

# 一种用于量化蛋白质中错义 突变的系统性影响的结构系 统生物学方法

张屹  
蔡景共

# 意义

- 测量非同义单核苷酸多态性的系统性影响在寻求个性化用药上是一个有重要意义的课题。
- 然而，去了解蛋白质结构水平上的变化如何最终影响细胞行为是一个非常困难的任务。这是因为在蛋白质和通路水平上复杂的信息被整合在一起。
- 由于被整合在一起的蛋白质和路径动力学来评价蛋白质上的错义突变的系统性影响的方法的实用性仍然是未知的。
- 因此通过数学模型来研究这种方法的实用性，并用实验数据与其相比较来研究错义突变。
- 事实上，在过去的二十年中通过数学模型方法的应用对细胞行为的理解已经显著提高，如常微分方程（ODE）和基于规则的模拟。

- 然而，整合蛋白质的动力学方面和它们的相关路径来调查错义突变的系统性影响依然在发展的早期。
- 2007年，Stein等人提议整合结构和路径信息以估算与生物学途径相关的关键动力学常数。
- 更近的，Kiel和Serrano研究c-Raf的Ras结合结构域（RafRBD）中的错义突变如何影响下游蛋白Erk的表达，通过调查的结构和构建一个描述Erk信号路径的ODE模型。
- Kiel和Serrano的工作表明整合在一起的蛋白结构分析和路径建模可以作为一个有效的方法来解释蛋白质中错义突变的生理学意义。

- 这里展示和讨论的是“系统影响因子”（SIF）的测量，其定义为自由能的改变（ $\Delta\Delta G$ ）和系统性控制（ $C_{pi}^S$ ）的一个功能，作为一个实用的方法来评估在特异系统中错义突变的相关效应。
- 对于出现在蛋白质的突变，其复合状态和非复合状态都被考虑进这个模型，我们通过取两种状态之间的最大 $\Delta\Delta G$ 来计算其的最大SIF。同时比较不同的有或没有两种状态的蛋白质使SIF值更具可比较性。



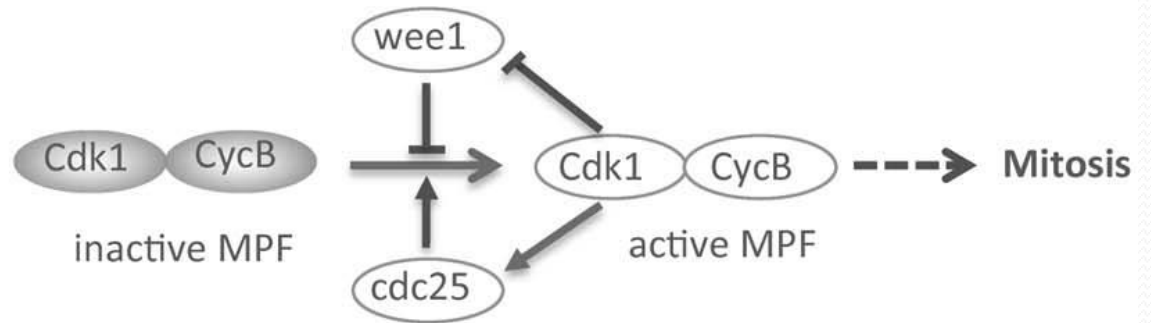
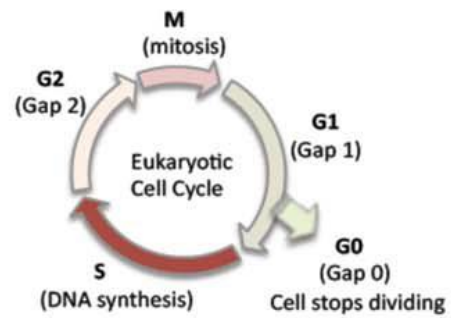
# 方法

- 本实验介绍两个案例研究：
- 解释裂殖酵母G<sub>2</sub>期到有丝分裂期（G<sub>2</sub>-M）过渡调控机制内突变的系统干扰。
- 与促分裂原活化蛋白激酶（MAPK）通路的突变相联的神经相关人类疾病的表型分类。

- 第一个基准系统是一个定义更明确的系统，因为错义突变的系统干扰可以被直接联系到酵母细胞的长度变化以研究基因型与表型的关系。
- 本实验用温度敏感型酵母菌株作为实验模型，其每个蛋白Cdk1或CycB（细胞周期素）都含有一个错义突变，在间隔期（间隔期紧随着分裂期）测量其细胞长度。
- 最后，在量化错义突变的系统性影响中的SIF值的实用性将通过计算SIF值和细胞长度的相关性来评估。

- G<sub>2</sub>-M过渡期控制细胞何时进入有丝分裂以及决定细胞在分裂为两个子细胞时的大小。
- 在裂殖酵母中，G<sub>2</sub>-M过渡期涉及Cdk<sub>1</sub>，CycB，Wee1和Cdc25。在G<sub>2</sub>期，Cdk<sub>1</sub>和CycB组成一个已知复合物作为有丝分裂促进因子（MPF），其引发G<sub>2</sub>-M过渡。
- MPF的活性通过蛋白激酶Wee1和蛋白磷酸酶Cdc25来调节：Wee1抑制MPF的活性通过磷酸化Cdk<sub>1</sub>，而Cdk<sub>1</sub>也通过磷酸化来负反馈调节Wee1。此外，Cdc25通过磷酸化激活MPF，反之亦然。
- Wee1-MPF-Cdc25控制系统增加了MPF的激活状态比率大于其非激活状态，最终促进了细胞进入有丝分裂。

