Reprogramming protein kinase substrate specificity through synthetic mutations

SUPPORTING INFORMATION

Instructions to recreate manuscript pLogos:

pLogos (*11*) depict residues proportional to the log-odds of their binomial probabilities with respect to a given background. In a pLogo, the most statistically significant residues appear closest to the x-axis, with residues above the x-axis indicating overrepresentation and those below the x-axis indicating underrepresentation. Given the existence of one or more different residues at a given substrate position, it is possible to compute conditional probabilities of all remaining amino acids and positions to determine significant positions *given* specific residues at specific positions. We refer to this as "fixing" a given residue at a given position, which allows the exploration of correlated or uncorrelated residues across positions in the kinase specificity motif. Fixed positions within the pLogo (e.g., the central position) are depicted on a grey background, and red horizontal lines denote the $p = 0.05$ significance threshold (after Bonferroni correction). pLogos can be scaled for clarity. For each pLogo, peptide sequences centered upon the phosphoacceptor P-site residue used as the foreground data were obtained by mapping phosphorylated tryptic peptide sequences (identified by tandem mass spectrometry) onto the *E. coli* proteome to retrieve the *in vivo* 15-residue context of the phosphoacceptor, flanked by its neighboring 7 residues in either direction (this is the same procedure as in *motif-x* analyses (*20*)). The *E. coli* background data set was generated through alignment of all unique serine- or threonine-centered 15-mers in the *E. coli* proteome.

For the *in silico* differential analysis with mutant kinases, the foreground is constructed in the same way, using tryptic peptides from mutant kinase (DYRK1A^{Q323G}, DYRK1A^{R328K},

DYRK1A^{QR-GK}) experiments. However, rather than using the *E. coli* proteome to construct a background data set, instead, the tryptic peptides from DYRK1A^{WT} were used, to determine the background frequencies for the corresponding mutant kinase. This allows one to analyze statistically different residues and positions in the kinase specificity motif resulting directly from the mutational changes.

Instructions to recreate manuscript pLogos

In order to fully explore the data in an interactive manner, readers may use the provided data to recreate any of the pLogos. Below are basic instructions:

1) Access the pLogo website (and register for an account if desired): plogo.uconn.edu
2) Paste desired foreground data set into the box on the left of the page. For convenience, the aligned 15mers (with negative controls subtracted) for each kinase are provided in Table S2. Raw identified peptides are also provided in Table S1.
3) Select "Protein" and then "e. coli k12" from the available backgrounds on the right of the page.
4) [OPTIONAL] If logged into a personal account, the user may add a job name.
5) The user then clicks the "generate pLogo" button in the center of the page.
6) Residues can be fixed or unfixed by clicking on them; however, residues that do not achieve statistical significance may not be fixed. Alternatively, users may fix significant residues by checking the corresponding box in the "statistics" tab to the left of the pLogo.
7) The zoom can be changed by clicking the "customize" tab, and either hitting the +/- buttons, or entering a value. Clicking "renormalize" will rescale the pLogo back to its default size.
8) For additional functionality and explanations, see:
   O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, & Schwartz D. (2013). pLogo: A probabilistic approach to visualizing sequence motifs. *Nat Methods* 10, 1211-1212.

SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. CK II pLogos. (**A** to **B**) pLogos (as described in Fig. 2) illustrate substrate preferences for (**A**) serine-centered and (**B**) threonine-centered CK II substrates, constructed from known literature-curated substrates (*12*).

Fig. S2. DYRK1A variants western blot. Western blot shows (**1**) empty pET45b vector negative control, and robust expression of (**2**) DYRK1A^{KD}, (**3**) DYRK1A^{WT}, (**4**) DYRK1A^{Q323G}, (**5**) DYRK1A^{R328K}, and (**6**) DYRK1A$^{QR\text{-}GK}$ at the expected molecular weight, using α-6XHis antibody.

Fig. S3. DYRK1A Pro-Q Diamond and Coomassie stained gels. (**A**) SDS-PAGE with Pro-Q Diamond staining reveals robust autophosphorylation and efficient phosphorylation of bacterial substrates over a wide molecular weight range for (**3**) DYRK1A^{WT}, (**4**) DYRK1A^{Q323G}, (**5**) DYRK1A^{R328K}, and (**6**) DYRK1A$^{QR\text{-}GK}$ constructs, relative to (**1**) empty vector pET45b or (**2**) DYRK1A^{KD} negative controls. (**B**) Coomassie staining provides loading control.

