric inhibitor alanine, and allosteric activator. PEP, ADP, alanine (labeled ALA), and fructose-1,6-bisphosphate (labeled FBP) are shown in spheres, colored in magenta, pink, orange, and red, respectively. The structure was assembled by superposing monomers from several structures of homologues of L-PYK with PEP, ADP, and alanine bound onto a tetrameric structure of human L-PYK with fructose-1,6-bisphosphate bound (PDB: 4IP7). B: The allosteric binding site of alanine. Alanine is shown in sticks and colored in orange. Residues that were mutated in experiment 1 are shown in sticks, and colored in pink. C: The binding site of fructose-1,6-bisphosphate (FBP). FBP is shown in sticks and colored in red. Interacting residues are shown in sticks and colored in blue

3 | RESULTS

In this assessment, four groups (53, 54, 55, and 56; Table 1) submitted a total of five prediction sets, of which two were from group 56, labeled 56_1 and 56_2. The methods utilized by each group are provided in the Supp. Materials as are the instructions and information provided to predictors at the time of the experiment.

Human L-PYK is a tetrameric enzyme with distinct binding sites for its reactants, pyruvate, and ADP, and its allosteric effectors, alanine, and Fru-1,6-BP. The structure of the tetramer is shown in Figure 2A, where molecules at the three sites are represented as spheres in each monomer. This composite structure was created by superposing monomers from structures containing alanine (PDB: 2G50, a structure of rabbit L-PYK) (Williams et al., 2006), PEP (PDB: 4HYV, *Trypanosoma brucei* pyruvate kinase) (Zhong et al., 2013), and ADP (PDB: 3GR4, human pyruvate kinase M2) (Hong et al., unpublished, DOI: 10.2210/pdb3gr4/pdb) onto each member of the tetrameric biological assembly of human L-PYK (PDB: 4IP7) (Holyoak et al., 2013). Experiment 1 consisted of 113 mutations spread across nine amino acid positions in or near the alanine-binding site (Fig. 2B): Arg55, Ser56, Asn82, Arg118, His476, Val481, Pro483, and Phe514. Experiment 2 consisted of alanine-scanning mutations across the entire protein, except wild-type positions that are Gly or Ala. The Fru-1,6-BP site is shown in Figure 2C.

3.2 | Prediction of L-PYK enzyme activity

The first challenge was to provide a probability that each enzyme was active. This was a binary outcome, not the level of activity. Even weakly active enzymes were considered active in the experiment. In both experiments, some mutants had no detectable activity, and these were labeled 0; the rest were labeled 1. The active mutants included some enzymes with very low but detectable activity. In experiment 1, 79.6% of mutants were active and 20.4% were inactive. In experiment 2, 88.8% of the mutants were active and 10.2% were inactive. Two of the groups (53 and 54) submitted real values between 0 and 1, instead of binary indicators. For these groups, we labeled all predictions with values ≥0.5 as active and the rest as inactive. Figure 3 shows the density functions of predicted enzyme activities. For experiment 1, two groups (55 and 56_2) predicted all mutants to be active (a value of 1) (Fig. 3, top row). This is not unreasonable since all of the mutations were in or near the alanine effector-binding site, which is distant from the active site.

Table 2 provides an assessment of the predictions of enzyme activity for each group for both experiments. We also included values obtained from the PolyPhen-2 server, which is commonly used to predict phenotypes of missense mutations (Adzhubei et al., 2010). Group 56 achieved the highest ACC in both experiments (ACC of 0.867 for group 56_1 in experiment 1; ACC of 0.894 for group 56_2 in experiment 2). Since the goal was to predict whether enzymes were active or inactive, rather than the level of activity, this is a successful result. In the case of experiment 1, predicting all mutants as active would result in an accuracy of 0.796, whereas in experiment 2, a value of 0.888 would be obtained. At least for experiment 1, group 56 achieved better predictions than the simple prediction that all mutants were active.

In most binary phenotype prediction assessments (Wei & Dunbrack, 2013), it is important to balance the success of positive predictions and/or experimental outcomes with negative predictions and/or experimental outcomes. One such measure is the BACC, which is the average of the rate of correctly predicting the experimentally active mutants (TPR) and the rate of correctly predicting the experimentally inactive mutants (TNR). For experiment 1, only groups 53 and 56_1 achieved BACC values above 0.5, with BACC = 0.768 and 0.755, respectively. A BACC of 0.50 is trivial to achieve, since if one predicts all of the phenotypes in one class, the BACC is automatically 0.50 (e.g., groups 55 and 56_2 for experiment 1). Groups 53 and 56_1 achieved their results in contrasting manners: group 53 has low TPR and high TNR, and group 56_1 has high TPR and low TNR. For experiment 2, which contained mutations across the entire protein and is therefore a more real-world prediction task, only group 53 has TPR and TNR > 0.5, resulting in a BACC of 0.745.

