

XXIII РОССИЙСКИЙ НАЦИОНАЛЬНЫЙ КОНГРЕСС «ЧЕЛОВЕК И ЛЕКАРСТВО»

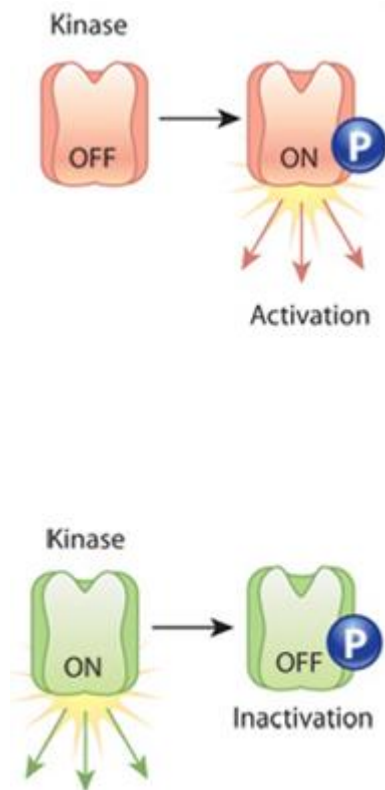
**Определение аминокислотных остатков,  
влияющих на селективность ингибиторов  
к протеинкиназам человека**

**Карасев Д.А.**

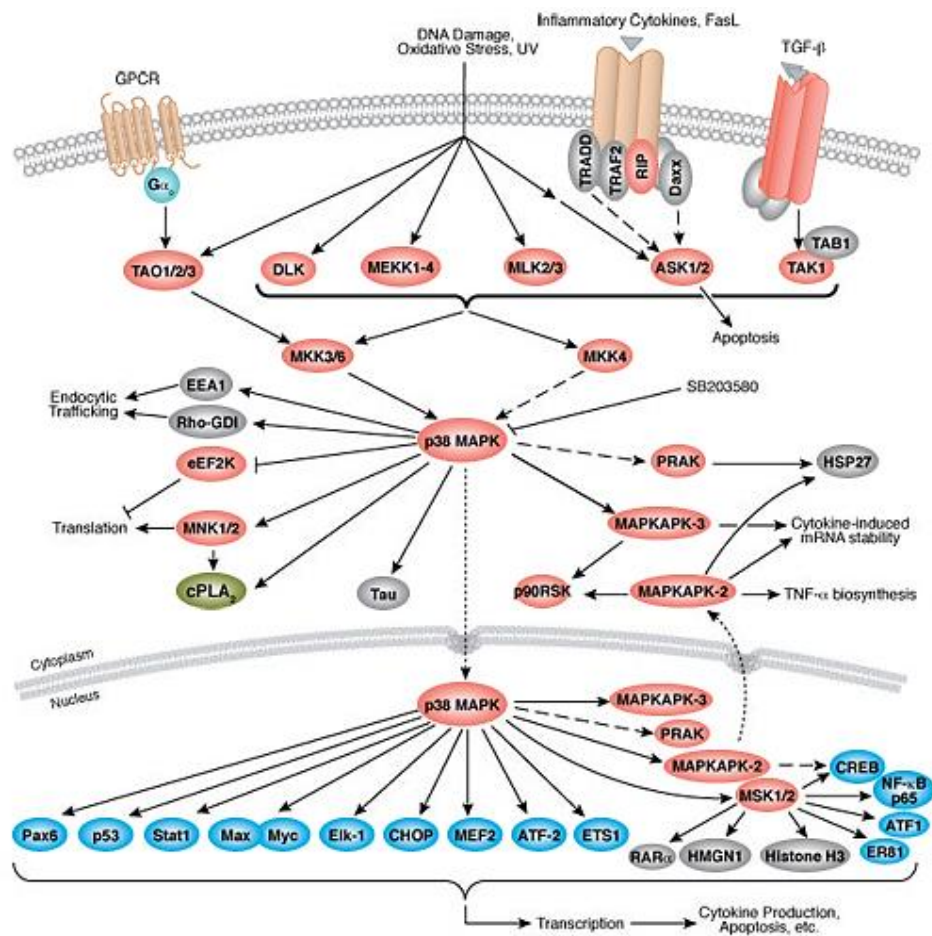
РНИМУ им. Н.И. Пирогова  
ИБМХ им. В.Н. Ореховича