Fig. S4. Additional DYRK1A^{WT} pLogos. (**A** to **C**) Substrate preferences for DYRK1A^{WT} visualized with pLogo (as described in Fig. 2). Threonine-centered pLogos constructed from either (**A**) known literature-curated substrates (*12*) or from (**B**) unbiased ProPeL experiments. (**C**) Tyrosine-centered pLogo constructed from ProPeL experiments. (**D** to **E**) pLogos for subset of DYRK1A^{WT} data, using three MS/MS runs for (**D**) serine-centered and (**E**) threonine-centered substrates shows agreement with the full data set (above (**B**) and Fig. 2).

Fig. S5. DYRK1A^{WT} pLogos with additional fixed positions. Substrate preferences for DYRK1A^{WT} visualized with pLogo (as described in Fig. 2), with variable fixed positions. Fixing

of additional positions allows for conditional probabilities to be calculated (see main text discussion). (**A**) S-fixed (reproduced from Fig. 2B for ease of comparison), (**B**) SP-fixed, (**C** to **F**) individual fixed upstream arginine (note multiple upstream arginines are not correlated in the motif), (**G**) SV-fixed, and (**H**) SA-fixed.

Fig. S6. DYRK1A^{WT} pLogo accurately discriminates substrates. Average position weight matrix (PWM) score using the DYRK1A^{WT} pLogo based on ProPeL data to score either known human DYRK1A substrates, equivalent random substrates selected from the human proteome, or known human substrates (*12*) of representative model kinases. Error bars represent 95% confidence intervals.

Fig. S7. Additional mutant DYRK1A pLogos. (**A** to **C**) pLogos (as described in Fig. 2) illustrate threonine-centered substrate preferences for (**A**) DYRK1A^{Q323G}, (**B**) DYRK1A^{R328K}, and (**C**) DYRK1A$^{QR\text{-}GK}$. (**D** to **F**) Differential pLogos display the relative changes in substrate specificity between mutant DYRK1A and DYRK1A^{WT} by using DYRK1A^{WT} as a background data set, and mutant foreground data sets were used for (**D**) DYRK1A^{Q323G}, (**E**) DYRK1A^{R328K}, and (**F**) DYRK1A$^{QR\text{-}GK}$.

Figure S1

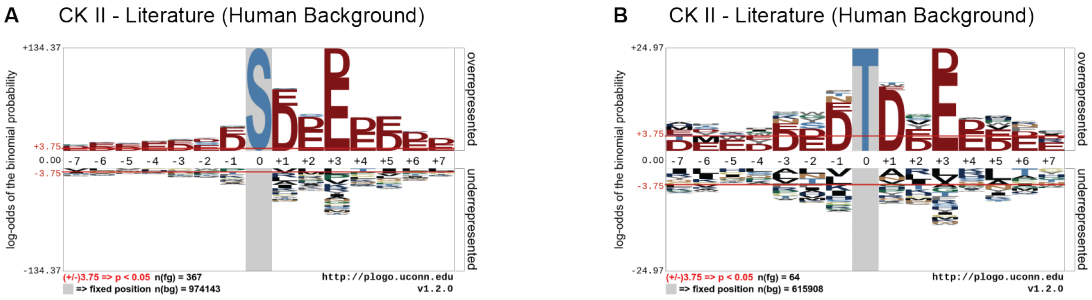
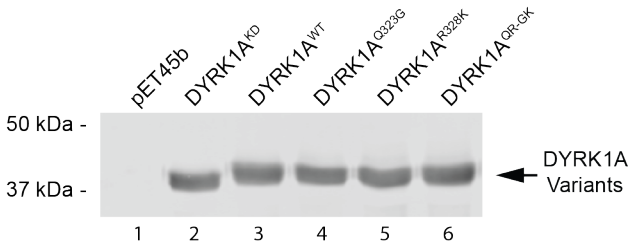


Fig. S1. CK II pLogos. (**A** to **B**) pLogos (as described in Fig. 2) illustrate substrate preferences for (**A**) serine-centered and (**B**) threonine-centered CK II substrates, constructed from known literature-curated substrates (*12*).

