Human Genetics
https://doi.org/10.1007/s00439-024-02724-8

ORIGINAL INVESTIGATION

Assessing the predicted impact of single amino acid substitutions in MAPK proteins for CAGI6 challenges

Paola Turina[1] · Maria Petrosino[2] · Carlos A. Enriquez Sandoval[1] · Leonore Novak[2] · Alessandra Pasquo[3] · Emil Alexov[4] · Muttaqi Ahmad Alladin[5] · David B. Ascher[6,7] · Giulia Babbi[1] · Constantina Bakolitsa[8] · Rita Casadio[1] · Jianlin Cheng[9] · Piero Fariselli[10] · Lukas Folkman[11] · Akash Kamandula[12] · Panagiotis Katsonis[13] · Minghui Li[14] · Dong Li[15] · Olivier Lichtarge[13] · Sajid Mahmud[9] · Pier Luigi Martelli[1] · Debnath Pal[5] · Shailesh Kumar Panday[4] · Douglas E. V. Pires[16] · Stephanie Portelli[6,7] · Fabrizio Pucci[15] · Carlos H. M. Rodrigues[6] · Marianne Rooman[15] · Castrense Savojardo[1] · Martin Schwersensky[15] · Yang Shen[17] · Alexey V. Strokach[18] · Yuanfei Sun[17] · Junwoo Woo[19] · Predrag Radivojac[12] · Steven E. Brenner[8,20,21] · Roberta Chiaraluce[2] · Valerio Consalvi[2] · Emidio Capriotti[1,22]

Received: 30 May 2024 / Accepted: 27 December 2024

Abstract

New thermodynamic and functional studies have been recently conducted to evaluate the impact of amino acid substitutions on the Mitogen Activated Protein Kinases 1 and 3 (MAPK1/3). The Critical Assessment of Genome Interpretation (CAGI) data provider, at Sapienza University of Rome, measured the unfolding free energy and the enzymatic activity of a set of variants (MAPK challenge dataset). Thermodynamic measurements for the denaturant-induced equilibrium unfolding of the phosphorylated and unphosphorylated forms of the MAPKs were obtained by monitoring the far-UV circular dichroism and intrinsic fluorescence changes as a function of denaturant concentration. These values have been used to calculate the change in unfolding free energy between the variant and wild-type proteins at zero concentration of denaturant ($\Delta\Delta G^{H_2O}$). The enzymatic activity of the phosphorylated MAPKs variants was also measured using Chelation-Enhanced Fluorescence to monitor the phosphorylation of a peptide substrate. The MAPK challenge dataset, composed of a total of 23 single amino acid substitutions (11 and 12 for MAPK1 and MAPK3, respectively), was used to assess the effectiveness of the computational methods in predicting the $\Delta\Delta G^{H_2O}$ values, associated with the variants, and categorize them as destabilizing and not destabilizing. The data on the enzymatic activity of the MAPKs mutants were used to assess the performance of the methods for predicting the functional impact of the variants. For the sixth edition of CAGI, thirteen independent research groups from four continents (Asia, Australia, Europe and North America) submitted > 80 sets of predictions, obtained from different approaches. In this manuscript, we summarized the results of our assessment to highlight the possible limitations of the available algorithms.

Introduction

The human kinome is composed of over five hundred different protein kinases (Manning et al. 2002), making it one of the largest gene families in eukaryotes. Protein kinases play a crucial role in various cell signaling processes and are implicated in numerous human diseases (Metz et al. 2018) and as a key drug target class in the oncology area and beyond (Attwood et al. 2021). Within elthe kinase tree, Mitogen Activated Protein Kinases 1 and 3 (MAPK1 and MAPK3), also known as extracellular regulated kinases (ERK2 and ERK1), regulate a variety of cellular processes and participate extensively in the control of cell-fate decisions (Varjosalo et al. 2013).

The MAPK signaling cascade is a central pathway involved in transmitting extracellular signals through sequential phosphorylation and activation of downstream kinases, which regulate and control many fundamental cellular processes (Lavoie et al. 2020). The dysregulation of such pathway is associated with various pathological conditions (Kim and Choi 2010), including cancer (Roskoski 2019), neurodegenerative diseases (Khezri et al. 2023), autoimmune diseases (Liu et al. 2021), and diabetes (Zhang et al. 2021).

Paola Turina, Maria Petrosino are co-first authors.

Extended author information available on the last page of the article

Published online: 20 February 2025

MAPK1 and MAPK3 are serine/threonine kinases activated downstream in the Ras/Raf/MEK/ERK signaling cascade. They share a high degree of sequence identity and similarity and have similar domain architecture. While they are often co-expressed and functionally redundant, recent studies suggest some functional differences between the two isoforms. MAPK1 and MAPK3 exhibit distinct conformational mobility upon activation (Ring et al. 2011) and possess differential stability and nuclear envelope crossing capabilities (Marchi et al. 2008). Additionally, MAPK3 is more resistant to the turnover induced by MAPK inhibitors compared to MAPK1 (Balmanno et al. 2023).

The primary sequence of MAPK1 and MAPK3 is prone to somatic missense mutations, particularly in cancer tissues. Investigating the biochemical and biophysical properties of wild-type and variant MAPK1 and MAPK3 is crucial for understanding the consequences of these cancer-associated mutations. Overall, understanding the functional characteristics, stability, and mutations in MAPK1 and MAPK3 may also help in elucidating their roles in cellular processes, diseases, and potential therapeutic strategies.

However, comprehensive studies on the effects of mutations on the biochemical and structural properties of MAPK1/3 have not been conducted so far on the purified proteins. The 23 mutations analyzed in this study were selected from the COSMIC database (Sondka et al. 2024) prioritizing those frequently observed in cancer tissues and those expected to alter the physical and chemical properties of the proteins due to changes in amino acid residues. The variants were distributed across the entire sequences of the two kinases (Fig. 1) to ensure comprehensive coverage of their primary structures. Particular attention was given to selecting mutations located in analogous positions within the closely identical architecture of the two proteins, allowing for direct comparisons between them (Petrosino et al. 2023).

Over the last decades, many methods for predicting the impact of amino acid substitutions on protein stability have been developed (Marabotti et al. 2020). In general, these tools, which predict the variation of folding free energy change upon mutation, implement different approaches including empirical energy functions and machine learning algorithms (Compiani and Capriotti 2013). Although predicting the impact of variants on protein stability can be a relevant approach for characterizing the genotype–phenotype relationships (Petrosino et al. 2021), the assessment and standardization of such tools is still a challenging task (Sanavia et al. 2020; Pancotti et al. 2022).

To overcome the above limitations, in the previous edition of the Critical Assessment of Genome Interpretation (CAGI), we introduced a challenge focused on predicting the

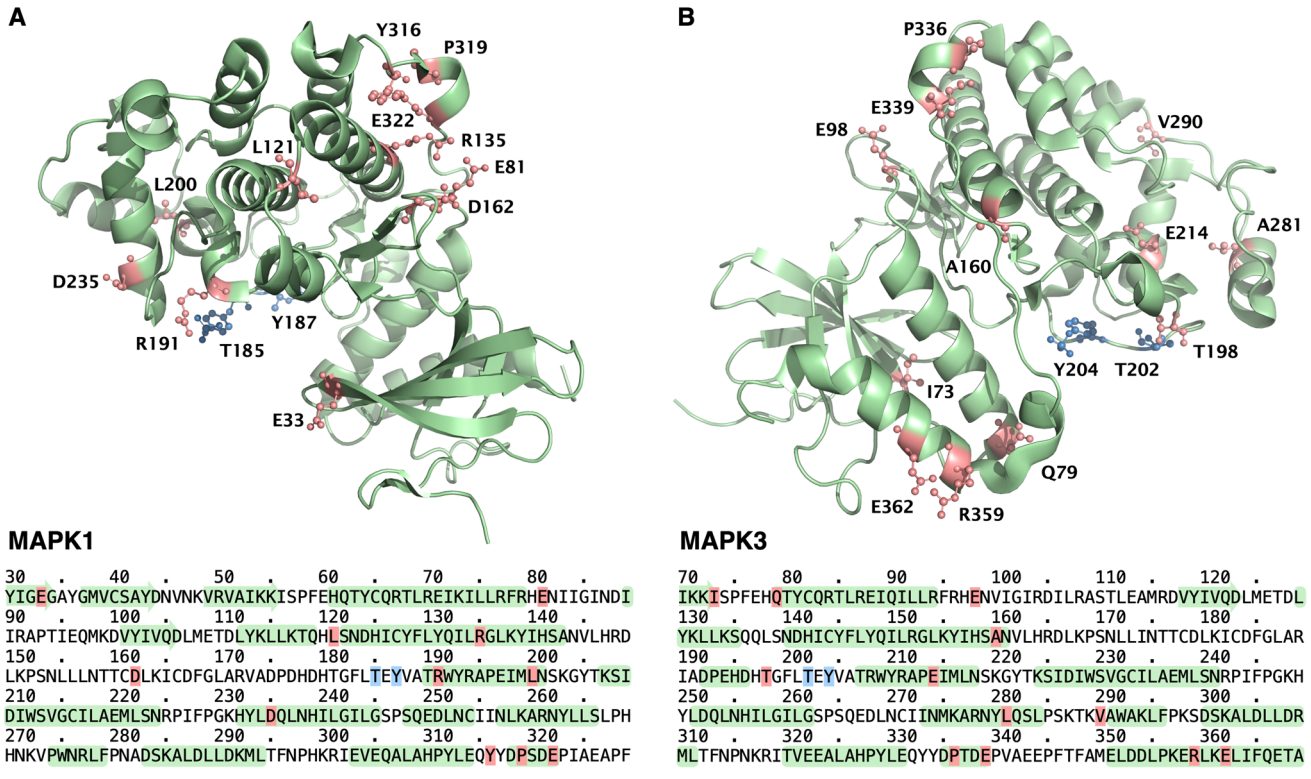

Fig. 1 Mapping of the 23 mutated sites of the MAPK challenges datasets (red) on the three-dimensional structure of the phosphorylated forms of MAPK1 (PDB: 5v60) and MAPK3 (PDB: 2zoq) shown in panels A and B, respectively. Residues in blue indicate the phosphorylated sites

impact of eight single amino acid variants on the stability of the human frataxin protein (Savojardo et al. 2019; Petrosino et al. 2019; Critical Assessment of Genome Interpretation Consortium 2024). As a follow-up to prior efforts, in the sixth edition of the CAGI experiments, we presented a more extensive challenge, aiming at evaluating the performance of computational methods in predicting the measured values of stability change for the 23 single amino acid variants in MAPK proteins, as well as their impact on the catalytic efficiency.