(A)



(B)

$$\frac{d}{dt} \text{CycB} = k_s - k_d \text{CycB}$$

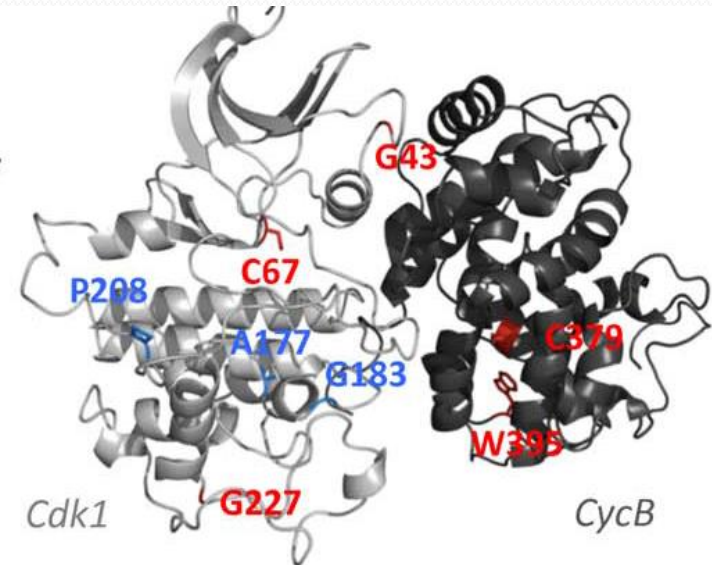
$$\frac{d}{dt} \text{MPF} = V_{25} \cdot (\text{CycB}_T - \text{MPF}) - V_{wee} \cdot \text{MPF} - k_d \text{MPF}$$

$$\frac{d}{dt} \text{Wee1} = \frac{k_{awee} \cdot (1 - \text{Wee1})}{J_{awee} + 1 - \text{Wee1}} - \frac{k_{iwee} \cdot \text{MPF} \cdot \text{Wee1}}{J_{iwee} + \text{Wee1}}$$

$$\frac{d}{dt} \text{Cdc25} = \frac{k_{a25} \cdot \text{MPF} \cdot (1 - \text{Cdc25})}{J_{a25} + 1 - \text{Cdc25}} - \frac{k_{i25} \cdot \text{Cdc25}}{J_{i25} + \text{Cdc25}}$$

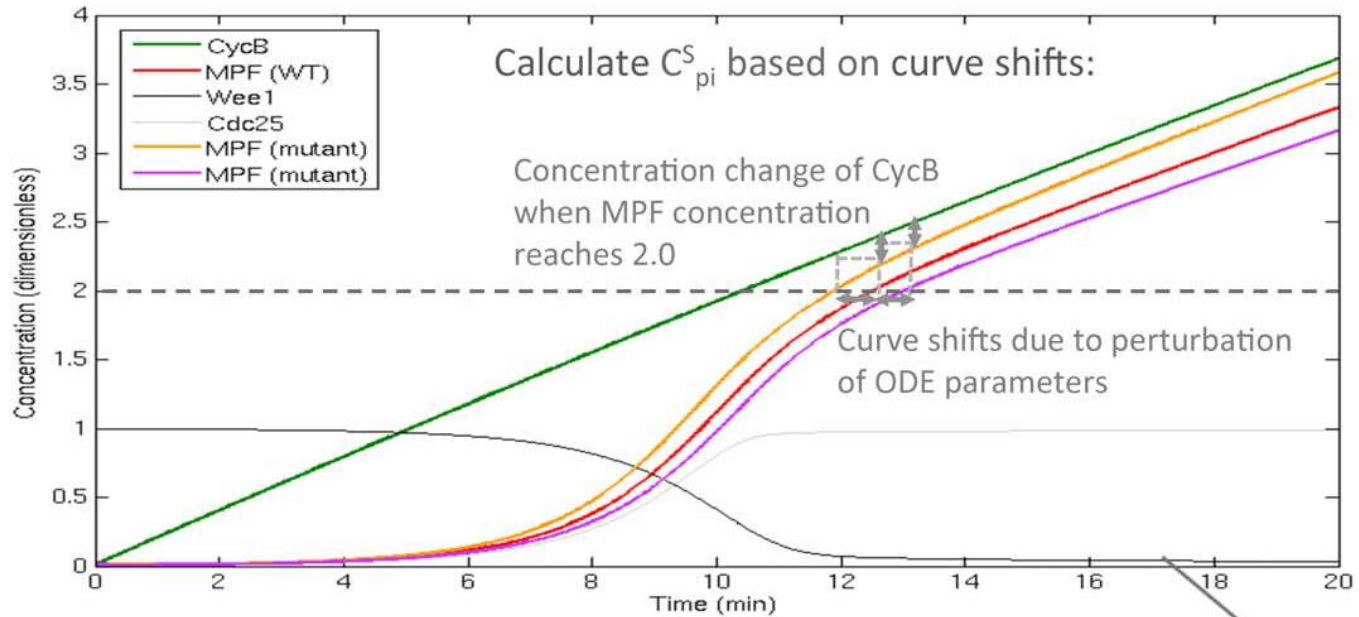
$$V_{25} = k'_{25} \cdot (1 - \text{Cdc25}) + k''_{25} \cdot \text{Cdc25}$$

$$V_{wee} = k'_{wee} \cdot (1 - \text{Wee1}) + k''_{wee} \cdot \text{Wee1}$$



Structural analysis and  $\Delta\Delta G$  calculation of individual mutations

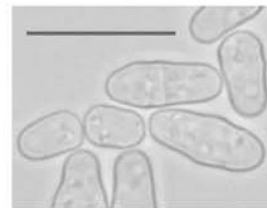
(C)