Similarly, the MCC and F1 values also balance positive and negative predictions and experimental values but in different ways than BACC (see *Materials and Methods*). F1, in particular, only includes positive predictions and experimental phenotypes and omits negative predictions and phenotypes. Since both data sets consisted of majority of active enzymes (80% and 88% for experiments 1 and 2, respectively), groups that predicted a larger fraction of the enzymes to be active did better in F1 (groups 55, 56_1, and 56_2) than the other groups. Group 54 predicted a majority of the mutants to be inactive in both experiments and thus achieved much lower values for F1 than the other groups.

We compared the results of CAGI groups with that of PolyPhen-2, a server that is commonly used to predict the phenotypes of missense mutations in proteins. PolyPhen-2, like other servers, predicts phenotypes to be deleterious or neutral, or "damaging" versus "benign." This is not necessarily directly associated with enzyme activity, since a deleterious mutation might affect protein expression or the ability to regulate the protein by allosteric mechanisms. Also, the inactive enzymes were only those with no activity, and not those with significant reduction in activity. In experiment 1, PolyPhen-2 predicted most mutants to be inactive, probably because the alanine-binding site is very highly conserved in L-PYK enzymes in order to retain the negative effector capability of alanine. This resulted in a BACC of 0.539. In experiment 2, mutations were spread across the protein and PolyPhen-2 does better, with a BACC of 0.674. Nevertheless, group 53 was able to achieve better results on all four measures of overall success in experiment 2.

As mentioned above, groups 53 and 54 provide real values (not binary values) for the enzyme activity. We speculated that a cutoff of 0.5 might not be ideal to turn their real values into binary predictions. We calculated BACC as function of the cutoff and found that for group 53, a value of 0.5 was still the best for both experiments. But for group 54, values of 0.3 for experiment 1 and 0.35 for experiment 2 provide better results. The values of BACC are 0.724 and 0.696, respectively, which are much better than the 0.5 cutoff (0.534 and 0.627, respectively). But this is only possible with reference to the experimental data, which would not be available in real-world situations. Since the density for predictions for group 54 were

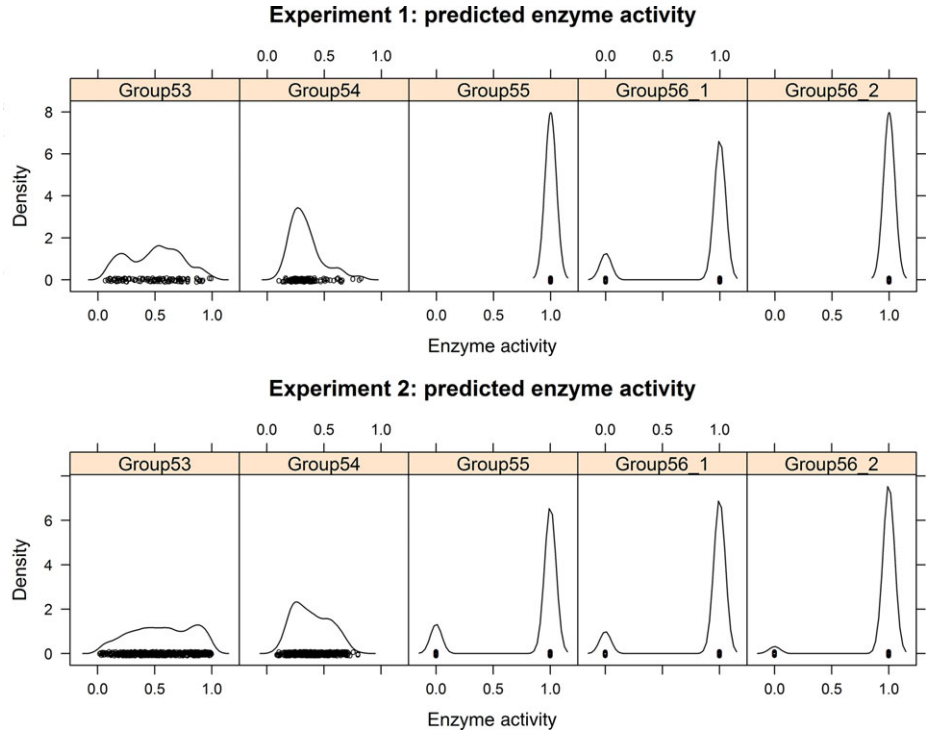

FIGURE 3 Kernel density estimates of five sets of predicted L-PYK enzyme activities