Materials and methods

Experimental measures

MAPK1 and MAPK3 proteins, along with their variants, were expressed in *E. coli* cells using N-terminally His-tagged constructs, as described (Petrosino et al. 2021). Equilibrium unfolding of both wild-type and variant MAPK1 and MAPK3 proteins was achieved by incubating them at increasing concentrations of guanidinium chloride (GdmCl) at 10 °C, and intrinsic fluorescence emission spectra and circular dichroism (CD) spectra were recorded for all proteins as described (Petrosino et al. 2021). Melting temperatures (T_m) were determined from the first derivative of the ellipticity changes at 222 nm. The analysis of spectral changes in far-UV CD ellipticity and intrinsic fluorescence emission involved fitting to the data different models, describing either a 2-state or 3-state unfolding process, depending on the detection of an intermediate state. Such models were used to extrapolate the unfolding free energy change at zero denaturant concentration (ΔG^{H_2O}) for each MAPK protein and variants.

Finally, the catalytic activity of phosphorylated MAPK1 and MAPK3 proteins, as well as their variants, was assessed using a fluorescence-based method with a substrate peptide. Kinetic studies were conducted to analyze enzyme activity in relation to substrate concentration and temperature. The temperature dependence of catalytic activity was investigated across a range of temperatures, between 10 and 45 °C. More detailed information on the experimental techniques and data analysis procedures is provided in (Petrosino et al. 2023).

MAPK challenge datasets

In the sixth CAGI edition, two challenges on MAPK1 and MAPK3 proteins were organized. The MAPK1 and MAPK3 challenge datasets consist of 11 and 12 coding variants, respectively, which were selected from the COSMIC database (Sondka et al. 2024). A representation of the mutated sites in the three-dimensional structures of the MAPK1 (PDB: 5V60) and MAPK3 (PDB: 2ZOQ) in the phosphorylated forms are displayed in Fig. 1.

To evaluate the impact of the variants on protein stability, the ΔG^{H_2O} difference between the variant and wild-type proteins ($\Delta\Delta G^{H_2O}$) was computed using the following equation:

$$\Delta G^{H_2O}_{N\to U}(mut) = \Delta G^{H_2O}_{N\to U}(mut) - \Delta G^{H_2O}_{N\to U}(wt). \quad (1)$$

The average experimental values of the $\Delta\Delta G^{H_2O}$ obtained by circular dichroism and fluorescence were used for the challenge.

The simplest protein unfolding process involves a 2-state transition, where the protein denatures directly from its native (N) to the unfolded (U) state. However, for many MAPK variants, at low concentration of denaturant, the equilibrium unfolding titration reveals the formation of an intermediate state (I). In such cases, the total unfolding ΔG^{H_2O} is calculated by summing the contributions of the two transitions:

$$\Delta G^{H_2O}_{N\to U} = \Delta G^{H_2O}_{N\to I} + \Delta G^{H_2O}_{I\to U}. \quad (2)$$

Given the difference between the folding mechanism of the MAPK variants, when possible, we fitted to the data of MAPK1 (Table S1) and MAPK3 (Table S2) both a 2-state and a 3-state unfolding model. To harmonize the experimental ΔG^{H_2O} values, utilized for each MAPK challenge, we considered 2-state and 3-state fitting models for MAPK1 and MAPK3, respectively. A comparison of the ΔG^{H_2O} values for the unphosphorylated and phosphorylated forms of MAPKs proteins is displayed in Fig. S1. The final sets of variants with the relative averages of $\Delta\Delta G^{H_2O}$, obtained by different experimental techniques (circular dichroism and fluorescence), and their experimental errors, for both the phosphorylated and the unphosphorylated forms, are reported in Table 1 (MAPK1) and Table 2 (MAPK3). A comparison of the $\Delta\Delta G^{H_2O}$ values for the phosphorylated and unphosphorylated forms of MAPK proteins is shown in Fig. 2.

To assess the performance of the teams on a dataset composed of different proteins, we combined MAPK1 and MAPK3 variants based on their folding mechanisms. For this task, we only considered the subset of data supported by at least one optimal fit to a 2-state and/or 3-state unfolding mechanism (Table S3).

The catalytic efficiency of each MAPK protein and variant was assessed using the k_{cat}/K_M ratio, where k_{cat} represents the rate at which substrate molecules are converted into products per unit time by a single enzyme molecule, and K_M is the Michaelis–Menten constant. The functional impact of the single amino acid substitution was estimated by dividing the catalytic efficiency of the mutant by that of

Table 1 Thermodynamic stability of MAPK1 protein and variants

Protein	Unphosphorylated		Phosphorylated	
	ΔG^{H_2O} (kcal/mol)	$\Delta\Delta G^{H_2O}$ (kcal/mol)	ΔG^{H_2O} (kcal/mol)	$\Delta\Delta G^{H_2O}$ (kcal/mol)
Wild-type	2.65 ± 0.14	–	2.59 ± 0.13	–
p.E33Q	2.89 ± 0.17	0.24 ± 0.22	2.27 ± 0.13	−0.32 ± 0.18
p.E81K	1.95 ± 0.15	−0.70 ± 0.21	2.45 ± 0.17	−0.14 ± 0.21
p.L121I	2.78 ± 0.17	0.13 ± 0.22	2.28 ± 0.12	−0.31 ± 0.18
p.R135K	2.50 ± 0.15	−0.15 ± 0.21	2.24 ± 0.12	−0.35 ± 0.18
p.D162G	2.48 ± 0.10	−0.17 ± 0.17	2.13 ± 0.10	−0.46 ± 0.16
p.R191H	3.64 ± 0.54	0.99 ± 0.56	2.43 ± 0.13	−0.16 ± 0.18
p.L200F	2.58 ± 0.11	−0.07 ± 0.18	2.61 ± 0.18	0.02 ± 0.22
p.D235V	2.47 ± 0.12	−0.18 ± 0.18	2.09 ± 0.13	−0.50 ± 0.18
p.Y316F	2.66 ± 0.17	0.01 ± 0.22	1.92 ± 0.11	−0.67 ± 0.17
p.P319S	2.20 ± 0.13	−0.45 ± 0.19	2.25 ± 0.15	−0.34 ± 0.20
p.E322V	2.14 ± 0.14	−0.51 ± 0.20	2.14 ± 0.12	−0.45 ± 0.18

The average unfolding free energy change at zero denaturant concentration (ΔG^{H_2O}) for the phosphorylated and unphosphorylated MAPK1 is calculated as the mean ΔG^{H_2O} values of fluorescence and circular dichroism experiments (see Table S1). The $\Delta\Delta G^{H_2O}$ relative to each variant is determined by Eq. 1.

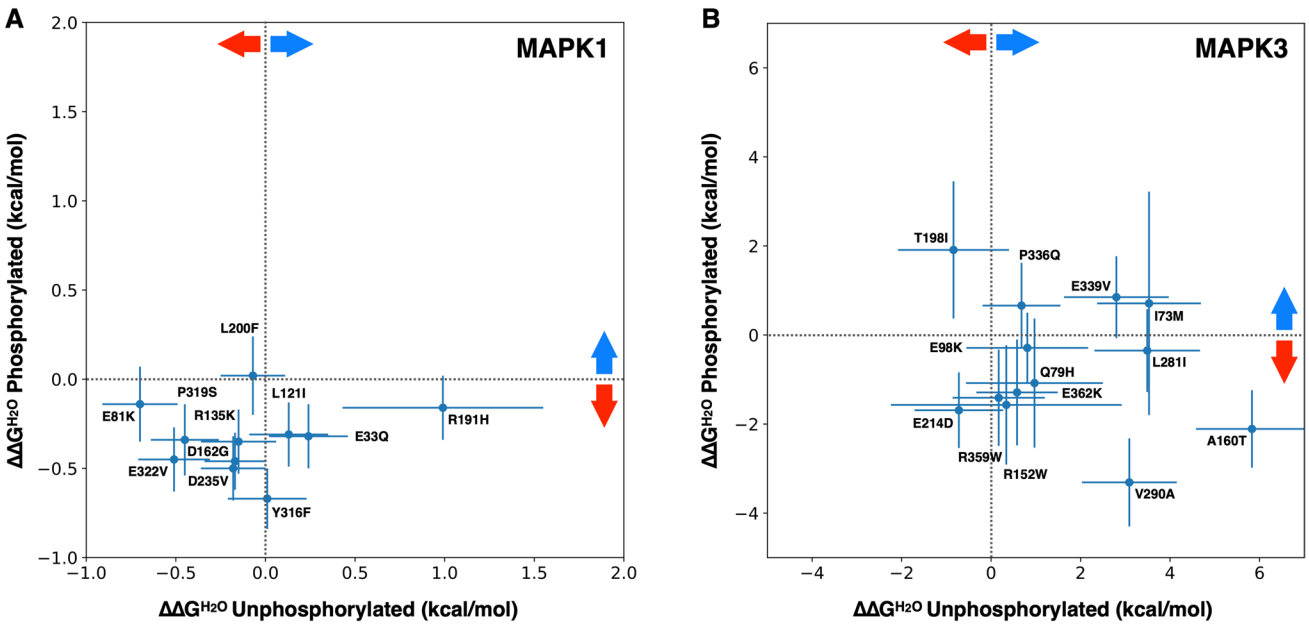

Fig. 2 Scatter plots of the experimental unfolding $\Delta\Delta G^{H_2O}$ values for the phosphorylated and unphosphorylated forms of MAPK1 (**A**) and MAPK3 (**B**) variants. Variants with $\Delta\Delta G^{H_2O}$ values above or on the right side of the dashed lines (indicated by blue arrows) are stabilizing. Destabilizing variants are below or on the left side of the dashed lines (indicated by red arrows)

the wild-type protein. The values of the catalytic efficiencies and their mutant-to-wildtype ratios are summarized in Table 3 for both MAPK1 and MAPK3.

Challenge participants and prediction methods

Thirteen groups contributed to the CAGI6 MAPK challenges, collectively submitting over 80 sets of predictions, 43 for MAPK1 and 40 for MAPK3. Here, we provide a brief description of the prediction sets submitted by the participating teams.