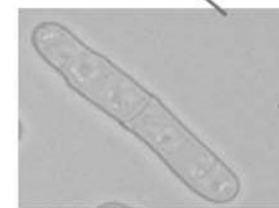
(D)

$$SIF \sim \Delta\Delta G \cdot C_{pi}^S$$

Inferring systemic consequences



Mutations with smaller SIF scores



Mutations with larger SIF scores



**Table 1.** Differential equations of the G2-M model.

$$\frac{d}{dt} \text{CycB} = k_5 - k_d \cdot \text{CycB}$$

$$\frac{d}{dt} \text{MPF} = V_{25} \cdot (\text{CycB}_T \cdot \text{MPF}) - V_{wee} \cdot \text{MPF} - k_d \cdot \text{MPF}$$

$$\frac{d}{dt} \text{Wee1} = \frac{k_{wee} \cdot (1 - \text{Wee1})}{J_{wee} + 1 - \text{Wee1}} - \frac{k_{wee} \cdot \text{MPF} \cdot \text{Wee1}}{J_{wee} + \text{Wee1}}$$

$$\frac{d}{dt} \text{Cdc25} = \frac{k_{a25} \cdot \text{MPF} \cdot (1 - \text{Cdc25})}{J_{a25} + 1 - \text{Cdc25}} - \frac{k_{d25} \cdot \text{Cdc25}}{J_{d25} + \text{Cdc25}}$$

$$V_{25} = k'_{25} \cdot (1 - \text{Cdc25}) + k''_{25} \cdot \text{Cdc25}$$

$$V_{wee} = k'_{wee} \cdot (1 - \text{Wee1}) + k''_{wee} \cdot \text{Wee1}$$

*Initial conditions (dimensionless):*

$$\text{CycB} = 0.01; \text{MPF} = 0.01; \text{Wee1} = 1.0; \text{Cdc25} = 0.01$$

$$k_5 = 0.2; k_d = 0.008;$$

$$k'_{25} = 0.008; k''_{25} = 0.89; k'_{wee} = 0.03; k''_{wee} = 0.18;$$

$$k_{wee} = 0.61; k_{iwee} = 0.71; k_{a25} = 0.80; k_{d25} = 0.35$$

*Michaelis constants (dimensionless):*

$$J_{wee} = 0.90; J_{iwee} = 0.21; J_{a25} = 0.19; J_{d25} = 0.93$$

- 该试验展示的模型（表1）是基于MPF活性的仿真模型。
- 模型有两个基本假定：
- 系统中Cdk1 (Cdk1<sub>T</sub>)的总数是一定的和过量的（远大于CycB的最初浓度）。
- 所有的CycB与Cdk1在其合成之后形成复合物，Cdk1牢固地结合于CycB又多于CycB：即， $CycB_T = MPF$ （MPF激活态）+ pre-MPF（MPF非激活态）。
- 该试验考虑将每一个错义突变作为一个对野生型的状态的干扰。每一个突变的系统性影响被设计为一个突变可能偏离野生型状态的程度。在该试验中，当CycB达到的浓度（无量纲）相当于活性MPF浓度为2.0时，进入有丝分裂就会发生。

- 假设细胞在间期不断地线性地生长，每个速率常数干扰的系统性影响通过当活性MPF= 2.0时CycB浓度的变化可以量：
- CycB浓度越高，有丝分裂的细胞就越长，反之亦然（图1C）。 $C_{pi}^S$ 表示CycB浓度的变化在每个速率常数的干扰的一致数量之下。 $C_{pi}^S$ 的符号表明CycB浓度变化的方向：正的 $C_{pi}^S$ 值相当于CycB浓度的增加，而负的 $C_{pi}^S$ 值表示相反的意思。
- $C_{pi}^S$ 的整体视图为G2-M期机制的计算表明，干扰CDC25相关的速率常数对MPF曲线有较大的影响，与相比Wee1相关的速率扰动常数相比（如图S2）。这表明了MPF激活的正面和负面的反馈回路之间的非对称关系，分别由CDC25和Wee1控制。