TABLE 2 Binary prediction results of L-PYK enzyme activity

	Experiment 1						Experiment 2					
Method	Group 53	Group 54	Group 55	Group 56_1	Group 56_2	PPH2	Group 53	Group 54	Group 55	Group 56_1	Group 56_2	PPH2
TPR	0.622	0.156	1	0.944	1	0.122	0.626	0.322	0.838	0.898	0.976	0.392
TNR	0.913	0.913	0	0.565	0	0.957	0.864	0.932	0.205	0.318	0.182	0.953
PPV	0.966	0.875	0.796	0.895	0.796	0.917	0.976	0.976	0.901	0.920	0.912	0.987
NPV	0.382	0.216	0	0.722	0	0.218	0.210	0.137	0.127	0.264	0.471	0.150
ACC	0.681	0.310	0.796	**0.867**	0.796	0.292	0.650	0.385	0.772	0.838	**0.894**	0.449
BACC	**0.768**	0.534	0.5	0.755	0.5	0.539	**0.745**	0.627	0.521	0.608	0.579	0.673
MCC	0.431	0.079	0	**0.561**	0	0.103	**0.301**	0.169	0.034	0.199	0.246	0.218
F1	0.757	0.264	0.887	**0.919**	0.887	0.217	0.762	0.484	0.868	0.907	**0.943**	0.562

Notes:
The highest score in each row for the four global measures is in bold and underlined.
0, inactive; 1, active.
TPR, true-positive rate; FPR, false-positive rate; TNR, true-negative rate; PPV, positive predictive value; NPV, negative predictive value; ACC, accuracy; BACC, balanced accuracy; MCC, Matthews correlation coefficient; F1, F1 score.

unimodal (Fig. 3), it was not possible to define a cutoff based on a minimum of density between a low-activity and a high-activity mode in the data.

3.3 | Prediction of allosteric inhibition of alanine ($Q_{ax\text{-}Ala}$)

The second challenge was to estimate the inhibitory allosteric effect of binding alanine, $Q_{ax\text{-}Ala}$ on binding of the substrate PEP. The density estimates of experimental $Q_{ax\text{-}Ala}$ values of two experiments are shown in Figure 4. The wild-type enzyme had a $Q_{ax\text{-}Ala}$ value of ~0.08 in both experiments. In experiment 1, 23 out of 90 mutants did not have measurable allosteric coupling, shown in a peak at $Q_{ax} = 1$ (Fig. 4, left). One possiblity is that alanine continues to bind to these mutant proteins, but that binding does not alter PEP affinity. In other cases, the $Q_{ax} = 1$ outcome is likely because the mutation eliminated binding of Ala to L-PYK altogether (at least to the maximum concentration tested in the experiments). In experiment 2, after excluding 37 mutants for which the allosteric coupling effect could not be measured, the $Q_{ax\text{-}Ala}$ values of 325 (83%) mutants were between 0 and 0.2, relatively similar to the wild-type enzyme.

A comparison by scatter plot of the experimental and the predicted $Q_{ax\text{-}Ala}$ values is shown in Figure 5. Group 55 provided only binary prediction for $Q_{ax\text{-}Ala}$. Group 56_1 and 56_2 provided identical values for both experiments. The scatter plots do not show any obvious correlations between the predicted and experimental $Q_{ax\text{-}Ala}$.