Figure S2



Western Blot

Figure S3

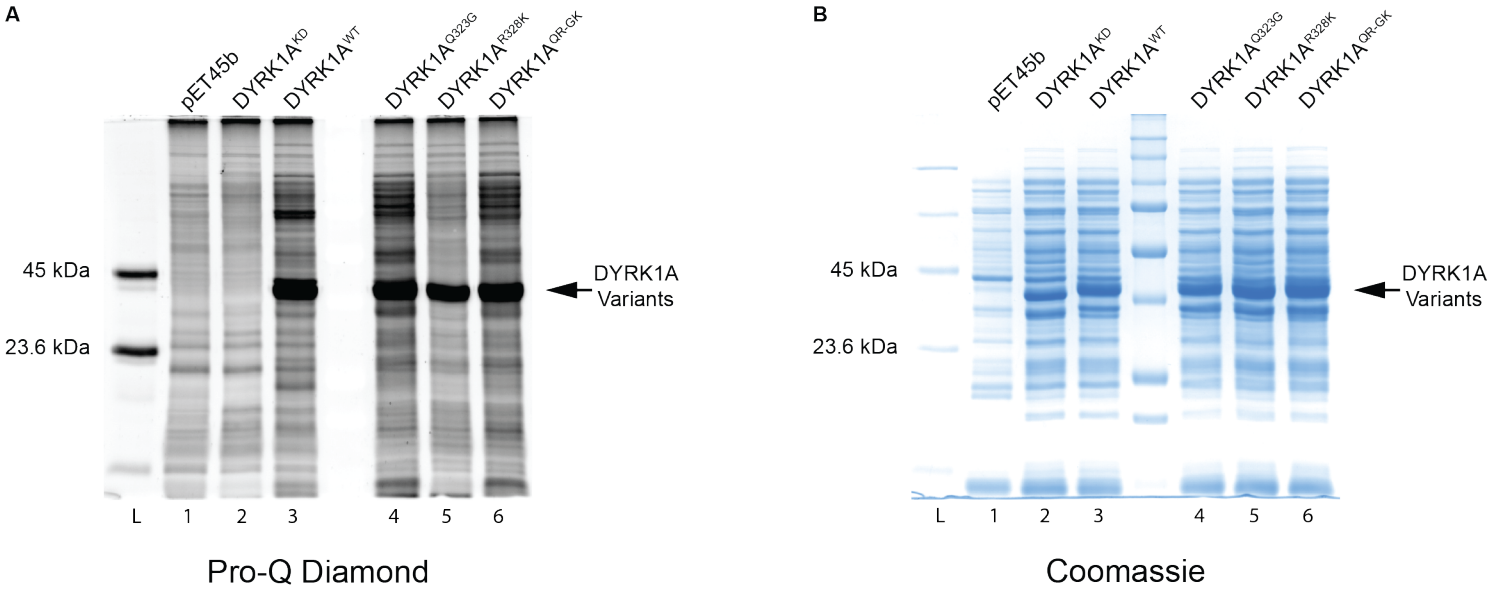
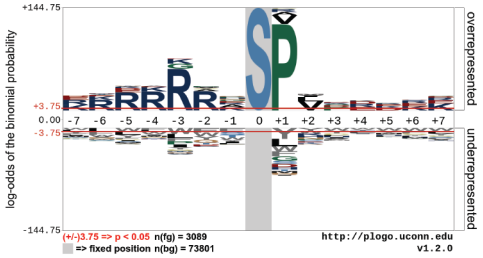
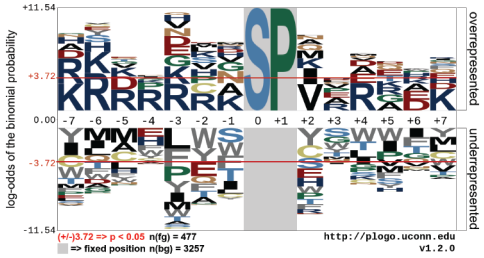