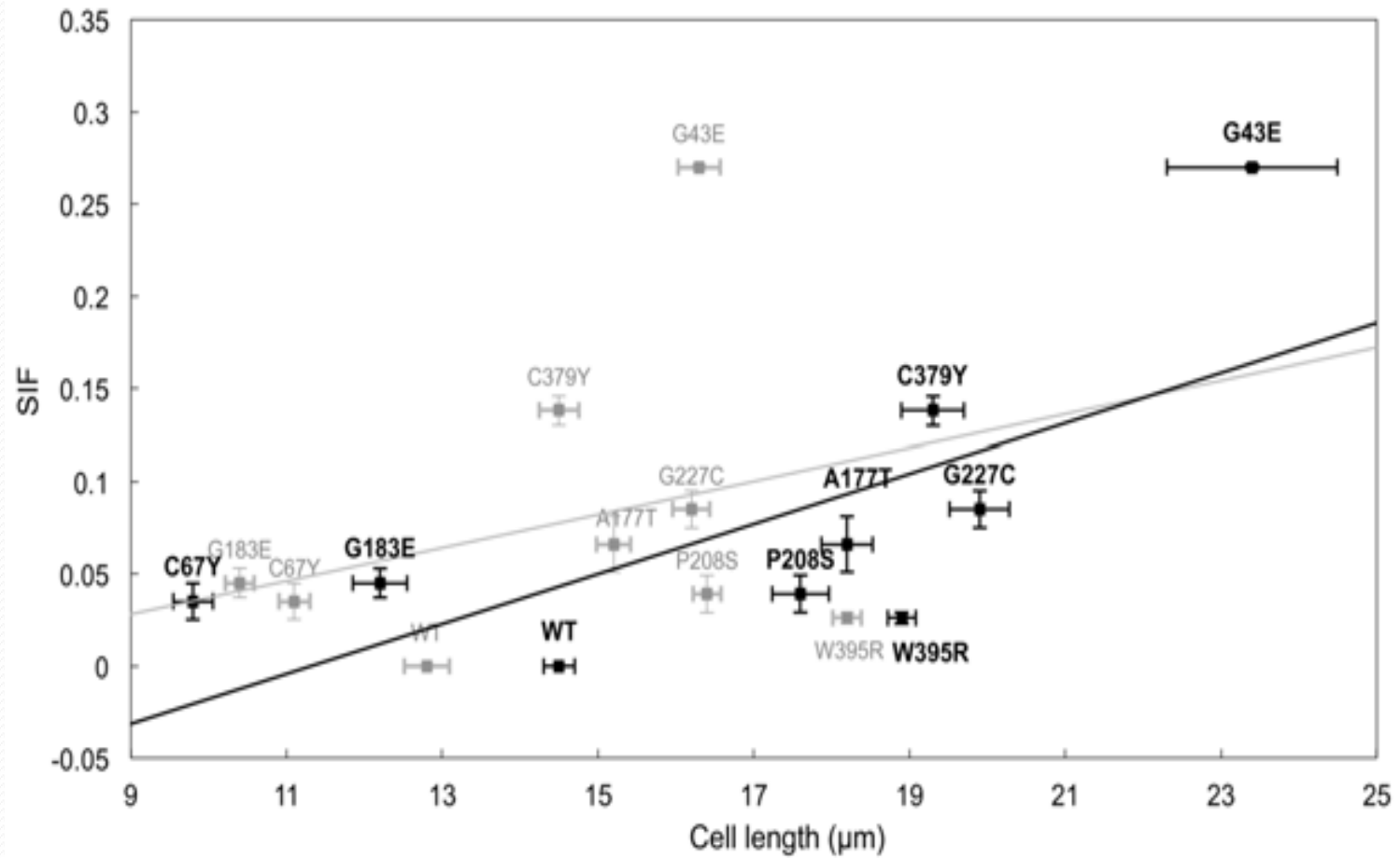
**Table 2.** *In vivo* length of the yeast trains in the G2-M model.

Strain Number	Strain Name	Mutated Protein	Residue Change	Cell Length ( $\mu\text{m}$ )		Cell Length ( $\mu\text{m}$ )	
				25°C		30°C	
				Mean	Stdev	Mean	Stdev
275	M85	Cdk1	G48F	16.3	1.5	18.4	6.0
468	3W	Cdk1	L87Y	11.1	1.1	9.8	1.4
8	3J	Cdk1	A177T	15.2	1.2	18.2	1.8
154	36YD0	Cdk1	G183E	10.4	1.0	12.2	1.9
374	17	Cdk1	D200S	16.4	1.0	17.6	2.0
515	M63	Cdk1	G277C	16.7	1.3	19.9	2.1
6	NA	CycB	L179Y	14.3	1.4	19.3	2.2
4052	NA	CycB	W205R	18.2	1.0	18.9	1.0
572	WT	NA	NA	12.8	1.6	14.5	1.1

doi:10.1371/journal.pcbi.1002798.t002

**Table 3.** *In silico* measurements of the mutant cells in the G2-M model.

Amino acid change	Target Protein	Impact Type <sup>a</sup>	$\Delta\Delta G$ (Cdk1/CycB) <sup>b</sup>	$\Delta\Delta G$ (MPFF) <sup>c</sup>	Maximum $\Delta\Delta G$ (kcal/mol) <sup>d</sup>		$C_{vi}^e$		SIF
					$k_{vi}$ (CycB) <sup>f</sup>	$k_{vi}$ (Cdk1) <sup>f</sup>	$J_{mut}/J_{wt}$ <sup>g</sup>		
G43E	Cek1	S	2.10	249 <sup>h</sup>	24.9	-	0.011	-	0.27
G67Y	Cek1	S	3.17	1.31	3.17	-	0.011	-	0.035
A177T	Cek1	F	5.97	3.65	5.97	-	-	0.011	0.066
G183C	Cek1	F	3.72	4.13	4.13	-	-	0.011	0.045
F208S	Cek1	F	3.56	7.35	3.56	-	-	0.011	0.039
G227C	Cek1	S	2.89	7.25	7.69	-	0.011	-	0.085
L329Y	CycB	S	31.93 <sup>h</sup>	34.58 <sup>h</sup>	34.56	0.004	-	-	0.138
W395R	CycB	S	6.57	6.15	6.57	0.004	-	-	0.076