Team 1: The *3 billion* group (South Korea), submitted 2 sets of predictions, both for MAPK1 and MAPK3, based on the 3Cnet algorithm (Won et al. 2021). The two versions of the method take as input either protein sequence features alone or combined with SNVBox database (Wong et al. 2011) and are trained on the VariBench

Table 2 Thermodynamic stability of MAPK3 protein and variants

Protein	Unphosphorylated		Phosphorylated	
	ΔG^{H_2O} (kcal/mol)	$\Delta\Delta G^{H_2O}$ (kcal/mol)	ΔG^{H_2O} (kcal/mol)	$\Delta\Delta G^{H_2O}$ (kcal/mol)
Wild-type	6.10 ± 0.72	–	7.79 ± 0.76	–
p.I73M	9.63 ± 0.91	3.53 ± 1.16	8.50 ± 2.39	0.71 ± 2.51
p.Q79H	7.07 ± 1.35	0.97 ± 1.53	6.71 ± 1.23	−1.08 ± 1.45
p.E98K	6.91 ± 1.15	0.81 ± 1.36	7.50 ± 0.21	−0.29 ± 0.79
p.R152W[a]	6.44 ± 2.48	0.34 ± 2.58	6.22 ± 1.10	−1.57 ± 1.34
p.A160T	11.93 ± 1.02	5.83 ± 1.25	5.68 ± 0.42	−2.11 ± 0.87
p.T198I	5.26 ± 1.01	−0.84 ± 1.24	9.70 ± 1.34	1.91 ± 1.54
p.E214D	5.38 ± 0.68	−0.72 ± 0.99	6.10 ± 0.38	−1.69 ± 0.85
p.L281I	9.59 ± 0.93	3.49 ± 1.18	7.44 ± 0.54	−0.35 ± 0.93
p.V290A	9.19 ± 0.78	3.09 ± 1.06	4.48 ± 0.63	−3.31 ± 0.99
p.P336Q	6.78 ± 0.48	0.68 ± 0.87	8.45 ± 0.59	0.66 ± 0.96
p.E339V	8.90 ± 0.92	2.80 ± 1.17	8.64 ± 0.51	0.85 ± 0.92
p.R359W	6.27 ± 0.73	0.17 ± 1.03	6.38 ± 0.77	−1.41 ± 1.08
p.E362K	6.68 ± 0.56	0.58 ± 0.91	6.50 ± 0.91	−1.29 ± 1.19

The average unfolding free energy change at zero denaturant concentration (ΔG^{H_2O}) for the phosphorylated and unphosphorylated MAPK3 is calculated as the mean ΔG^{H_2O} values of fluorescence and circular dichroism experiments (see Table S2). The $\Delta\Delta G^{H_2O}$ relative to each variant is determined by Eq. 1

[a]The experimental data for the MAPK3 p.R152W variant was not included in the assessment, as it was released after the end of the challenge

(Sasidharan Nair and Vihinen 2013) datasets to predict the unfolding $\Delta\Delta G^{H_2O}$ values. Team 1 submitted the same predictions for phosphorylated and unphosphorylated MAPKs.

Team 2: The *AIBI-CAGI6* group at the Texas A&M University (USA) submitted six sets of predictions, both for MAPK1 and MAPK3. Team 2 pretrained Bidirectional Encoder Representations from Transformers (BERT) on different PFAM representative proteomes (Mistry et al. 2021). The prediction models were fine-tuned on MAPK target family sequences, as described in (Sun and Shen 2023). The natural log of the ratio between the probability of occurrence of the mutant and wild-type residues in the mutated site is used for calculating the $\Delta\Delta G^{H_2O}$ predictions (the same for phosphorylated and unphosphorylated MAPK), while the predicted ratio as such is employed for assessing the catalytic efficiency ratio between mutant and wild-type proteins.

Team 3: The *Alexov Lab*, at the Clemson University (USA), predicted the values of $\Delta\Delta G^{H_2O}$ for both MAPK challenges using the SAAFEC-SEQ method (Li et al. 2021), a gradient boosting algorithm integrating physicochemical properties, sequence features, and evolutionary information. The SAAFEC-SEQ is a sequence-based method and does not require a 3D structure of the protein. Additionally, it does not account for chemical modifications like phoshorylation as it does not alter the protein sequence, which is the only input for the method. Thus, identical predictions were submitted for the phosphorylated and unphosphorylated forms of MAPK1 and MAPK3 mutants.

Team 4: The *BioSig* group (Team 4) at the University of Melbourne (Australia) submitted six sets of $\Delta\Delta G^{H_2O}$ predictions for both MAPK challenges utilizing DUET (Pires et al. 2014a), ENCoM (Frappier et al. 2015), mCSM (Pires et al. 2014b), SDM (Pandurangan et al. 2017), DynaMut (Rodrigues et al. 2018) and DynaMut2 (Rodrigues et al. 2021).

Team 5: The *3BIO-B* group at the Université Libre de Bruxelles (Belgium) computed the $\Delta\Delta G^{H_2O}$ predictions for both the unphosphorylated and the phosphorylated forms of MAPK1 and MAPK3, using three structure-based models: PoPMuSiC (Dehouck et al. 2011), PoPMuSiCsym (Pucci et al. 2018) and a combination of PoPMuSiC and MAESTRO (Laimer et al. 2015). The predictions of the ratio between the catalytic efficiency of the mutant and wild-type proteins are obtained by rescaling the predictions of the SNPMuSiC score (Ancien et al. 2018).

Team 6: The *Li Lab* at the Soochow University (China), submitted six prediction sets for MAPK1 and three for MAPK3. The $\Delta\Delta G^{H_2O}$ values were predicted, using the PremPS algorithm (Chen et al. 2020). The predictions were computed using different structures for the unphosphorylated and the phosphorylated forms of MAPK1 and MAPK3, as well as alternative structural conformations obtained by molecular dynamics minimization.

Team 7: The *EASE-MM* group submitted one set of predictions for each MAPK protein generated using EASE-

Table 3 Catalytic efficiency (k_{cat}/K_M) of MAPK1 and MAPK3 proteins and variants

MAPK1 catalytic efficiency				
Protein	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} μM^{-1})	$r_{kcat/KM}$
Wild-type	1.81 ± 0.29	2.20E + 00	1.22E + 00 ± 1.95E−01	–
p.E33Q	3.90 ± 0.67	1.98E + 00	5.08E−01 ± 8.72E−02	4.18E−01 ± 9.80E−02
p.E81K	2.64 ± 0.72	1.24E + 00	4.70E−01 ± 1.28E−01	3.86E−01 ± 1.22E−01
p.L121I	7.12 ± 2.09	7.20E − 02	1.01E−02 ± 2.97E−03	8.32E−03 ± 2.78E−03
p.R135K	6.19 ± 1.90	28.30	4.57E + 00 ± 1.41E + 00	3.76E + 00 ± 1.30E + 00
p.D162G	1.43 ± 0.47	9.30E−04	6.50E−04 ± 2.14E−04	5.35E−04 ± 1.96E−04
p.R191H	55.00 ± 0.02	2.00E−02	3.64E−04 ± 1.32E−07	2.99E−04 ± 4.80E−05
p.L200F	3.23 ± 0.40	2.43E−02	7.52E−03 ± 9.32E−04	6.19E−03 ± 1.25E−03
p.D235V	3.97 ± 0.67	3.10E−01	7.81E−02 ± 1.32E−02	6.42E−02 ± 1.49E−02
p.Y316F	1.57 ± 0.30	6.83E + 00	4.35E + 00 ± 8.31E−01	3.58E + 00 ± 8.90E−01
p.P319S	4.26 ± 1.11	3.62E−01	8.50E−02 ± 2.21E−02	6.99E − 02 ± 2.10E−02
p.E322V	2.81 ± 0.68	3.80E + 00	1.35E + 00 ± 3.27E−01	1.11E + 00 ± 3.23E−01
MAPK3 catalytic efficiency				
Protein	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} μM^{-1})	$r_{kcat/KM}$
Wild-type	3.17 ± 0.59	2.05	0.65 ± 0.12	–
p.I73M	5.30 ± 0.80	1.71	0.32 ± 0.05	0.50 ± 0.12
p.Q79H	3.31 ± 0.44	0.16	0.05 ± 0.01	0.07 ± 0.02
p.E98K	2.46 ± 0.43	2.65	1.08 ± 0.19	1.67 ± 0.43
p.R152W[a]	11.14 ± 1.80	1.80	0.16 ± 0.03	0.25 ± 0.06
p.A160T	3.41 ± 0.72	2.06	0.60 ± 0.13	0.93 ± 0.26
p.T198I	3.27 ± 0.45	10.74	3.28 ± 0.45	5.08 ± 1.18
p.E214D	8.51 ± 1.20	0.49	0.06 ± 0.01	0.09 ± 0.02
p.L281I	2.23 ± 0.37	0.93	0.42 ± 0.07	0.64 ± 0.16
p.V290A	14.49 ± 1.60	1.27	0.09 ± 0.01	0.14 ± 0.03
p.P336Q	3.52 ± 0.49	0.80	0.23 ± 0.03	0.35 ± 0.08
p.E339V	2.91 ± 0.59	1.07	0.37 ± 0.07	0.57 ± 0.16
p.R359W	3.17 ± 0.48	1.47	0.46 ± 0.07	0.72 ± 0.17

$r_{kcat/KM}$ is the ratio between the k_{cat}/K_M (mutant) and k_{cat}/K_M (wild-type)

[a]The experimental data for the MAPK3 p.R152W variant was not included in the assessment, as it was released after the end of the challenge

MM (Folkman et al. 2016). EASE-MM predicts protein stability changes based on protein sequence. The EASE-MM team submitted $\Delta\Delta G^{H_2O}$ values only for the unphosphorylated forms of MAPK1 and MAPK3.

Team 8: The *Bioinformatics and Machine Learning (BML) Laboratory* at the Missouri University (USA) submitted one set of $\Delta\Delta G^{H_2O}$ predictions, including both MAPK1 and MAPK3 in the unphosphorylated forms. These predictions were computed using graph convolutional neural networks trained on the ProThermDB database (Nikam et al. 2021), leveraging both sequence and 3D structure features.

Team 9: The *Lichtarge Lab* at the Baylor College of Medicine (USA), submitted six prediction sets, calculated using the Evolutionary Action (EA) method (Katsonis and Lichtarge 2014) and newer beta versions integrating multiple fitness predictors, alongside solvent accessibility calculations from PDB structures. It was assumed that protein stability correlates with solvent accessibility of the wild-type residue or its quadratic function. Additionally, it was estimated that the catalytic efficiency ratio is proportional to a quadratic function of the EA score. The six prediction sets were calculated by combining various prediction methods and structural features from different PDB structures. The method returned identical predictions for phosphorylated and unphosphorylated MAPKs.

Team 10: The *Biocomputing Group* at the University of Bologna (Italy) submitted 2 sets of $\Delta\Delta G^{H_2O}$ predictions for each MAPK protein. The $\Delta\Delta G^{H_2O}$ values were calculated combining the predictions of INPS3D (Savojardo et al. 2016), PoPMuSiC 2.1 (Dehouck et al. 2011), and FoldX (Guerois et al. 2002). Identical predictions were submitted for both the phosphorylated and the unphosphorylated forms of MAPK1 and MAPK3 mutants.