# Протеинкиназы – ключевые звенья регуляции многих физиологических процессов

Обратимое фосфорилирование универсальный механизм регуляции активности белков

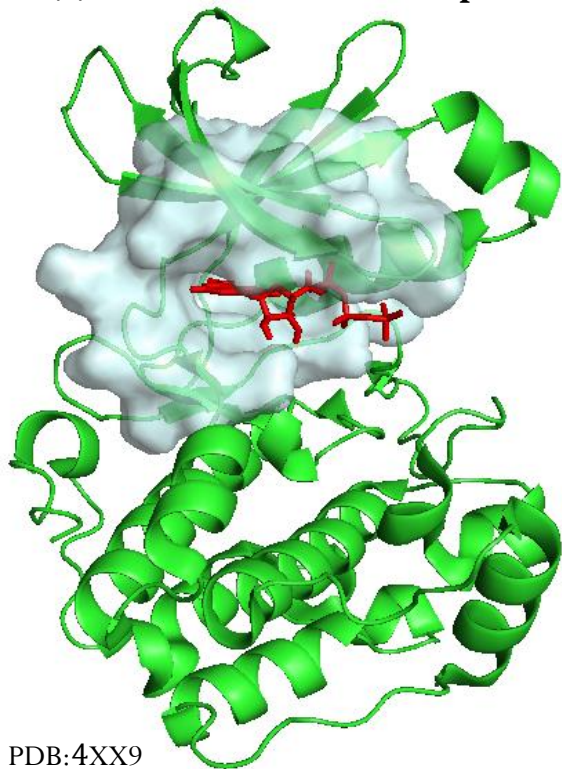


## Регуляторный каскад MAPK киназ



# Селективность протеинкиназ к низкомолекулярным ингибиторам

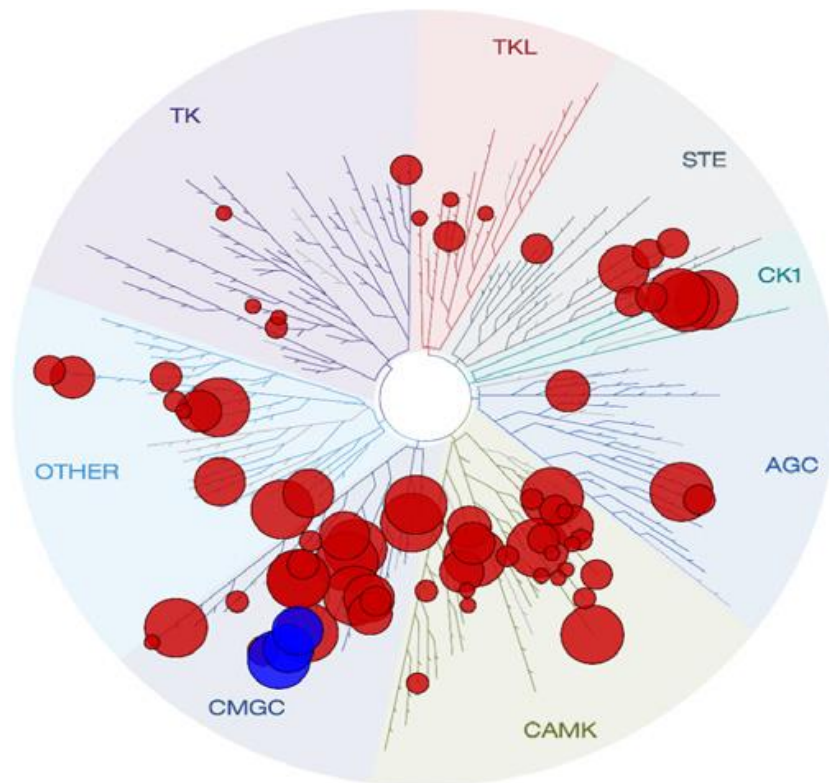
АТФ-связывающий карман привлекательная мишень для дизайна ингибиторов



PDB:4XX9

Протеинкиназа PDK1 с АТФ в АТФ-связывающем кармане.