- SIF和系统干扰（SP）之间的联系，可以统计建立回归： $SP \sim SIF \propto \Delta\Delta G \cdot C_{pi}^S$ 。
- 方程中 $\Delta\Delta G$ 是由靶蛋白的一个突变引起的自由能变化。
- $C_{pi}^S$ 是控制系数，反映蛋白（该试验中蛋白质CycB）的浓度变化的敏感程度是特定的参数。
- 因此，对于G<sub>2</sub>-M模型的每个SIF值的量级表示影响突变的程度可以反映在CycB的数量上，其决定细胞何时进入有丝分裂，从而决定细胞的长度。

- G<sub>2</sub>-M过渡（表2）的系统干扰是通过检查四个在Cdk1（C67Y和G183E突变除外）的温度敏感突变和在CycB的两个温度敏感突变MPF的功能衰减的效果来研究的。
- 当温度升高时,这些突变对蛋白质的稳定性或功能的影响是更明显的,且使细胞在一个比其野生型状态更大的长度时分裂。

- MPF的模型结构表明，在Cdk1的突变G43E是位于MPF的界面，从而有可能对MPF复合体的稳定有显著影响。
- 在Cdk1的突变A177T，G183E及P208S位于或靠近活性位点，因此有可能造成功能的影响；在Cdk1的C67Y，G227C和在CycB的突变W395R在这些蛋白质的外围，因此主要是结构上的相关。
- 在CycB的突变C379Y内的疏水核心，可能通过使CycB的结构不稳定而对MPF复合体有相当大的影响（图1B）。

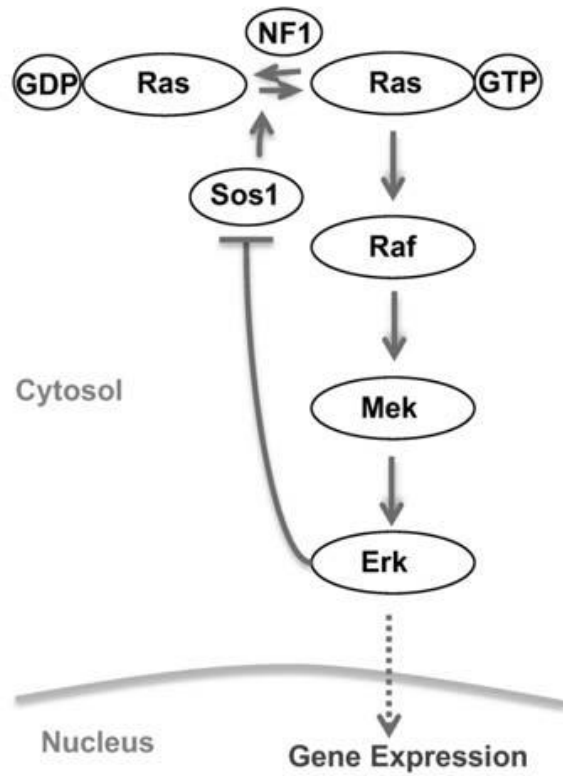
- 首先，一个目标生物系统（在这种情况下，在细胞周期的G<sub>2</sub>-M过渡）被选择，而报告蛋白的表达谱能用于衡量系统性的行为，是确定的。
- 这里使用的报告蛋白是MPF蛋白复合物。然后，基因突变映射到三维蛋白质结构和链接到在ODEs中的相关参数。对于每一个突变，其 $\Delta\Delta G$ 近似为引入相关联的ODE参数的干扰的大小。
- 然后计算出一个基于采样构象的平均 $\Delta\Delta G$ 接着，报告蛋白的表达谱的灵敏度通过计算 $C_{pi}^S$ 探讨每一个ODE参数。
- 最后，每一个突变的系统性后果是通过在 $\Delta\Delta G$ 和 $C_{pi}^S$ 的基础上计算其SIF得分，来推断的。在G<sub>2</sub>-M模型的情况下，较大的SIF反映了一个更大的延迟一个细胞进入有丝分裂。

- 第二个基准系统是一个更复杂的例子，其目标突变分布在不同的蛋白中（H-Ras,Ras-1,Braf和Me），导致对临床症状（此例中为神经心脏面部皮肤综合征）预测的复杂化。
- 为了确定一个简单的ODE模型能否被用于推断特异性突变的系统性干扰，该实验构建了一个只包括12个参数来计算值的简化的ODE模型。
- 然后，根据其预测SIF值把突变分为亚组，记录该分类是否能解释不同疾病的根本机制。