Table 4 Assessment of the best $\Delta\Delta G^{H_2O}$ predictions submitted by each team for phosphorylated and unphosphorylated MAPK3 variants

MAPK1 unphosphorylated										
Team	Model	r_P	r_S	r_{KT}	RMSE	MAE	BQ_2	MCC	AUC	<Rank>
BioSig	2	0.771	0.309	0.200	0.320	0.241	0.786	0.571	0.679	3.5
Strokach	2	0.232	0.245	0.127	0.507	0.418	0.750	0.624	0.750	5.0
Li Lab	3	0.574	0.591	0.418	0.752	0.640	0.571	0.239	0.786	5.3
3billion	1	0.217	0.345	0.200	0.465	0.352	0.429	−0.239	0.821	6.5
3BIO-B	3	0.195	0.182	0.200	0.709	0.606	0.500	0.000	0.786	8.5
Alexov Lab	1	0.367	0.245	0.127	1.064	0.961	0.500	0.000	0.464	13.8
CompBiomed UNITO	1	−0.041	−0.164	−0.091	0.559	0.447	0.411	−0.179	0.500	13.8
Lichtarge Lab	1	−0.020	0.155	0.127	3.788	3.395	0.500	0.000	0.750	14.8
Bologna Biocomputing	1	−0.223	−0.251	−0.205	0.638	0.493	0.536	0.069	0.571	14.8
BML	1	−0.356	−0.545	−0.345	0.634	0.495	0.500	0.000	0.214	18.0
AIBI-CAGI6	4	−0.234	0.027	0.055	6.852	5.466	0.429	−0.239	0.643	20.9
EASE-MM	1	−0.560	−0.545	−0.418	0.940	0.734	0.482	−0.039	0.321	21.9
MAPK1 phosphorylated										
Team	Model	r_P	r_S	r_{KT}	RMSE	MAE	BQ_2	MCC	AUC	<Rank>
BioSig	2	0.121	0.091	0.055	0.419	0.373	0.800	0.346	1.000	6.0
3billion	2	0.201	0.236	0.200	0.570	0.510	0.600	0.149	0.900	6.5
Strokach	3	0.613	0.127	0.055	0.626	0.608	0.500	0.000	1.000	7.3
Li Lab	3	0.507	0.545	0.418	1.005	0.944	0.550	0.100	0.700	8.0
3BIO-B	3	0.047	0.055	0.055	0.399	0.340	0.450	−0.100	0.200	8.4
Alexov Lab	1	0.430	0.391	0.273	1.265	1.217	0.500	0.000	0.600	10.9
CompBiomed UNITO	2	−0.365	−0.109	−0.073	0.488	0.387	0.650	0.194	0.600	11.3
Lichtarge Lab	1	−0.029	0.091	0.055	3.537	3.139	0.500	0.000	0.100	15.4
Bologna Biocomputing	1	−0.378	−0.324	−0.262	0.475	0.407	0.250	−0.289	0.350	15.4
BML	1	−0.362	−0.400	−0.309	0.716	0.692	0.500	0.000	0.000	17.1
AIBI-CAGI6	6	−0.309	−0.173	−0.127	5.512	4.464	0.400	−0.149	0.600	19.9

The eight measures of performance are defined in supplementary materials. Teams were allowed to submit multiple prediction sets using different approaches. The 'Model' column indicates the prediction set that achieved the best performanc

r_P, r_S, r_{KT} Pearson, Spearman, Kendall rank, correlation coefficients, *RMSE* root mean square error, *MAE* mean absolute error, BQ_2 balanced overall accuracy, *MCC* Matthews correlation coefficient, *AUC* area under the receiver operating characteristic curve, <*Rank*> the average rank is computed as the mean of the ranks achieved across the eight performance scores

Team 11: The *Pal Lab* at the Indian Institute of Science, Bangalore (India) submitted two sets of predictions of the catalytic efficiency for both MAPK proteins. The predictions were calculated by performing Molecular Dynamics (MD) simulations using the CGMM forcefield (Bhadra and Pal 2014). The variation of the Root Mean Square Fluctuation (ΔRMSF) between the mutant and the wild-type was used for predicting the catalytic efficiency ratio. MD simulations were run on different PDB structures.

Team 12: The *Strokach* group at the University of Toronto (Canada) submitted six sets of predictions for each MAPK protein. The predictions of $\Delta\Delta G^{H_2O}$ and catalytic efficiency were generated using four algorithms: ELASPIC2 (Strokach et al. 2021), ProteinSolver (Strokach et al. 2020), ProtBert (Elnaggar et al. 2022), and Rosetta (Park et al. 2016). The ELAPSIC2 predictions were computed both with the experimental and the AlphaFold predicted structure (Jumper et al. 2021). The same predictions were returned for both the unphosphorylated and the phosphorylated forms of the MAPKs variants.

Team 13: The *CompBiomed* group at the University of Torino (Italy) submitted two sets of $\Delta\Delta G^{H_2O}$ predictions for both MAPKs computed by using DDGun3D (Montanucci et al. 2019) and ACDC-NN-Seq (Benevenuta et al. 2021; Pancotti et al. 2021). Both methods returned the same predictions for the unphosphorylated and the phosphorylated forms of the MAPKs variants.

The supplementary materials provide a detailed description of the methods and procedures adopted by each team to predict the impact of the single amino acid substitutions.

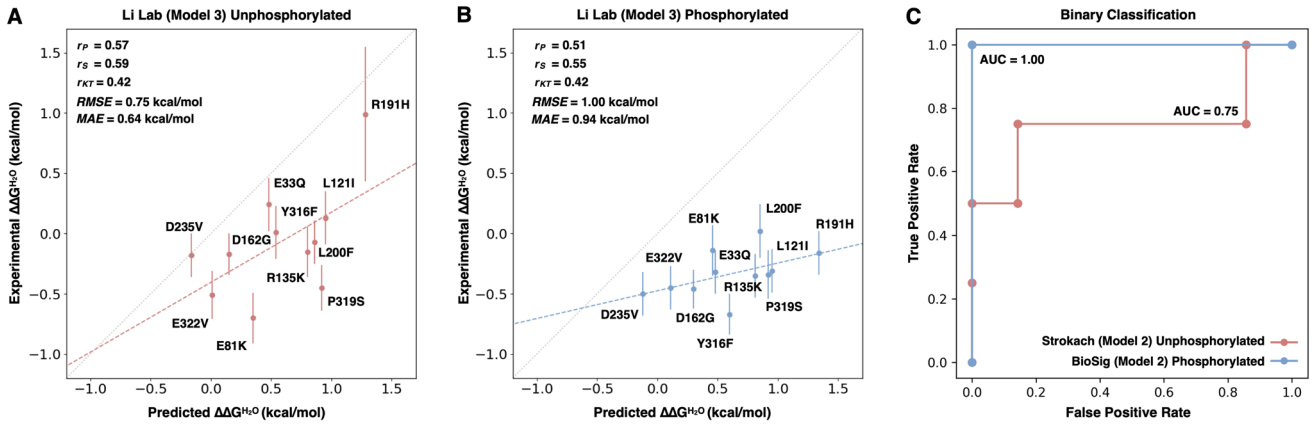

Fig. 3 Evaluation of the MAPK1 variant predictions. Linear regression curves of the highest correlating predictions of the unfolding $\Delta\Delta G^{H_2O}$ for unphosphorylated and phosphorylated MAPK1 variants, shown in panels (**A**) and (**B**) respectively, were submitted by the *Li Lab*. Receiver Operating Characteristic (ROC) curves of the best binary classification predictions of not destabilizing variants (unfolding $\Delta\Delta G^{H_2O} \geq 0.0$ kcal/mol) for unphosphorylated and phosphorylated MAPK1 variants (**C**) were provided by *Strokach* and *BioSig* teams respectively. r_P, r_S, r_{KT} = Pearson, Spearman, Kendall rank, correlation coefficients. *RMSE* Root Mean Square Error, *MAE* Mean Absolute Error, *AUC* Area Under the ROC curve

A comprehensive summary of all submissions is presented in Table S4.

Prediction assessment

To assess the predictions for the MAPK challenges, we employed a total of eight performance measures, comprising five from the regression tasks and three from the classification tasks (section Measures of Performance in Supplementary Materials). To compare the predicted and experimental values of $\Delta\Delta G^{H_2O}$ of each protein variant, we calculated three types of correlations (Person, Spearman and Kendall rank) and two types of errors (Root Mean Square Error and the Mean Absolute Error). Furthermore, we considered a threshold of 0.0 kcal/mol for classifying variants as destabilizing (unfolding $\Delta\Delta G^{H_2O} < 0.0$ kcal/mol) or not destabilizing (unfolding $\Delta\Delta G^{H_2O} \geq 0.0$ kcal/mol). Using this threshold for the binary classification task, we scored the predictions by calculating the balanced accuracy (BQ_2), the Matthews correlation coefficient (MCC) and the Area Under ROC Curve (AUC). Finally, we ranked all the submissions by considering each one of the eight performance measures and calculating the average value of the ranks, which were used to select the best predictions. In the assessment process, we evaluated the performance of the method in predicting the $\Delta\Delta G^{H_2O}$ of each MAPKs separately, and also the $\Delta\Delta G^{H_2O}$ of both MAPKs together. For the latter task, we combined the MAPK1 and MAPK3 variants based on their 2-state or 3-state unfolding mechanism.

We also evaluated the participants' performance in the binary classification task of predicting the catalytic efficiency ratio ($r_{kcat/KM}$). Specifically, we scored their ability to identify the variants that reduce the catalytic efficiency ($r_{kcat/KM} < 1$). For this task, the submitted predictions are interpreted as the likelihood for a variant to have no functional impact. Thus, the variants with scores lower than 0.5 were predicted to decrease catalytic efficiency. For this classification task, predictions were evaluated based on the calculation of BQ_2, MCC and AUC scores.

The definitions of the eight measures of performance, considered for this assessment, are reported in the supplementary materials.

Results

Assessment and performance evaluation on the MAPK ΔΔG predictions

In our assessment, we evaluated the performance of the participants in predicting the value of $\Delta\Delta G^{H_2O}$. For this task, we calculated five scores, three of which measure the correlations between experimental and predicted data (r_P, r_S and r_{KT}), the other two the prediction errors (RMSE and MAE). Additionally, we integrated regression measures with three classification scores (BQ_2, MCC, and AUC), derived from a threshold of 0.0 kcal/mol to distinguish destabilizing variants ($\Delta\Delta G^{H_2O} < 0$) from non-destabilizing ones ($\Delta\Delta G^{H_2O} \geq 0$). The performance in the regression tasks, along with the corresponding fitting curves for the best predictions from each team on both unphosphorylated and phosphorylated forms of MAPK1 and MAPK3, are shown in Figs. S2–S5. Furthermore, the classification scores of the Receiver Operating Characteristic curves are presented in Figs. S6–S9.