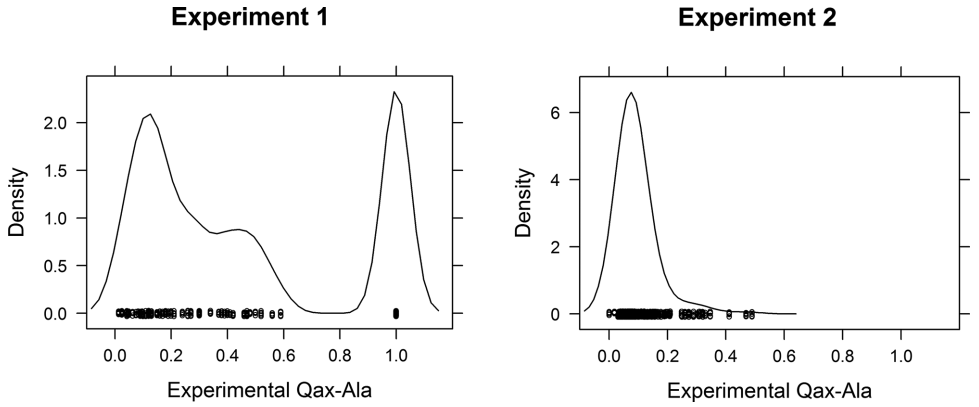

FIGURE 4 Kernel density estimates of experimental $Q_{ax\text{-}Ala}$ values of experiments 1 and 2

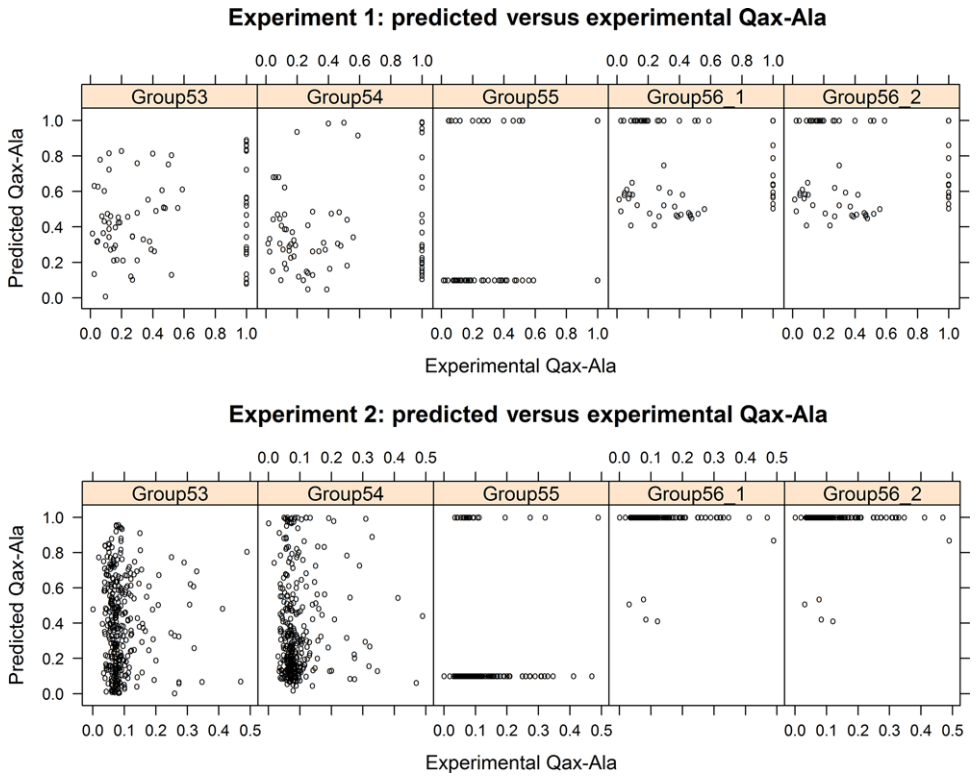

FIGURE 5 Scatter plot of the experimental $Q_{ax\text{-}Ala}$ versus the predicted $Q_{ax\text{-}Ala}$ values

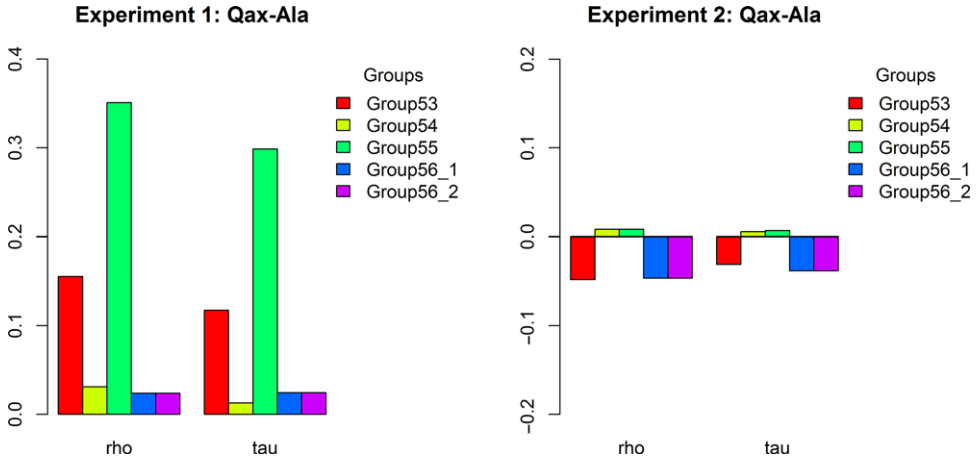

FIGURE 6 Correlations represented by Spearman's ρ and Kendall's τ between the predicted and experimental $Q_{ax\text{-}Ala}$ values of two experiments