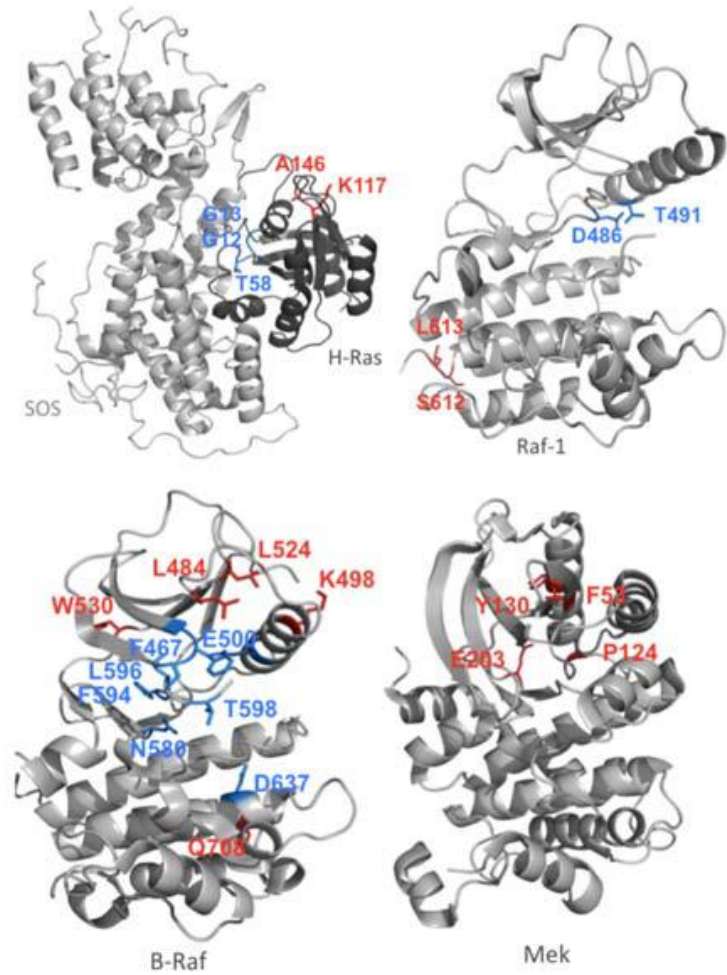
- MAPK信号通路在细胞的生长，增殖，分化（图3A）中起着至关重要的作用。
- 它的三层MAPK级联，即Raf-Mek-Erk，是一种高度保守的系统结构，调节通路的信号转导机制的类开关行为。
- 这个级联重要特点表现为，其自己作为代表来评估亲本通路的行为。



(A)

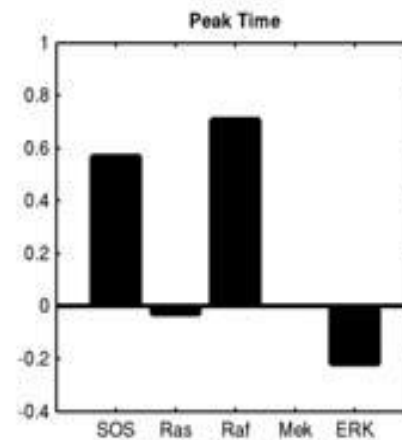
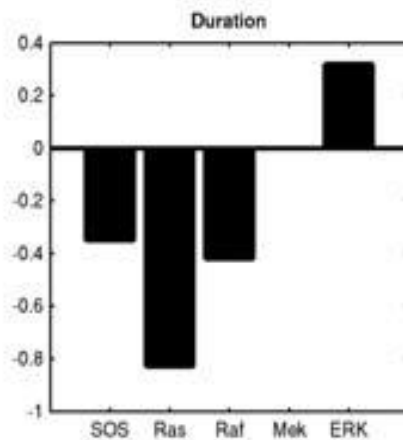
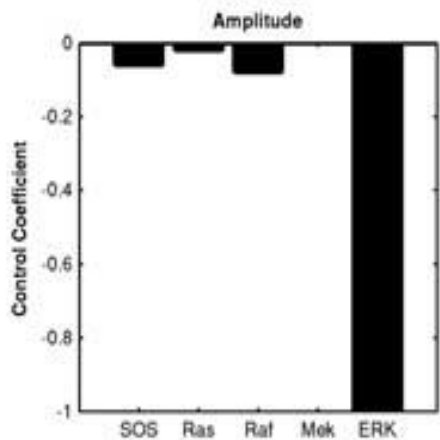


(B)

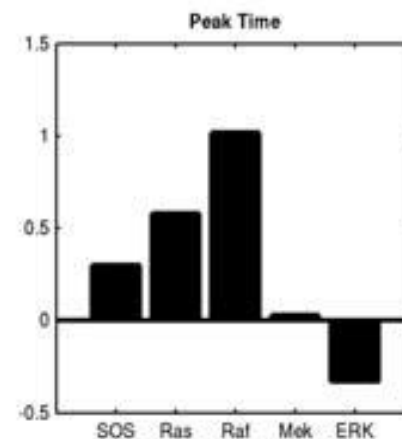
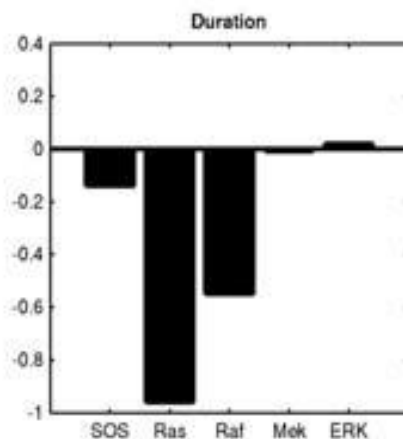
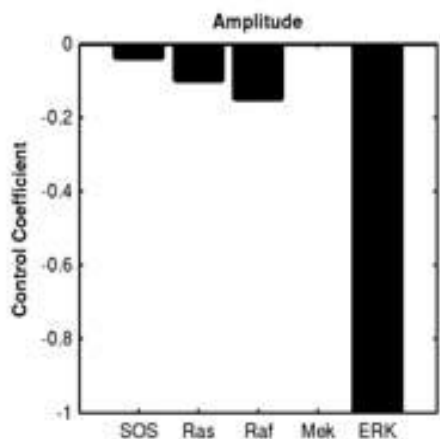