Table 5 Assessment of the best $\Delta\Delta G^{H_2O}$ predictions submitted by each team for phosphorylated and unphosphorylated MAPK3 variants

MAPK3 unphosphorylated										
Team	Model	r_P	r_S	r_{KT}	RMSE	MAE	BQ_2	MCC	AUC	<Rank>
Alexov Lab	1	0.606	0.615	0.455	1.977	1.445	0.750	0.674	0.800	1.4
Li Lab	2	0.209	0.231	0.212	2.163	1.645	0.500	0.000	0.450	4.3
CompBiomedUNITO	2	−0.056	0.131	0.132	2.625	1.992	0.600	0.158	0.450	7.5
BioSig	2	0.201	0.070	0.076	2.461	1.910	0.450	−0.135	0.375	7.9
Strokach	5	0.135	0.105	0.030	3.041	2.356	0.500	0.000	0.850	8.8
3billion	2	0.017	−0.028	−0.091	2.705	2.061	0.500	0.000	0.800	9.5
AIBI-CAGI6	5	0.073	0.196	0.212	5.027	4.005	0.600	0.200	0.850	9.9
BML	1	0.133	0.112	0.061	3.121	2.463	0.500	0.000	0.500	10.3
EASE-MM	1	0.024	−0.140	−0.061	3.128	2.474	0.550	0.135	0.600	13.1
3BIO-B	2	−0.134	−0.343	−0.198	2.841	2.198	0.600	0.200	0.200	14.9
Lichtarge Lab	1	−0.275	−0.056	−0.046	5.582	4.638	0.500	0.000	0.800	16.8
Bologna Biocomputing	1	−0.683	−0.599	−0.519	3.824	2.542	0.450	−0.076	0.425	21.5
MAPK3 phosphorylated										
Team	Model	r_P	r_S	r_{KT}	RMSE	MAE	BQ_2	MCC	AUC	<Rank>
BioSig	1	0.695	0.741	0.606	1.062	0.818	0.875	0.816	0.812	1.1
3BIO-B	3	0.646	0.580	0.424	1.199	0.993	0.625	0.426	0.906	3.8
Bologna Biocomputing	1	0.540	0.521	0.422	1.281	1.122	0.562	0.120	0.625	6.5
CompBiomedUNITO	2	0.692	0.471	0.395	1.416	1.228	0.562	0.120	0.719	7.0
3billion	1	0.458	0.510	0.394	1.369	1.122	0.500	0.000	0.812	7.3
Strokach	4	0.296	0.462	0.394	3.667	2.642	0.500	0.000	0.625	13.4
BML	1	−0.255	−0.098	−0.091	1.449	1.205	0.500	0.000	0.500	13.5
AIBI-CAGI6	5	0.085	0.077	0.091	3.232	2.697	0.375	−0.316	0.531	15.3
Li Lab	1	−0.344	−0.434	−0.303	2.015	1.599	0.562	0.213	0.500	18.9
Lichtarge Lab	5	−0.078	−0.287	−0.242	3.600	2.824	0.500	0.000	0.219	19.8
Alexov Lab	1	−0.590	−0.448	−0.333	2.162	1.724	0.375	−0.426	0.438	21.9

The eight measures of performance are defined in supplementary materials. Teams were allowed to submit multiple prediction sets using different approaches. The 'Model' column indicates the prediction set that achieved the best performance

r_P, r_S, r_{KT} Pearson, Spearman, Kendall rank, correlation coefficients, *RMSE* root mean square error, *MAE* mean absolute error, BQ_2 balanced overall accuracy, *MCC* Matthews correlation coefficient, *AUC* area under the receiver operating characteristic curve, <*Rank*> the average rank is computed as the mean of the ranks achieved across the eight performance scores

Assessment of the MAPK1 stability change predictions

For MAPK1, the $\Delta\Delta G^{H_2O}$ values predicted by each team were compared with the experimental values, derived by fitting to the data a 2-state unfolding equation. The performance achieved by the best prediction submitted by each team is summarized in Table 4. Overall, for the majority of the teams, the performance in regression mode exhibits low values in terms of Spearman and Kendall rank correlation coefficients. However, despite not ranking first, the predictions submitted by Team 6 (*Li Lab*) achieved an average of approximately 0.5 for all correlation coefficients, both for unphosphorylated and phosphorylated MAPK1. The regression curves between predicted and experimental $\Delta\Delta G^{H_2O}$ are displayed in Fig. 3A and B. In the evaluation of $\Delta\Delta G^{H_2O}$ predictions in classification mode, Team 12 (*Strokach* group) and Team 4 (*BioSig*) showed the highest performance for the unphosphorylated and phosphorylated forms of MAPK1, respectively. Specifically, the predictions from the *Strokach* group for $\Delta\Delta G^{H_2O}$ for not destabilizing variants in unphosphorylated MAPK1 achieved a balanced overall accuracy (BQ_2) and Area Under the ROC Curve (AUC) of 0.75. On the other hand, for the phosphorylated MAPK1, the *BioSig* team reached a BQ_2 of 0.8 and an AUC of 1.0. The AUC curves featuring the best predictions in classification mode for the *Strokach* and *BioSig* teams are illustrated in Fig. 3C.

Assessment of the MAPK3 stability change predictions

For MAPK3, predicted $\Delta\Delta G^{H_2O}$ values were compared with experimental values derived from fitting to the data

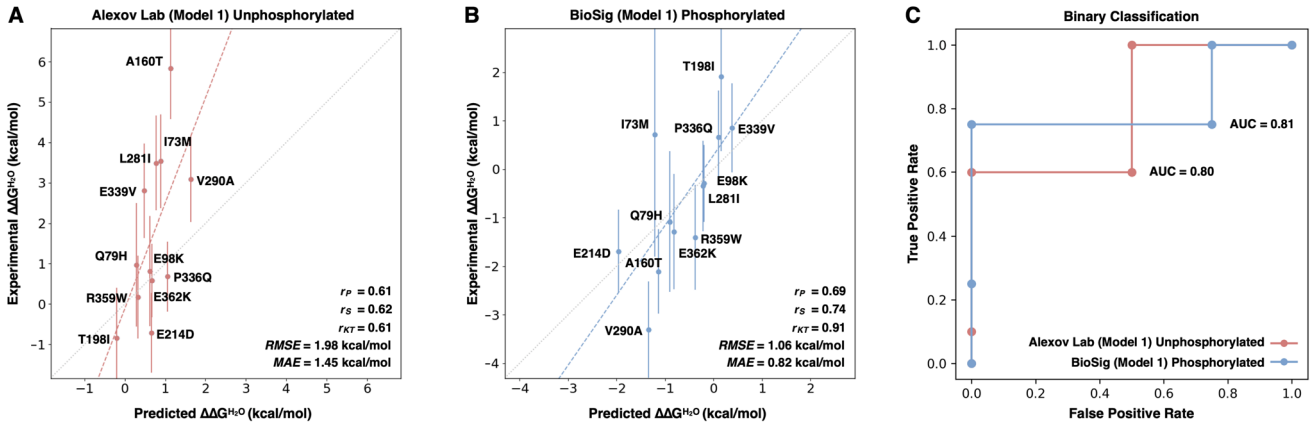

Fig. 4 Evaluation of the MAPK3 variant predictions. Linear regression curves of the most accurate predictions of the unfolding $\Delta\Delta G^{H_2O}$ for unphosphorylated and phosphorylated MAPK3 variants, shown in panels (**A**) and (**B**), were submitted by the *Alexov Lab* and *BioSig* team respectively. Receiver Operating Characteristic (ROC) curves of the best binary classification predictions of not destabilizing variants (unfolding $\Delta\Delta G^{H_2O} \geq 0.0$ kcal/mol) for unphosphorylated and phosphorylated MAPK3 variants (**C**) were provided by the *Alexov Lab* (red) and *BioSig* team (blue) respectively. r_P, r_S, r_{KT} Pearson, Spearman, Kendall rank, correlation coefficients, *RMSE* Root Mean Square Error, *MAE* Mean Absolute Error, *AUC* Area Under the ROC curve

a 3-state unfolding equation. The performance of the best prediction from each team is summarized in Table 5. Overall, the performance in regression mode for the most accurate methods achieved Pearson and Spearman correlation coefficients above 0.6 for both the unphosphorylated and phosphorylated forms of MAPK3. The best predictions for unphosphorylated MAPK3 were provided by *Alexov Lab* (Fig. 4A), while, for the phosphorylated form, the *BioSig* team achieved Pearson and Spearman correlation coefficients of approximately 0.7 (Fig. 4B). In classification mode, both teams achieved an AUC above 0.8 (Fig. 4C). Specifically, *Alexov Lab*'s $\Delta\Delta G^{H_2O}$ predictions for not destabilizing variants in unphosphorylated MAPK3 achieved a BQ_2 of 0.75 and a Matthews Correlation Coefficient (MCC) of 0.67, while *BioSig*'s predictions on phosphorylated MAPK3 reached a BQ_2 of 0.87 and an MCC of 0.82.

Assessment of the predictions on combined ΔΔG datasets

To assess the teams' performance in predicting the impact of variants, we also combined data on MAPK1 and MAPK3, based on their unfolding mechanisms and phosphorylation states. To generate more reliable datasets, we only considered the subset of $\Delta\Delta G^{H_2O}$ values obtained from optimal fitting with 2-state and/or 3-state unfolding mechanisms, in the absence of preliminary assumptions of the mechanism type. The results show that good performance is achieved when considering unphosphorylated MAPK variants folding through a 2-state mechanism, and phosphorylated MAPK variants folding through a 3-state mechanism. The performance of the teams on these two subsets is summarized in Table 6. The best regression scores for unphosphorylated MAPKs were provided by *Li Lab*. On a dataset composed of 12 variants (9 from MAPK1 and 3 from MAPK3), the Pearson (r_P) and Spearman (r_S) correlation coefficients were above 0.7, and the RMSE was below 1.0 kcal/mol (Fig. 5A). For the phosphorylated MAPKs, the best regression scores were achieved by the *BioSig* group, with r_P and r_S correlation coefficients above 0.65 and an RMSE of approximately 1.4 kcal/mol (Fig. 5B) on a dataset composed of 15 variants (3 from MAPK1 and 12 from MAPK3). On both datasets, *BioSig* also reached the best performance in classification mode, with AUCs of 0.71 and 0.87 for not destabilizing variants in unphosphorylated and phosphorylated MAPKs, respectively (Fig. 5C).