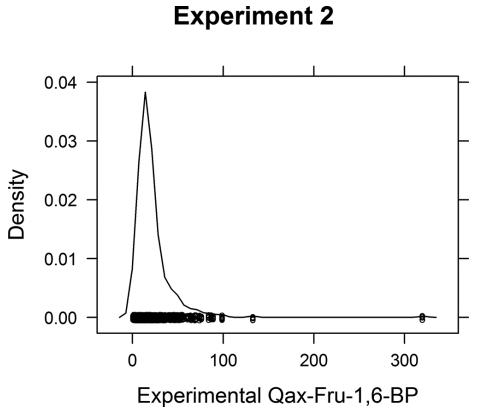

FIGURE 7 Kernel density estimate of experimental $Q_{\text{ax-Fru-1,6-BP}}$ from experiment 2

We calculated Spearman's ρ and Kendall's τ coefficients as nonparametric tests of the correlation of the predictions with the experiments, since the data and predicted values are not unimodal or normally distributed. Only group 55 in experiment 1 achieves a favorable correlation, with $\rho = 0.351$ and $\tau = 0.299$ with P values of 0.002 for both (Fig. 6). All of the other P values are in the range of 0.17–0.88, which implies there is no correlation between the predicted and experimental $Q_{\text{ax-Ala}}$ values. If we treat the experimental $Q_{\text{ax-Ala}}$ values as binary for experiment 1 (Fig. 4, left), we can calculate binary assessment measures such as TPR, TNR, and so on. We did this for group 55, which provided binary prediction values (0.1 and 1.0) with the following results (where positive indicates $Q_{\text{ax-Ala}} = 1$): TPR = 17/23 = 0.739; TNR = 39/55 = 0.709; BACC = 0.724. This is better than random and explains the positive correlation coefficients.

The results for experiment 2 are negatively correlated for three of the groups, and only very weak positive correlations were achieved by groups 54 and 55 (Fig. 6, right). The P values are in the range of 0.38–0.88.

3.4 | Prediction of allosteric activation of Fru-1,6-BP ($Q_{\text{ax-Fru-1,6-BP}}$)

Participants were asked to predict the allosteric effect of Fru-1,6-BP binding to L-PYK for the mutants created in experiment 2 and were told that the maximum value in the experiments was 320. The wild-type protein has a $Q_{\text{ax-Fru-1,6-BP}}$ value of 14.2. The density estimate of experimental $Q_{\text{ax-Fru-1,6-BP}}$ values is shown in Figure 7, showing that the vast majority of mutants had values between 0 and 60. The scatter plots of the predicted $Q_{\text{ax-Fru-1,6-BP}}$ versus experimental $Q_{\text{ax-Fru-1,6-BP}}$ show that groups 53 and 54 provided real values over the full range of the experimental values and group 55 provided discrete values (1, 50, 250, and 320), whereas group 56 provided an approximate wild-type value of 15.3 for most of the mutants and other values for 18 mutants in the range from 1 to 28.3 (Fig. 8).

We calculated Spearman's ρ and Kendall's τ to evaluate the correlations between predicted and experimental $Q_{\text{ax-Fru-1,6-BP}}$ values (Fig. 9). Only group 55 has positive correlations, both very marginal (both ρ and $\tau \sim 0.05$, with P value of 0.2). All others have negative correlations, especially for group 53 and 54. The P values of group 53 are 7.5E-05 for ρ and 8.98E-05 for τ, and the P values of group 54 are 0.0003 for both ρ and τ.

4 | DISCUSSION

We may summarize the results of the CAGI experiment on L-PYK as follows. Groups 53 and 56 had good predictions of the L-PYK enzyme activity in experiments 1 and 2 as measured by BACC (group 53) and ACC (group 56). In these cases, the results were better than that achieved by PolyPhen-2. Group 54 had good predictions only if we set a new cutoff for binary enzyme activity from their real-valued results in both experiments 1 and 2.