- 为了基准简化和初始模型的行为，该试验进行了三个目标蛋白Erk的分析：振幅（最大激活），持续时间（直到下降至最大激活的50%）和峰值时间（最大激活的时间）。
- 关键蛋白的初始浓度在两种模式是多种多样的，他们对控制目标蛋白Erk的效果被比较（关键的蛋白质包括ShcGS（Shc: Src同源蛋白和胶原蛋白结构域蛋白），GS, Grb2（生长因子受体结合蛋白2），SOS（交换因子同源蛋白），RAS, RAF, MEK和ERK）。
- 其结果是，在这两种模式的控制系数表现出同样的模式跨所有三个目标数量（图4A, B），这表明该简化模型不牺牲初始模型的整体动态的来实现它的简单结构。

(A)



(B)



**Table 4.** Differential equations of the reduced MAPK model.

$$\frac{d}{dt} \text{ShcGS} = -c_{12} \cdot \text{Erk}^* \cdot \text{ShcGS} \left( 1 - \frac{\text{RasGDP}}{c_{11} + \text{RasGDP}} \right)$$

$$\frac{d}{dt} \text{RasGDP} = \frac{c_{10} \cdot \text{ShcGS} \cdot \text{RasGDP}}{c_{11} + \text{RasGDP}} - c_1 \cdot \text{RasGTP}$$

$$\frac{d}{dt} \text{RasGTP} = \frac{c_{10} \cdot \text{ShcGS} \cdot \text{RasGDP}}{c_{11} + \text{RasGDP}} + c_1 \cdot \text{RasGTP}$$

$$\frac{d}{dt} \text{Raf} = -c_2 \cdot \text{Raf} \cdot \text{RasGTP} + c_3 \cdot \text{Raf}^*$$

$$\frac{d}{dt} \text{Raf}^* = c_2 \cdot \text{Raf} \cdot \text{RasGTP} - c_3 \cdot \text{Raf}^*$$

$$\frac{d}{dt} \text{Mek} = -c_4 \cdot \text{Raf}^* + c_5 \cdot \text{Mek}^*$$

$$\frac{d}{dt} \text{Mek}^* = c_4 \cdot \text{Raf}^* - c_5 \cdot \text{Mek}^*$$

$$\frac{d}{dt} \text{Erk} = -\frac{c_6 \cdot \text{Erk} \cdot \text{Mek}^*}{c_7 + \text{Erk}} + \frac{c_8 \cdot \text{Erk}^*}{c_9 + \text{Erk}^*}$$

$$\frac{d}{dt} \text{Erk}^* = \frac{c_6 \cdot \text{Erk} \cdot \text{Mek}^*}{c_7 + \text{Erk}} - \frac{c_8 \cdot \text{Erk}^*}{c_9 + \text{Erk}^*}$$

Initial conditions (molecules cell<sup>-1</sup>):

ShcGS = 20,000; RasGDP = 20,000; RasGTP = 0; Raf = 10,000;

Raf\* = 0; Mek = 360,000; Mek\* = 0; Erk = 750,000; Erk\* = 0

Rate constants (molecules<sup>-1</sup> cell min<sup>-1</sup>):

c<sub>2</sub> = 7.7 · 10<sup>-4</sup>; c<sub>3</sub> = 8.3; c<sub>4</sub> = 4 · 10<sup>5</sup>; c<sub>10</sub> = 15; c<sub>12</sub> = 4 · 10<sup>-6</sup>

Rate constants (molecules cell<sup>-1</sup>):

c<sub>7</sub> = 9 · 10<sup>4</sup>; c<sub>9</sub> = 6 · 10<sup>5</sup>; c<sub>11</sub> = 1.53 · 10<sup>3</sup>

Rate constants (min<sup>-1</sup>):