Assessment and performance evaluation on the catalytic efficiency ratio

The MAPK challenges also include the prediction of the catalytic efficiency ratio ($r_{kcat/KM}$) between mutant and wild-type proteins. For assessing the performance on this task, we only consider the evaluation of the predictions in classification mode, applying a threshold on the experimental $r_{kcat/KM}$ equal to 1. For the predictions returned by the teams, the classification threshold is set to 0.5. Using this classification criteria, we evaluated the performance of six teams and found that *AIBI-CAGI6* (Team 2) and *3BIO-B* (Team 5) provided the most accurate predictions of $r_{kcat/KM}$ for MAPK1 and MAPK3, respectively

Table 6 Assessment of the best $\Delta\Delta G^{H_2O}$ predictions submitted by each team for phosphorylated and unphosphorylated MAPK variants: MAPK1 and MAPK3 variants are combined into a single set based on their unfolding mechanism

Unphosphorylated MAPKs with 2-state unfolding mechanism										
Team	Model	r_P	r_S	r_{KT}	RMSE	MAE	BQ_2	MCC	AUC	<Rank>
BioSig	2	0.596	0.305	0.168	0.305	0.224	0.786	0.598	0.714	4.3
Li Lab	3	0.723	0.739	0.565	0.592	0.522	0.571	0.255	0.743	4.3
Strokach	2	0.173	0.186	0.107	0.458	0.347	0.700	0.529	0.686	6.8
Alexov Lab	1	0.429	0.382	0.260	0.829	0.691	0.571	0.255	0.600	8.9
3BIO-B	3	0.262	0.315	0.382	0.710	0.637	0.429	−0.255	0.743	9.1
3billion	1	0.081	0.147	0.107	0.459	0.346	0.429	−0.255	0.714	9.9
Lichtarge Lab	1	0.226	0.511	0.382	3.737	3.394	0.500	0.000	0.800	11.4
Bologna Biocomputing	1	−0.154	−0.158	−0.116	0.588	0.439	0.586	0.169	0.614	13.0
BML	1	0.137	−0.319	−0.107	0.693	0.600	0.500	0.000	0.400	13.9
CompBiomed UNITO	1	−0.138	−0.385	−0.229	0.549	0.422	0.414	−0.169	0.371	16.4
AIBI-CAGI6	2	−0.493	0.035	0.076	5.338	3.728	0.700	0.529	0.657	18.3
EASE-MM	1	−0.595	−0.571	−0.412	0.962	0.770	0.457	−0.098	0.286	21.6
Phosphorylated MAPKs with 3-state unfolding mechanism										
Team	Model	r_P	r_S	r_{KT}	RMSE	MAE	BQ_2	MCC	AUC	<Rank>
BioSig	1	0.661	0.682	0.543	1.377	1.040	0.917	0.866	0.870	1.0
Bologna Biocomputing	1	0.590	0.578	0.473	1.560	1.310	0.667	0.327	0.704	4.8
3BIO-B	3	0.538	0.520	0.364	1.570	1.263	0.583	0.327	0.843	5.3
3billion	1	0.524	0.450	0.333	1.645	1.350	0.583	0.327	0.722	7.6
CompBiomed UNITO	2	0.530	0.399	0.341	1.702	1.428	0.583	0.185	0.722	8.3
Strokach	4	0.380	0.546	0.448	3.486	2.548	0.500	0.000	0.704	12.8
BML	1	−0.222	−0.111	−0.105	1.824	1.479	0.500	0.000	0.407	15.9
AIBI-CAGI6	5	0.033	−0.061	0.029	3.464	2.987	0.333	−0.408	0.463	18.5
Li Lab	1	−0.507	−0.504	−0.390	2.206	1.769	0.500	0.000	0.370	21.4
Alexov Lab	1	−0.575	−0.513	−0.364	2.307	1.844	0.417	−0.327	0.389	22.4
Lichtarge Lab	3	−0.359	−0.457	−0.371	4.389	3.404	0.500	0.000	0.204	24.0

The table reports the performance calculated for unphosphorylated MAPK variants with a 2-state unfolding mechanism and phosphorylated MAPK variants with a 3-state unfolding mechanism. Teams were allowed to submit multiple prediction sets using different approaches. The 'Model' column indicates the prediction set that achieved the best performance

r_P, r_S, r_{KT} Pearson, Spearman, Kendall rank, correlation coefficients, *RMSE* Root Mean Square Error, *MAE* mean absolute error, BQ_2 balanced overall accuracy, *MCC* Matthews correlation coefficient, *AUC* area under the receiver operating characteristic curve, <*Rank*> the average rank is computed as the mean of the ranks achieved across the eight performance scores

(Table 7). Specifically, the *AIBI-CAGI6* team achieved a BQ_2 of 0.77 and an MCC of 0.83 for predicting the catalytic efficiency ratio of MAPK1 mutants, while the *3BIO-B* team reached a BQ_2 of 0.75 and an AUC of 0.80 for predicting the catalytic efficiency ratio of MAPK3 mutants. The ROC curves calculated on both MAPK prediction sets are shown in Fig. 6.

The classification and regression scores for all the prediction models submitted by the teams, for both unphosphorylated and phosphorylated MAPK proteins, are summarized in Supplementary File 1. All the submitted predictions are provided in Supplementary File 2.

Discussion

The assessment of the MAPK challenges in the CAGI6 experiment provided an opportunity to evaluate the performance of several variant annotation methods for predicting the impact of single amino acid variations on protein stability and catalytic function. Overall, the MAPK challenges were more complex than the Frataxin challenge from CAGI5. Unlike Frataxin, which folds via a 2-state mechanism, the unfolding mechanism for both MAPK proteins can vary, potentially leading to the appearance of an intermediate state in the unfolding curve. Furthermore, the

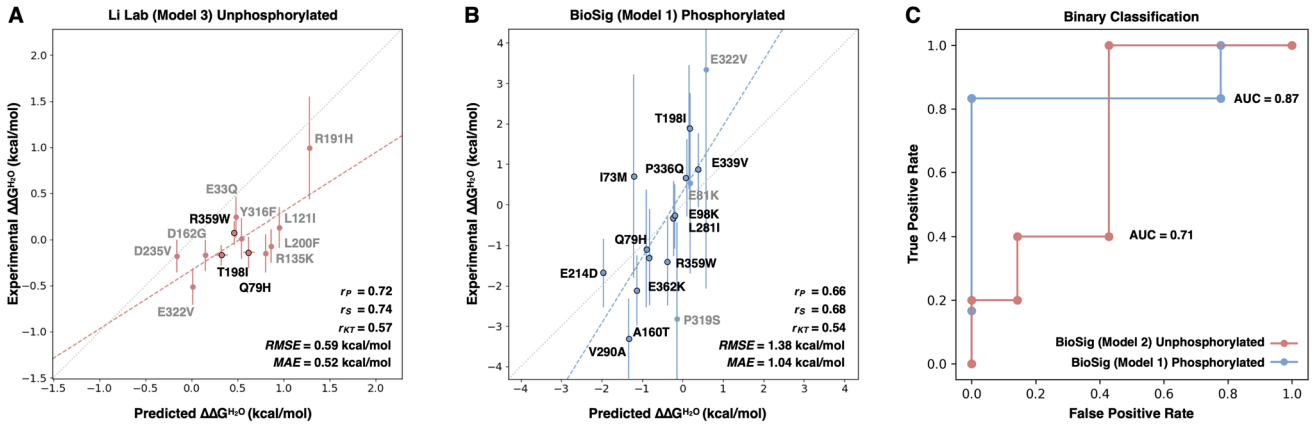

Fig. 5 Evaluation of MAPK Variant Predictions: Variants of both MAPKs were combined based on their unfolding 2-state or 3-state model, with MAPK1 and MAPK3 variants indicated in gray and black, respectively. **A** Linear regression curves for the most accurate predictions (*Li Lab*) of the unfolding $\Delta\Delta G^{H_2O}$ for unphosphorylated MAPK variants with a 2-state folding mechanism. **B** Linear regression curves for the most accurate predictions (*BioSig*) of the unfolding $\Delta\Delta G^{H_2O}$ for phosphorylated MAPK variants with a 3-state folding mechanism. **C** Receiver Operating Characteristic (ROC) curves for the best binary classification predictions (*BioSig*) of not destabilizing variants (unfolding $\Delta\Delta G^{H_2O} \geq 0.0$ kcal/mol) for both unphosphorylated and phosphorylated variants, provided by the *BioSig* team. r_P, r_S, r_{KT} Pearson, Spearman, Kendall rank, correlation coefficients, *RMSE* Root Mean Square Error, *MAE* Mean Absolute Error, *AUC* Area Under the ROC curve

Table 7 Assessment of the best binary classification predictions of the catalytic efficiency ratio ($r_{kcat/KM}$) for MAPK1 and MAPK3 variants, submitted by each group

MAPK1 catalytic efficiency predictions					
Team	Model	BQ_2	MCC	AUC	<Rank>
AIBI-CAGI6	1	0.771	0.542	0.833	1.0
Lichtarge Lab	5	0.708	0.386	0.583	3.0
Strokach	2	0.542	0.083	0.792	4.0
Pal Lab	1	0.438	−0.194	0.750	6.0
3BIO-B	2	0.417	−0.149	0.208	8.7
MAPK3 catalytic efficiency predictions					
Team	Model	BQ_2	MCC	AUC	<Rank>
3BIO-B	1	0.750	0.378	0.800	1.0
Lichtarge Lab	5	0.500	0.000	0.450	2.7
Strokach	4	0.500	0.000	0.450	2.7
Pal Lab	1	0.500	0.000	0.400	3.0
AIBI-CAGI6	4	0.400	−0.200	0.200	5.3

Variants with $r_{kcat/KM} < 1$ are those decreasing the catalytic efficiency. The submitted $r_{kcat/KM}$ are interpreted as the likelihood that each variant has no functional impact. Consequently, variants with predicted $r_{kcat/KM}$ lower than 0.5 are considered to decrease the catalytic efficiency. The three measures of performance for binary classifiers are defined in supplementary materials. Teams were allowed to submit multiple prediction sets using different approaches. The 'Model' column indicates the prediction set that achieved the best performance

BQ_2 balanced overall accuracy, *MCC* Matthews correlation coefficient, *AUC* area under the receiver operating characteristic curve, <*Rank*> the average rank is computed as the mean of the ranks achieved across the three classification scores

phosphorylation state is generally not considered in stability change prediction tools.