For the prediction of allosteric effects of alanine and fructose, groups 55 and 53 had positive correlations for the $Q_{\text{ax-Ala}}$ challenge in experiment 1, but only group 55 had a statistically significant positive correlation. No group had statistically significant, positive correlations for their predictions of $Q_{\text{ax-Ala}}$ or $Q_{\text{ax-Fru-1,6-BP}}$ in experiment 2.

At the conclusion of this experiment, we are left to contemplate why the overall success of predicting allosteric effects was underwhelming. This consideration is particularly valuable given the indications of success of computational approaches reported in the literature. As noted, the only statistically significant result for predicting allosteric data was for group 55 on the $Q_{\text{ax-Ala}}$ challenge in experiment 1. This group used a very simple model that considered the distance each wild-type residue was from bound Ala (as modeled from the structure of human pyruvate kinase M2) and the severity of the mutation from wild type (as determined by scores from a substitution matrix). It is likely that they correctly predicted many of the mutations that abrogated Ala binding altogether ($Q_{\text{ax-Ala}} = 1$), rather than quantitatively predicting the effect of the mutations on the diverse values of $Q_{\text{ax-Ala}}$ of the remaining mutations ($Q_{\text{ax-Ala}} < 1$). It is not likely that their distance-based method would extend readily to the general problem of predicting allosteric effects, especially for residues not in or near the binding site. The results for experiment 2, where mutations were made throughout the protein, confirm this.

It is also clear from the experiment that methods that predominantly used evolutionary considerations (groups 53 and 54) were not able to predict the effects of mutation on allosteric behavior. Group 53 used the evolutionary action of each mutation, a number that can be calculated from phylogenetic sequence analysis (Katsonis & Lichtarge, 2014). Group 54 used covariation of amino acids in pairs of positions within a multiple sequence alignment of homologues of L-PYK (Jones et al., 2015).

Group 56 calculated the binding affinity of each effector to each mutant with docking calculations (Shin et al., 2013), and made the assumption that Q_{ax} was directly proportional to these values. In fact, $Q_{\text{ax}} = K_{\text{ix}}/K_{\text{ix/a}}$ where K_{ix} is the equilibrium dissociation constant of the effector X and $K_{\text{ix/a}}$ is the equilibrium dissociation constant of the effector X when the substrate A is bound. The approximation is not unreasonable given the experimental data from experiment 2: the Pearson and Kendall correlation coefficients between the experimental values of Q_{ax} and K_{ix} for alanine are 0.73 and 0.59, respectively, and for Fru-1,6-BP they are 0.80 and 0.64, respectively (all P values $< 1.0 \times 10^{-15}$).

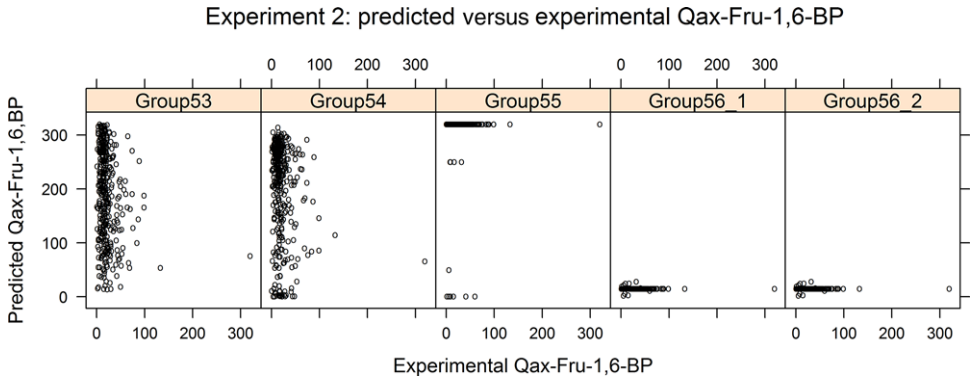

FIGURE 8 Scatter plot of the predicted versus experimental $Q_{ax\text{-}Fru\text{-}1,6\text{-}BP}$ values from experiment 2