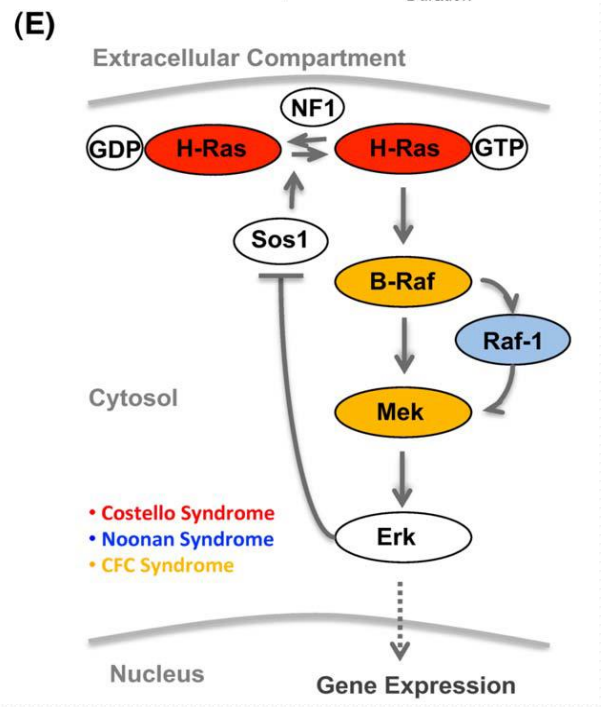
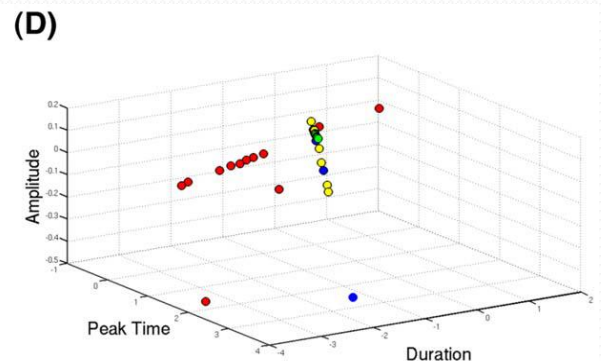
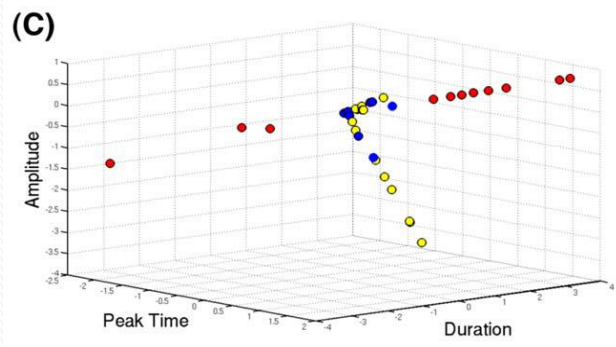
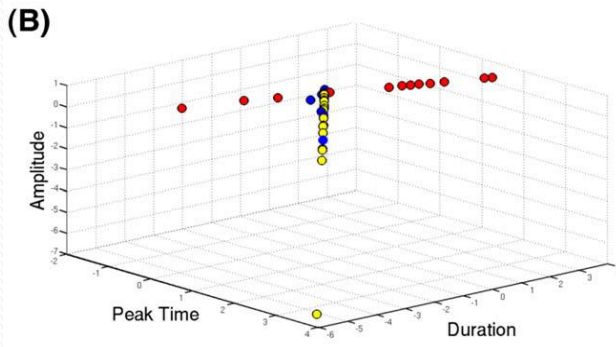
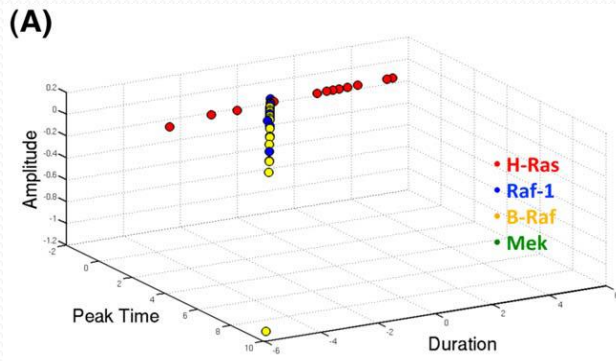
c<sub>1</sub> = 69; c<sub>3</sub> = 14; c<sub>4</sub> = 50; c<sub>5</sub> = 0.78

- 通过收集和研究与神经心面部皮肤综合征相关的40个突变（表S1），如图3B所示，所有的突变可以被映射到的H-ras, Raf-1, B-Raf和MEK晶体结构，每一个突变被归类为功能上的突变或结构上的突变主要根据它在靶蛋白的位置。
- 不同于酵母G2-M模型的错义突变，对MAPK信号转导途径中的突变生理结果的非定量测量，可以被用来计算相关SIF值。

- 因此，作为以间接的方式来评价突变和临床症状之间的关系的方法，每个突变被表示为通过根据对野生型Erk表达曲线：测量的幅度，持续时间和高峰时间的差异，的系统性影响，计算的3个SIF值。
- 与每一个突变相对应的SIFs的轨迹作为这三个目标数量的功能，在Raf1，B-Raf和MEK的突变更可能被类似的区域重叠，而在H-ras的突变往往分布在一个非常不同的轨迹相对于其他的突变方向（图5A）。



- 如图5B和5C所示，两个参数表示H-ras突变的轨迹不同于其他突变，这表明，H-ras分离对初始浓度和参数空间的变化是不敏感的。作为一个基准，来自原始模型的三维SIF分数也呈现了（图5D）。一致的是，H-ras突变被分配到一个完全不同的组。
- 突变SIF分数的仔细检查显示，H-ras 突变干扰了MAPK通路用一个完全不同于在Raf-1, B-Raf和MEK（图5A-D）中的突变的方式：HRAS突变倾向于显性影响ERK的表达谱的持续时间，而其他的突变主要影响表达谱的振幅，后跟对峰值的时间影响较小和一个更小的效果对ERK活化的持续时间。生理上，这表明HRas突变的细胞响应是不同与其他突变。



# 分析

- 在这项工作中，该实验提出了SIF功能作为一种有效的测量对系统性影响的错义突变。
- SIF值以简单的方式反映的事实，即蛋白质是在细胞中功能单元，其相互网络调节细胞行为。
- SIF值反映酵母G<sub>2</sub>-M模型的体内表型，没有额外的参数在功能和结构上地区分重要的突变。
- 这表明，虽然它们以不同的方式改变蛋白的行为，功能和结构的突变会干扰途径到类似程度。

# 讨论

- 该实验表明，结合多级知识来调查错义突变在细胞的行为中的影响是有利的。
- 蛋白质结构预测技术的进步将特别地造成SIF的计算更加可行，因为它需要承载目标错义突变的蛋白质的结构信息。
- 通过ODE模拟通路动力学，提供了一个方便的平台来利用蛋白质结构的信息。
- 但是，ODE的应用意味着两个主要限制：一个是在公共资源的蛋白表达的时间过程数据的可用性，另一个是目前这是相比相对低的和稀疏的基因表达数据。随着越来越多的高通量的时间过程数据变得可用，这将缓和。