Thus, in the first part of the assessment, we harmonized the data by fitting specific unfolding models to the experimental results. Specifically, we considered all MAPK1 and MAPK3 variants to follow 2-state and 3-state unfolding mechanisms, respectively. Some submitted predictions performed well, particularly among those for the MAPK3 challenge, where *Alexov Lab* and the *BioSig* team achieved the best results. The assessment of the MAPK1 challenge was more complex. Most MAPK1 variants exhibited a 2-state unfolding mechanism, with many of them showing

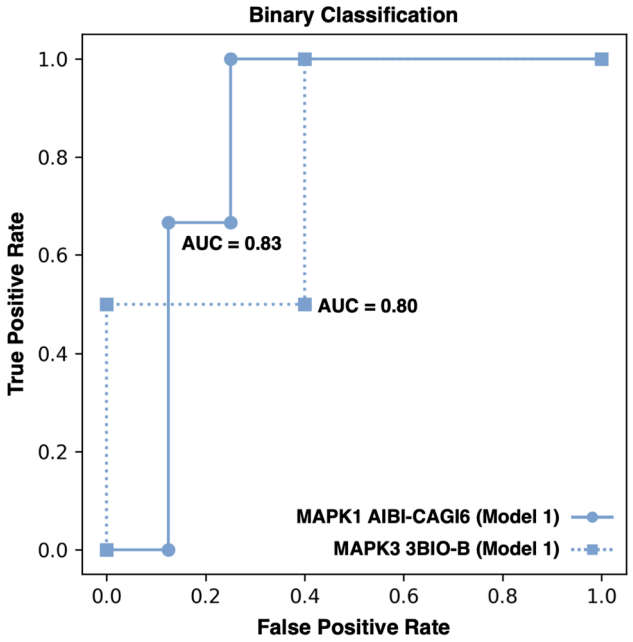

Fig. 6 Receiver Operating Characteristic (ROC) curves of the binary classifiers of the catalytic efficiency rate ($r_{\mathrm{kcat/Km}}$). Best predictions of efficiency decreasing variants ($r_{\mathrm{kcat/Km}} < 1.0$) for MAPK1 and MAPK3 were provided by the *AIBI-CAGI6* (solid line) and *3BIO-B* (dashed line) teams, respectively. The classification threshold for predicted $r_{\mathrm{kcat/Km}}$ is set to 0.5. *AUC* Area Under the ROC curve

only marginal changes in stability, compared to the wild-type. Consequently, ranking correlation coefficients (r_S, r_{KT}) tended to have low values. The most accurate predictions, with balanced correlation coefficients, were submitted by the *Li Lab*. However, such predictions did not perform well in the binary classification task, where the best results were achieved by the *BioSig* and *Strokach* teams.

In the second part of the assessment, we evaluated the performance of the methods on combined datasets by merging MAPK variants with homogeneous apparent folding mechanisms. Our analysis, which included variants from both MAPK1 and MAPK3, confirmed that *BioSig* and *Li Lab* achieved the highest accuracy. Notably, the strong performance levels observed when considering unphosphorylated MAPKs folding through an apparent 2-state mechanism and phosphorylated MAPKs folding through an apparent 3-state mechanism support the hypothesis of a relationship between phosphorylation state and folding mechanism. However, the bias in the datasets towards variants of a specific protein makes it challenging to prove this hypothesis.

Finally, the most difficult task of the MAPK challenges was predicting the catalytic efficiency rate. For this task, predictions as a regression task yielded poor results, thus we evaluated performance as a classification task only. Out of the five teams that submitted predictions, the *3BIO-B* and *AIBI-CAGI6* teams achieved the best performances.

In conclusion, the MAPK challenges represent an advancement over the previous Frataxin challenge, where participants were asked to predict the functional and structural impact of protein variants. The potential presence of an intermediate state in the unfolding process increased the complexity of the challenge. Nonetheless, some methods provided good predictions in both regression and classification modes, with relatively low RMSE. We expect that the data generated by the MAPK challenges could be reused for training and testing new methods for predicting the impact of protein variants.

Conclusions

In the MAPK challenges, computational approaches based on ensemble methods like gradient boosting and random forests were particularly effective in binary classification tasks, such as distinguishing destabilizing from non-destabilizing variants. Teams like *BioSig* and *Strokach* excelled in these tasks, with *BioSig* achieving a perfect AUC of 1.0 for phosphorylated MAPK1. For regression tasks, energy-based models, combined with machine learning techniques, performed best. Using this approach, the *Alexov Lab* achieved high correlation coefficients for MAPK3 by integrating evolutionary conservation of the protein, mutation site, and its sequence neighborhood along with physicochemical properties of the wild-type and mutant amino acids with machine learning. This combination allowed for more accurate $\Delta\Delta G^{\mathrm{H_2O}}$ predictions, particularly for the phosphorylated MAPK3 form.

Although deep learning models, such as large language models (LLMs), have proven powerful in addressing many bioinformatics problems, they did not achieve the best performance in the MAPK challenges. Instead, hybrid approaches that combine energy models with machine learning proved the most successful. The strong performance of *BioSig* and *3BIO-B* in predicting the impact of variants on protein stability and catalytic efficiency, respectively, suggests that integrating these methods provides a significant advantage.

In summary, no single approach outperformed all others. However, a combination of ensemble methods, energy-based models, and machine learning delivered the best performance, in predicting changes in protein stability and catalytic efficiency resulting from mutations.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-024-02724-8.

Acknowledgements We acknowledge all the CAGI organizers for their support in the organization of the MAPK Challenges.

Author contributions E.C., P.T., R.C., V.C., A.P., M.P. conceptualized the study. R.C., V.C., A.P., M.P. generated the experimental data and performed the preliminary analysis. E.C. P.T C.A.E.S analyzed the prediction and performed the assessment. E.C., P.T wrote the original draft; S.E.B, P.R., C.B., A.K. helped to define and organize the CAGI challenge All authors have read and agreed to the published version of the manuscript. E.A., M.A.A., D.B.A., G.B R.C., J.C., P.F., L.F., P.K., M.L, D.L, O.L., S.M, P.L.M., D.P, S.K.P., D.E.V.P, S.P., F.P., C.H.M.R., M.R., C.S., M.S., Y.S., A.V.S., Y.S., J.W. generated and/or submitted the prediction to the MAPK1 and MAPK3 challenges. All authors have read and agreed to the published version of the manuscript.

Funding E.C. acknowledges support from *"Ministero dell'Istruzione, dell'Università e della Ricerca"*, MIUR-PRIN-201744NR8S—Integrative tools for defining the molecular basis of the diseases: Computational and Experimental methods for Protein Variant Interpretation. D.A. acknowledges support from NHMRC GNT1174405 and The Victorian Government's Operational Infrastructure Support Program. J.C. acknowledges support from NIGMS R01GM093123 and NSF DBI2308699. M.L. acknowledges support from National Natural Science Foundation of China, Grant/Award Number: 32070665. P.L.M. acknowledges support from *"Ministero dell'Istruzione, dell'Università e della Ricerca"* by ELIXIR-IT, the Italian research infrastructure for life-science data, ELIXIRNextGenIT Grant Code IR0000010. O.L. acknowledges support from the Grant Number AG068214. E.A. acknowledges support from NIH R01GM093937 and R35GM151964. Y.S. acknowledges support from NIGMS R35GM124952. F.P., M.R. and M.S. acknowledge support from the F.R.S.-FNRS. The CAGI experiment coordination is supported by NIH U41 HG007346 (to S.E.B.) and the CAGI conference by NIH R13 HG006650.

Data availability The classification and regression scores for all the prediction models submitted by the teams, for both unphosphorylated and phosphorylated MAPK proteins, are summarized in Supplementary File 1. All the submitted predictions are provided in Supplementary File 2.

Declarations

Conflict of interest The authors declare no competing interests.

References

Attwood MM, Fabbro D, Sokolov AV et al (2021) Trends in kinase drug discovery: targets, indications and inhibitor design. Nat Rev Drug Discov 20:839–861. https://doi.org/10.1038/s41573-021-00252-y

Balmanno K, Kidger AM, Byrne DP et al (2023) ERK1/2 inhibitors act as monovalent degraders inducing ubiquitylation and proteasome-dependent turnover of ERK2, but not ERK1. Biochem J 480:587–605. https://doi.org/10.1042/BCJ20220598

Benevenuta S, Pancotti C, Fariselli P et al (2021) An antisymmetric neural network to predict free energy changes in protein variants. J Phys Appl Phys 54:245403. https://doi.org/10.1088/1361-6463/abedfb

Bhadra P, Pal D (2014) De novo inference of protein function from coarse-grained dynamics. Proteins 82:2443–2454. https://doi.org/10.1002/prot.24609

Chen Y, Lu H, Zhang N et al (2020) PremPS: predicting the impact of missense mutations on protein stability. PLoS Comput Biol 16:e1008543. https://doi.org/10.1371/journal.pcbi.1008543

Compiani M, Capriotti E (2013) Computational and theoretical methods for protein folding. Biochemistry 52:8601–8624. https://doi.org/10.1021/bi4001529

Critical Assessment of Genome Interpretation Consortium (2024) CAGI, the critical assessment of genome interpretation, establishes progress and prospects for computational genetic variant interpretation methods. Genome Biol 25:53. https://doi.org/10.1186/s13059-023-03113-6

Dehouck Y, Kwasigroch JM, Gilis D, Rooman M (2011) PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinform 12:151. https://doi.org/10.1186/1471-2105-12-151

Elnaggar A, Heinzinger M, Dallago C et al (2022) ProtTrans: toward understanding the language of life through self-supervised learning. IEEE Trans Pattern Anal Mach Intell 44:7112–7127. https://doi.org/10.1109/TPAMI.2021.3095381

Folkman L, Stantic B, Sattar A, Zhou Y (2016) EASE-MM: sequence-based prediction of mutation-induced stability changes with feature-based multiple models. J Mol Biol 428:1394–1405. https://doi.org/10.1016/j.jmb.2016.01.012

Frappier V, Chartier M, Najmanovich RJ (2015) ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability. Nucleic Acids Res 43:W395-400. https://doi.org/10.1093/nar/gkv343

Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320:369–387

Jumper J, Evans R, Pritzel A et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Katsonis P, Lichtarge O (2014) A formal perturbation equation between genotype and phenotype determines the evolutionary action of protein-coding variations on fitness. Genome Res 24:2050–2058. https://doi.org/10.1101/gr.176214.114

Khezri MR, Yousefi K, Esmaeili A, Ghasemnejad-Berenji M (2023) The role of ERK1/2 pathway in the pathophysiology of Alzheimer's disease: an overview and update on new developments. Cell Mol Neurobiol 43:177–191. https://doi.org/10.1007/s10571-022-01191-x

Kim EK, Choi E-J (2010) Pathological roles of MAPK signaling pathways in human diseases. Biochim Biophys Acta 1802:396–405. https://doi.org/10.1016/j.bbadis.2009.12.009

Laimer J, Hofer H, Fritz M et al (2015) MAESTRO-multi agent stability prediction upon point mutations. BMC Bioinformatics 16:116. https://doi.org/10.1186/s12859-015-0548-6

Lavoie H, Gagnon J, Therrien M (2020) ERK signalling: a master regulator of cell behaviour, life and fate. Nat Rev Mol Cell Biol 21:607–632. https://doi.org/10.1038/s41580-020-0255-7