A
pET45b
DYRK1A^KD
DYRK1A^WT
DYRK1A^Q323G
DYRK1A^R328K
DYRK1A^QR-GK
45 kDa
23.6 kDa
DYRK1A Variants
L 1 2 3 4 5 6
Pro-Q Diamond
B
pET45b
DYRK1A^KD
DYRK1A^WT
DYRK1A^Q323G
DYRK1A^R328K
DYRK1A^QR-GK
45 kDa
23.6 kDa
DYRK1A Variants
L 1 2 3 4 5 6
Coomassie

Figure S4

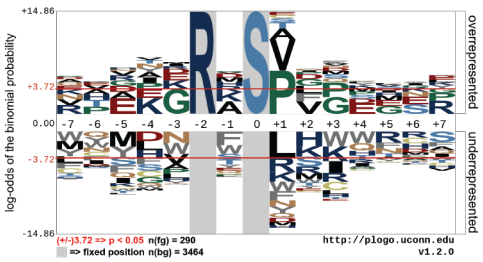
A DYRK1A^{WT} S-fixed (*E. coli* Background)



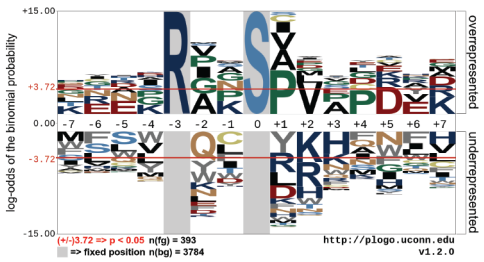
B DYRK1A^{WT} SP-fixed (*E. coli* Background)



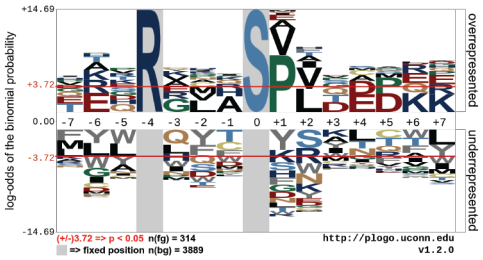
C DYRK1A^{WT} RxS-fixed (*E. coli* Background)



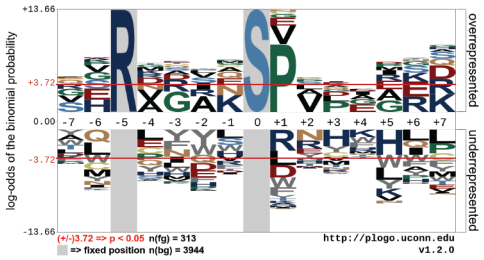
D DYRK1A^{WT} RxxS-fixed (*E. coli* Background)



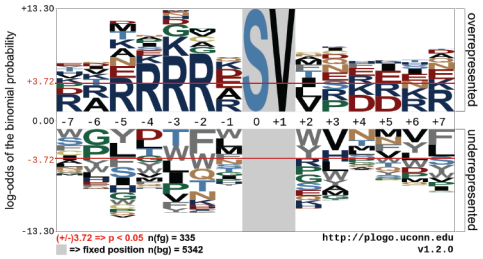
E DYRK1A^{WT} RxxxS-fixed (*E. coli* Background)



F DYRK1A^{WT} RxxxxS-fixed (*E. coli* Background)



G DYRK1A^{WT} SV-fixed (*E. coli* Background)



H DYRK1A^{WT} SA-fixed (*E. coli* Background)

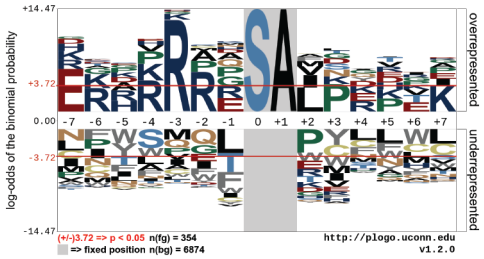
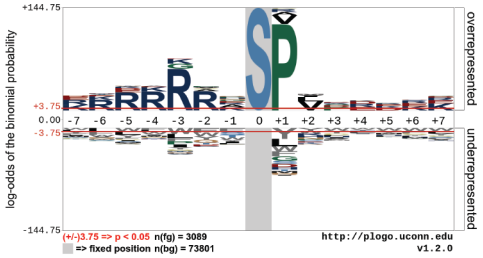
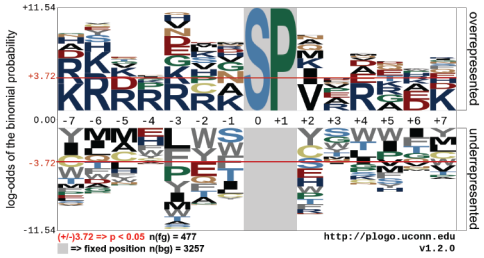