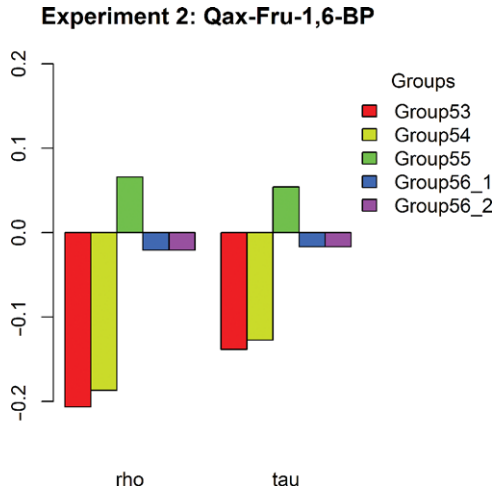

FIGURE 9 Correlations represented by Spearman's ρ and Kendall's τ between the predicted and experimental $Q_{ax\text{-}Fru\text{-}1,6\text{-}BP}$ values in experiment 2

Group 56 only performed docking calculations to mutations in the binding sites of alanine and Fru-1,6-BP, and submitted values for all other positions of 1.0 for $Q_{ax\text{-}Ala}$ (no inhibition of PEP-binding by Ala) and 15.3 for $Q_{ax\text{-}Fru\text{-}1,6\text{-}BP}$ (the experimental value). This resulted in only eight mutations with $Q_{ax\text{-}Ala}$ not equal to 1.0, only five of which had experimental values available. If we restrict the calculation of correlation coefficients to these five values, the *P* values for the Spearman and Kendall correlation coefficients are greater than 0.8, and the values of rho and tau are 0.1 and 0, respectively. For $Q_{ax\text{-}Fru\text{-}1,6\text{-}BP}$, group 56 produced values for 17 mutations adjacent to the Fru-1,6-BP site, only 11 of which had enough enzyme activity to measure $Q_{ax\text{-}Fru\text{-}1,6\text{-}BP}$. The correlation coefficients with $Q_{ax\text{-}Fru\text{-}1,6\text{-}BP}$ were both ~0.2 with *P* values of ~0.5. Unless docking calculations are able to discern changes in binding affinity of the effector (in the presence or absence of the substrate) for sites far from their binding sites, it is not possible to determine whether such calculations provide valuable information on allosteric behavior.

It is clear from the quality of predictions in this study that additional approaches are needed. Many of the methods reported in the literature involve molecular dynamics simulations that are very computationally intensive (Blacklock & Verkhivker, 2014; Hertig et al., 2016; Weinkam et al., 2012). Several simulations of other forms of pyruvate kinase (Naithani et al., 2015) and mutants thereof have been performed (Kalaiarasan et al., 2015). However, whether such methods could be used in a predictive fashion has yet to be determined. The current data set could be used to benchmark such methods, if a sufficient number of mutants can be simulated.

Allosteric regulation is sometimes presented as a Rube Goldberg-type mechanism initiated by the effector associating with the enzyme/protein (binding causes change A; change A causes change B; change B causes change C, etc.). However, the definition for allostery based on an energy cycle (Fig. 1) implies that allostery is an equilibrium mechanism (Carlson & Fenton, 2016). As such, the allosteric mechanism would be a comparison of changes in the fully equilibrated enzyme forms represented in Figure 1 and not a Rube Goldberg mechanism that would be associated with a kinetics mechanism. Calculations of this sort remain a challenge for computational approaches to predicting the effects of mutations on allosteric regulation.

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249.

Blacklock, K., & Verkhivker, G. M. (2014). Computational modeling of allosteric regulation in the hsp90 chaperones: A statistical ensemble analysis of protein structure networks and allosteric communications. *PLoS Computational Biology*, *10*(6), e1003679.

Carlson, G. M., & Fenton, A. W. (2016). What mutagenesis can and cannot reveal about allostery. *Biophysical Journal*, *110*(9), 1912–1923.

Collier, G., & Ortiz, V. (2013). Emerging computational approaches for the study of protein allostery. *Archives of Biochemistry and Biophysics*, *538*(1), 6–15.

Feher, V. A., Durrant, J. D., Van Wart, A. T., & Amaro, R. E. (2014). Computational approaches to mapping allosteric pathways. *Current Opinion in Structural Biology*, *25*, 98–103.

Fenton, A. W. (2008). Allostery: An illustrated definition for the 'second secret of life'. *Trends in Biochemical Sciences*, *33*(9), 420–425.

Fenton, A. W. (Ed.). (2012). *Allostery: Methods and protocols*). New York: Humana Press: Springer Science.

Fenton, A. W., & Alontaga, A. Y. (2009). The impact of ions on allosteric functions in human liver pyruvate kinase. *Methods in Enzymology*, *466*, 83–107.