Li G, Panday SK, Alexov E (2021) SAAFEC-SEQ: a sequence-based method for predicting the effect of single point mutations on protein thermodynamic stability. Int J Mol Sci 22:606. https://doi.org/10.3390/ijms22020606

Liu S, Ma H, Zhang H et al (2021) Recent advances on signaling pathways and their inhibitors in rheumatoid arthritis. Clin Immunol Orlando Fla 230:108793. https://doi.org/10.1016/j.clim.2021.108793

Manning G, Whyte DB, Martinez R et al (2002) The protein kinase complement of the human genome. Science 298:1912–1934. https://doi.org/10.1126/science.1075762

Marabotti A, Scafuri B, Facchiano A (2020) Predicting the stability of mutant proteins by computational approaches: an overview. Brief Bioinform. https://doi.org/10.1093/bib/bbaa074

Marchi M, D'Antoni A, Formentini I et al (2008) The N-terminal domain of ERK1 accounts for the functional differences with ERK2. PLoS ONE 3:e3873. https://doi.org/10.1371/journal.pone.0003873

Metz KS, Deoudes EM, Berginski ME et al (2018) Coral: clear and customizable visualization of human kinome data. Cell Syst 7:347-350.e1. https://doi.org/10.1016/j.cels.2018.07.001

Mistry J, Chuguransky S, Williams L et al (2021) Pfam: the protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913

Montanucci L, Capriotti E, Frank Y et al (2019) DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. BMC Bioinform. https://doi.org/10.1186/s12859-019-2923-1

Nikam R, Kulandaisamy A, Harini K et al (2021) ProThermDB: thermodynamic database for proteins and mutants revisited after 15 years. Nucleic Acids Res 49:D420–D424. https://doi.org/10.1093/nar/gkaa1035

Pancotti C, Benevenuta S, Repetto V et al (2021) A deep-learning sequence-based method to predict protein stability changes upon genetic variations. Genes 12:911. https://doi.org/10.3390/genes12060911

Pancotti C, Benevenuta S, Birolo G et al (2022) Predicting protein stability changes upon single-point mutation: a thorough comparison of the available tools on a new dataset. Brief Bioinform 23:bbab555. https://doi.org/10.1093/bib/bbab555

Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL (2017) SDM: a server for predicting effects of mutations on protein stability. Nucleic Acids Res 45:W229–W235. https://doi.org/10.1093/nar/gkx439

Park H, Bradley P, Greisen P et al (2016) Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. J Chem Theory Comput 12:6201–6212. https://doi.org/10.1021/acs.jctc.6b00819

Petrosino M, Pasquo A, Novak L et al (2019) Characterization of human frataxin missense variants in cancer tissues. Hum Mutat 40:1400–1413. https://doi.org/10.1002/humu.23789

Petrosino M, Novak L, Pasquo A et al (2021) Analysis and interpretation of the impact of missense variants in cancer. Int J Mol Sci 22:5416. https://doi.org/10.3390/ijms22115416

Petrosino M, Novak L, Pasquo A et al (2023) The complex impact of cancer-related missense mutations on the stability and on the biophysical and biochemical properties of MAPK1 and MAPK3 somatic variants. Hum Genom 17:95. https://doi.org/10.1186/s40246-023-00544-x

Pires DEV, Ascher DB, Blundell TL (2014a) DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res 42:W314-319. https://doi.org/10.1093/nar/gku411

Pires DEV, Ascher DB, Blundell TL (2014b) mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinform Oxf Engl 30:335–342. https://doi.org/10.1093/bioinformatics/btt691

Pucci F, Bernaerts KV, Kwasigroch JM, Rooman M (2018) Quantification of biases in predictions of protein stability changes upon mutations. Bioinform Oxf Engl 34:3659–3665. https://doi.org/10.1093/bioinformatics/bty348

Ring AY, Sours KM, Lee T, Ahn NG (2011) Distinct patterns of activation-dependent changes in conformational mobility between ERK1 and ERK2. Int J Mass Spectrom 302:101–109. https://doi.org/10.1016/j.ijms.2010.08.020

Rodrigues CH, Pires DE, Ascher DB (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res 46:W350–W355. https://doi.org/10.1093/nar/gky300

Rodrigues CHM, Pires DEV, Ascher DB (2021) DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations. Protein Sci Publ Protein Soc 30:60–69. https://doi.org/10.1002/pro.3942

Roskoski R (2019) Targeting ERK1/2 protein-serine/threonine kinases in human cancers. Pharmacol Res 142:151–168. https://doi.org/10.1016/j.phrs.2019.01.039

Sanavia T, Birolo G, Montanucci L et al (2020) Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. Comput Struct Biotechnol J 18:1968–1979. https://doi.org/10.1016/j.csbj.2020.07.011

Sasidharan Nair P, Vihinen M (2013) VariBench: a benchmark database for variations. Hum Mutat 34:42–49. https://doi.org/10.1002/humu.22204

Savojardo C, Fariselli P, Martelli PL, Casadio R (2016) INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinforma Oxf Engl 32:2542–2544. https://doi.org/10.1093/bioinformatics/btw192

Savojardo C, Petrosino M, Babbi G et al (2019) Evaluating the predictions of the protein stability change upon single amino acid substitutions for the FXN CAGI5 challenge. Hum Mutat 40:1392–1399. https://doi.org/10.1002/humu.23843

Sondka Z, Dhir NB, Carvalho-Silva D et al (2024) COSMIC: a curated database of somatic variants and clinical data for cancer. Nucleic Acids Res 52:D1210–D1217. https://doi.org/10.1093/nar/gkad986

Strokach A, Becerra D, Corbi-Verge C et al (2020) Fast and flexible protein design using deep graph neural networks. Cell Syst 11:402-411.e4. https://doi.org/10.1016/j.cels.2020.08.016

Strokach A, Lu TY, Kim PM (2021) ELASPIC2 (EL2): combining contextualized language models and graph neural networks to predict effects of mutations. J Mol Biol 433:166810. https://doi.org/10.1016/j.jmb.2021.166810

Sun Y, Shen Y (2023) Structure-informed protein language models are robust predictors for variant effects. Res Sq. https://doi.org/10.21203/rs.3.rs-3219092/v1

Varjosalo M, Keskitalo S, Van Drogen A et al (2013) The protein interaction landscape of the human CMGC kinase group. Cell Rep 3:1306–1320. https://doi.org/10.1016/j.celrep.2013.03.027

Won D-G, Kim D-W, Woo J, Lee K (2021) 3Cnet: pathogenicity prediction of human variants using multitask learning with evolutionary constraints. Bioinform Oxf Engl 37:4626–4634. https://doi.org/10.1093/bioinformatics/btab529

Wong WC, Kim D, Carter H et al (2011) CHASM and SNVBox: toolkit for detecting biologically important single nucleotide mutations in cancer. Bioinform Oxf Engl 27:2147–2148. https://doi.org/10.1093/bioinformatics/btr357

Zhang Y, Jin D, Kang X et al (2021) Signaling pathways involved in diabetic renal fibrosis. Front Cell Dev Biol 9:696542. https://doi.org/10.3389/fcell.2021.696542

Authors and Affiliations

Paola Turina[1] · Maria Petrosino[2] · Carlos A. Enriquez Sandoval[1] · Leonore Novak[2] · Alessandra Pasquo[3] · Emil Alexov[4] · Muttaqi Ahmad Alladin[5] · David B. Ascher[6,7] · Giulia Babbi[1] · Constantina Bakolitsa[8] · Rita Casadio[1] · Jianlin Cheng[9] · Piero Fariselli[10] · Lukas Folkman[11] · Akash Kamandula[12] · Panagiotis Katsonis[13] · Minghui Li[14] · Dong Li[15] · Olivier Lichtarge[13] · Sajid Mahmud[9] · Pier Luigi Martelli[1] · Debnath Pal[5] · Shailesh Kumar Panday[4] · Douglas E. V. Pires[16] · Stephanie Portelli[6,7] · Fabrizio Pucci[15] · Carlos H. M. Rodrigues[6] · Marianne Rooman[15] · Castrense Savojardo[1] · Martin Schwersensky[15] · Yang Shen[17] · Alexey V. Strokach[18] · Yuanfei Sun[17] · Junwoo Woo[19] · Predrag Radivojac[12] · Steven E. Brenner[8,20,21] · Roberta Chiaraluce[2] · Valerio Consalvi[2] · Emidio Capriotti[1,22]

✉ Roberta Chiaraluce
roberta.chiaraluce@uniroma1.it

✉ Valerio Consalvi
valerio.consalvi@uniroma1.it

✉ Emidio Capriotti
emidio.capriotti@unibo.it

[1] Department of Pharmacy and Biotechnology, University of Bologna, 40126 Bologna, Italy

[2] Department of Biochemical Sciences “A. Rossi Fanelli”, Sapienza University of Roma, 00185 Rome, Italy

[3] Diagnostics and Metrology Laboratory FSN-TECFIS-DIM, ENEA CR Frascati, 00044 Frascati, Italy

[4] Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA

[5] Department of Computational and Data Sciences, Indian Institute of Science, Bangaluru 560012, India

[6] Computational Biology and Clinical Informatics, Baker Heart and Diabetes Institute, Melbourne, VIC 3004, Australia

[7] School of Chemistry and Molecular Biosciences, Australian Centre for Ecogenomics, University of Queensland, St Lucia, QLD 4072, Australia

[8] Department of Plant and Microbial Biology and Center for Computational Biology, University of California, Berkeley, CA 94720, USA

[9] Department of Electrical Engineering and Computer Science, NextGen Precision Health Institute, University of Missouri, Columbia, MO 65211, USA

[10] Department of Medical Sciences, University of Torino, 10126 Torino, Italy

[11] Institute for Integrated and Intelligent Systems, Griffith University, Southport, QLD, 4222, Australia

[12] Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA

[13] Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[14] School of Biology and Basic Medical Sciences, Suzhou Medical College of Soochow University, Suzhou 215123, Jiangsu, China

[15] Computational Biology and Bioinformatics, Université Libre de Bruxelles, 1050 Brussels, Belgium

[16] School of Computing and Information Systems, The University of Melbourne, Melbourne, VIC 3053, Australia

[17] Department of Electrical and Computer Engineering Texas, A&M University, College Station, TX 77843, USA

[18] Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada

[19] 3 Billion, Seoul, South Korea

[20] Biophysics Graduate Group, University of California, Berkeley, Berkeley, CA 94720, USA

[21] Center for Computational Biology, University of California, Berkeley, Berkeley, CA 94720, USA

[22] Computational Genomics Platform, IRCCS University Hospital of Bologna, 40138 Bologna, Italy