Figure S5

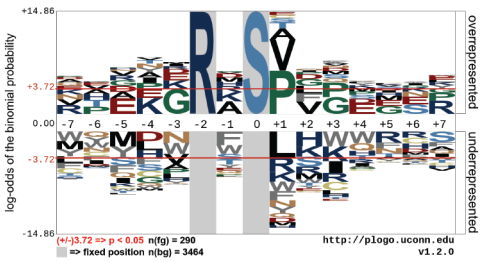
A DYRK1A^{WT} S-fixed (*E. coli* Background)



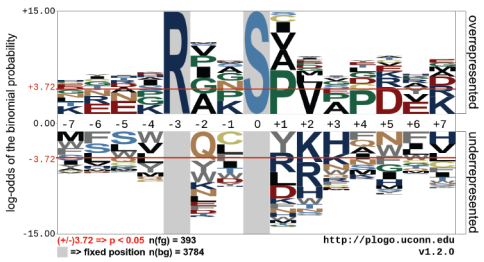
B DYRK1A^{WT} SP-fixed (*E. coli* Background)



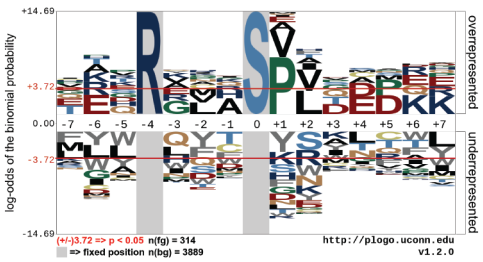
C DYRK1A^{WT} RxS-fixed (*E. coli* Background)



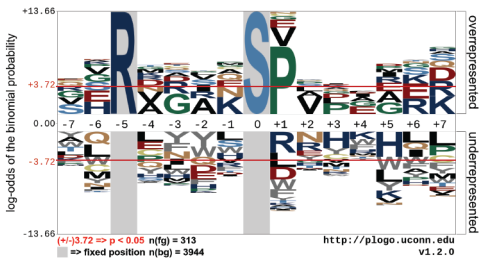
D DYRK1A^{WT} RxxS-fixed (*E. coli* Background)



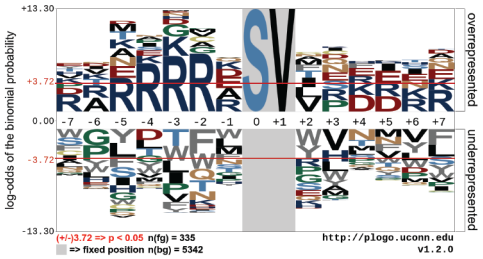
E DYRK1A^{WT} RxxxS-fixed (*E. coli* Background)



F DYRK1A^{WT} RxxxxS-fixed (*E. coli* Background)



G DYRK1A^{WT} SV-fixed (*E. coli* Background)



H DYRK1A^{WT} SA-fixed (*E. coli* Background)