Fenton, A. W., & Hutchinson, M. (2009). The pH dependence of the allosteric response of human liver pyruvate kinase to fructose-1,6-bisphosphate, ATP, and alanine. *Archives of Biochemistry and Biophysics*, *484*, 16–23.

Fenton, A. W., Johnson, T. A., & Holyoak, T. (2010). The pyruvate kinase model system, a cautionary tale for the use of osmolyte perturbations to support conformational equilibria in allostery. *Protein Science*, *19*, 1796–1800.

Hertig, S., Latorraca, N. R., & Dror, R. O. (2016). Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. *PLOS Computational Biology*, *12*(6), e1004746.

Holyoak, T., Zhang, B., Deng, J., Tang, Q., Prasannan, C. B., & Fenton, A. W. (2013). Energetic coupling between an oxidizable cysteine and the phosphorylatable N-terminus of human liver pyruvate kinase. *Biochemistry*, *52*(3), 466–476.

Ishwar, A., Tang, Q., & Fenton, A. W. (2015). Distinguishing the interactions in the fructose 1,6-bisphosphate binding site of human liver pyruvate kinase that contribute to allostery. *Biochemistry*, *54*(7), 1516–1524.

Jones, D. T., Singh, T., Kosciolek, T., & Tetchner, S. (2015). MetaPSICOV: Combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics*, *31*(7), 999–1006.

Kalaiarasan, P., Kumar, B., Chopra, R., Gupta, V., Subbarao, N., & Bamezai, R. N. (2015). In silico screening, genotyping, molecular dynamics simulation and activity studies of SNPs in pyruvate kinase M2. *PLOS ONE*, *10*(3), e0120469.

Katsonis, P., & Lichtarge, O. (2014). A formal perturbation equation between genotype and phenotype determines the Evolutionary Action of protein-coding variations on fitness. *Genome Research*, *24*(12), 2050–2058.

Lensink, M. F., Velankar, S., & Wodak, S. J. (2017). Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins*, *85*, 359–377.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, *405*(2), 442–451.

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., & Tramontano, A. (2016). Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins*, *84* (Suppl 1), 4–14.

Naithani, A., Taylor, P., Erman, B., & Walkinshaw, M. D. (2015). A molecular dynamics study of allosteric transitions in Leishmania mexicana pyruvate kinase. *Biophysical Journal*, *109*(6), 1149–1156.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reinhart, G. D. (1983). The determination of thermodynamic allosteric parameters of an enzyme undergoing steady-state turnover. *Archives of Biochemistry and Biophysics*, *224*(1), 389–401.

Reinhart, G. D. (1988). Linked-function origins of cooperativity in a symmetrical dimer. *Biophysical Chemistry*, *30*(2), 159–172.

Reinhart, G. D. (2004). Quantitative analysis and interpretation of allosteric behavior. *Methods in Enzymology*, *380*, 187–203.

Shin, W. H., Kim, J. K., Kim, D. S., & Seok, C. (2013). GalaxyDock2: Protein-ligand docking using beta-complex and global optimization. *Journal of Computational Chemistry*, *34*(30), 2647–2656.

Weber, G. (1972). Ligand binding and internal equilibria in proteins. *Biochemistry*, *11*(5), 864–878.

Wei, Q., & Dunbrack, R. L. Jr. (2013). The role of balanced training and testing data sets for binary classifiers in bioinformatics. *PLOS ONE*, *8*(7), e67863.

Weinkam, P., Pons, J., & Sali, A. (2012). Structure-based model of allostery predicts coupling between distant sites. *Proceedings of the National Academy of Sciences*, *109*(13), 4875–4880.

Williams, R., Holyoak, T., McDonald, G., Gui, C., & Fenton, A. W. (2006). Differentiating a ligand's chemical requirements for allosteric interactions from those for protein binding. Phenylalanine inhibition of pyruvate kinase. *Biochemistry*, *45*(17), 5421–5429.

Zhong, W., Morgan, H. P., McNae, I. W., Michels, P. A., Fothergill-Gilmore, L. A., & Walkinshaw, M. D. (2013). 'In crystallo' substrate binding triggers major domain movements and reveals magnesium as a co-activator of Trypanosoma brucei pyruvate kinase. *Acta Crystallographica Section D: Biological Crystallography*, *69*(Pt 9), 1768–1779.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Xu, Q, Tang, Q, Katsonis, P, Lichtarge, O, et al. Benchmarking predictions of allostery in liver pyruvate kinase in CAGI4. *Human Mutation*. 2017;38:1123–1131. https://doi.org/10.1002/humu.23222