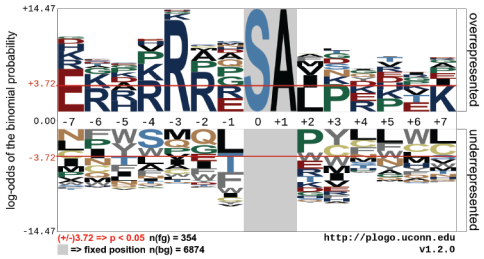


Figure S6

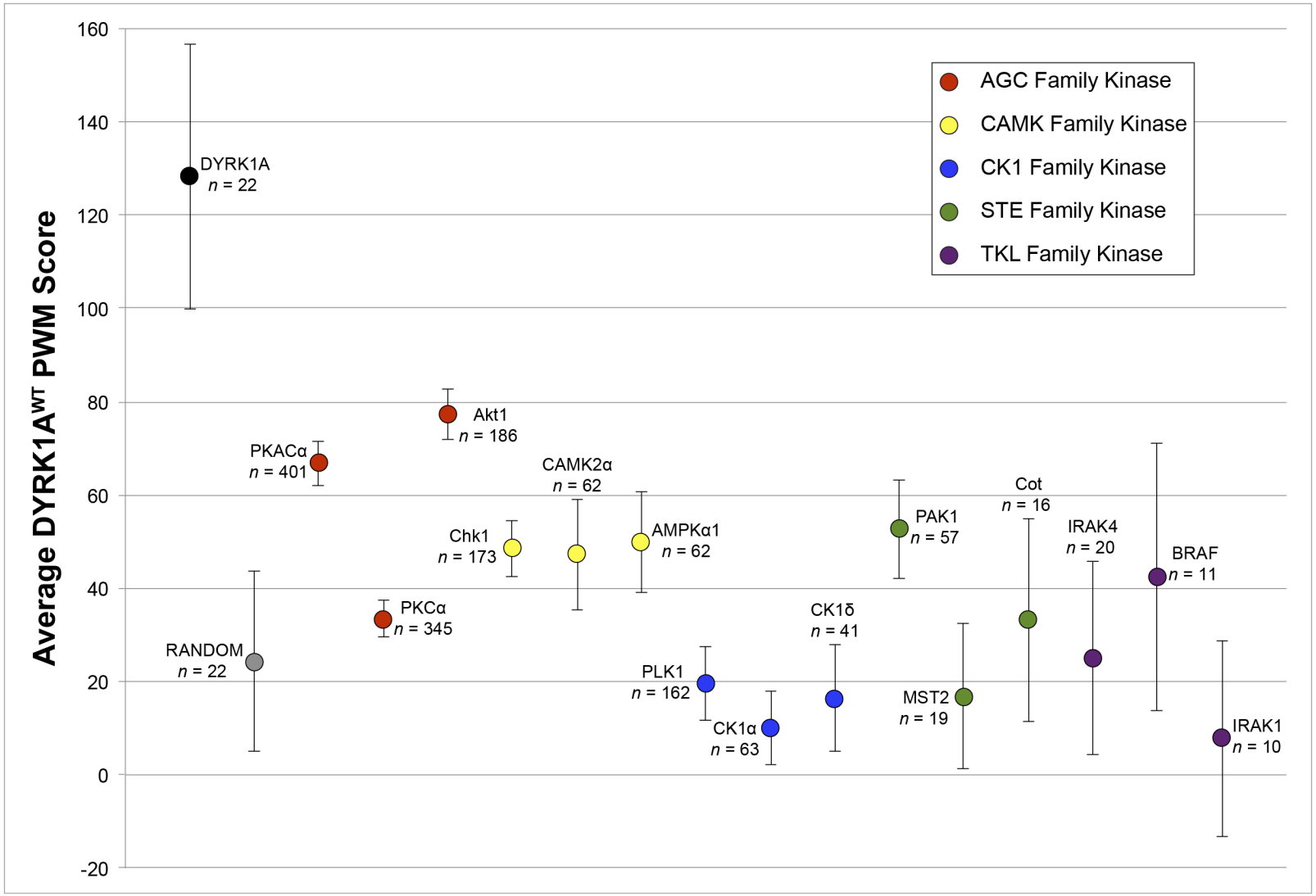
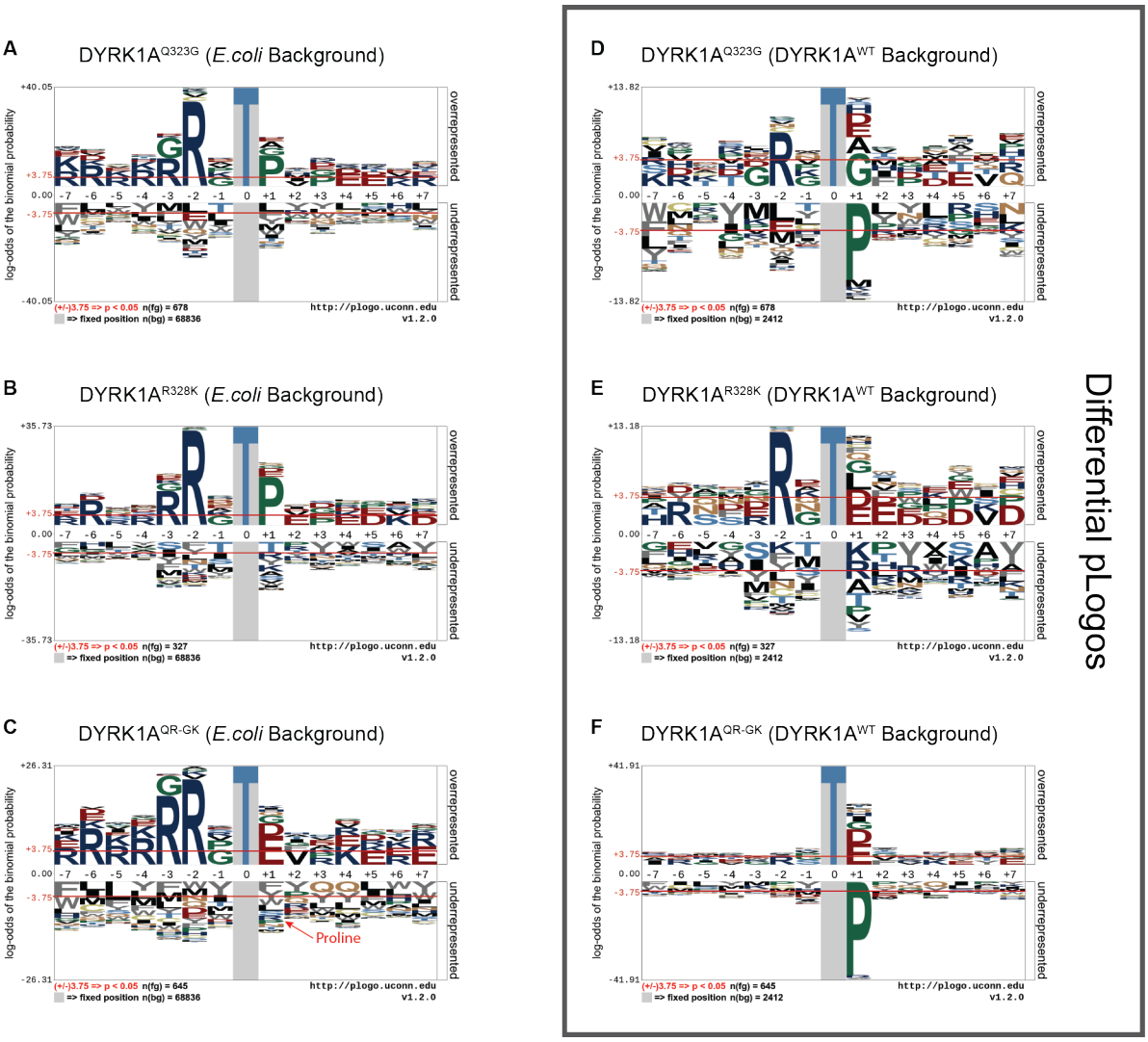

Average DYRK1A^WT PWM Score
160
140
120
100
80
60
40
20
0
-20
AGC Family Kinase
CAMK Family Kinase
CK1 Family Kinase
STE Family Kinase
TKL Family Kinase
DYRK1A
n = 22
PKACα
n = 401
Akt1
n = 186
CAMK2α
n = 62
Chk1
n = 173
AMPKα1
n = 62
PAK1
n = 57
Cot
n = 16
IRAK4
n = 20
BRAF
n = 11
RANDOM
n = 22
PKCα
n = 345
PLK1
n = 162
CK1δ
n = 41
MST2
n = 19
CK1α
n = 63
IRAK1
n = 10

Figure S7


A
DYRK1A^Q323G (E.coli Background)
B
DYRK1A^R328K (E.coli Background)
C
DYRK1A^QR-GK (E.coli Background)
Proline
D
DYRK1A^Q323G (DYRK1A^WT Background)
E
DYRK1A^R328K (DYRK1A^WT Background)
F
DYRK1A^QR-GK (DYRK1A^WT Background)
Differential pLogos
log-odds of the binomial probability
overrepresented
underrepresented
http://plogo.uconn.edu
v1.2.0