Селективность протеинкиназ к слабо коррелирует с филогенией

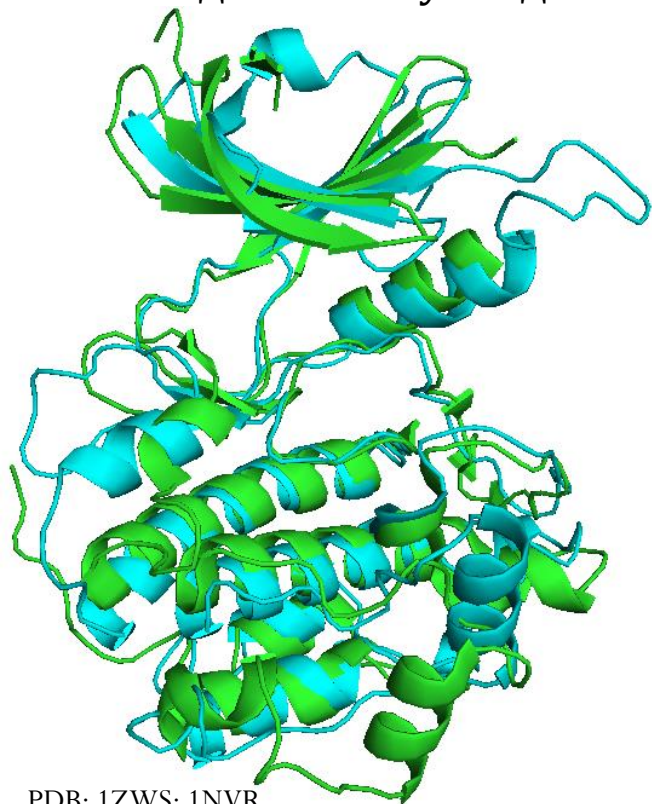


Ингибирования различных членов киназного семейства низкомолекулярным ингибитором

*Palmer et al. 2013, Bioorg Med Chem Lett.;23(5):1486-92*

# Исследование селективности протеинкиназ *in silico*

Протеинкиназы имеют сходный тип укладки



PDB: 1ZWS; 1NVR

RMS = 1.520

Сходство последовательностей - 43.2%

Фрагмент множественного выравнивания киназного домена 200 протеинкиназ человека.

```
QVYDQLVRIAKVGGTEELYGYLKKYHIDLDPRFNDILGHSRKRWFNFIHSEN-----RHLYSPEALDLDKLLRYDHCQRLLTAKEAMERPY-----
GVDQLVEIIVKVLGTPTRREQIREMNNENYTEEFKQIKAHF---WTKVKEKSR-----TPPEAIALCSSLLEVTTPSSRLSPLEACANSE-----
SKHEHLAMMERILGPIFQHMIOKTRKRKYFHNDRLDDEHSSAGRYVRRRCKP-----LKEFMLCHD-----EEHEKFLDLVRRMLEVDFTQRIITLDEALQHPF-----
SKHEHLAMMERILGPIPKHMIQKTRKRKYFHNDRLDDEHSSAGRYVRRRCKP-----LKEFMLSD-----VEHERLFDLQKMLEVDPAKRITLREALKHPF-----
NREHLAMMERILGPIFSRMIRKTRKQKYFVRGRLDMDENTSAGRYVRENCKP-----LRYVLTSEA-----EEHHOLFDLIESMLEVEPAKRLLTGEALQHPF-----
EYDQMNRIIVEVLGIFFAAMLDCQPKARKYERLPGGGVLRRLRKELRKDYQGPGTRRLOEVLGVITGGPGGRBAGEPGHSPADYLRFDLVLRMLEVEPAARISPLGALQHG-----
-DIPFEHDEEILRQVFFRQR-----VSSEQHLLRWCLALRPSDRPTFEEIQNHPW-----
-DIPFEQDEEILRGRLLFRRR-----VSPECCQLRWCLSLRPSERPSLDQIAAHP-----
-DIPFERDQEIIEAELHFPAH-----VSPECCALIRRCLEPKFSRPSLEEILLDFW-----
PYDSMGTPFQQLKQVVEEPPQLP-----ADKFSAEFVDFTSCLKKNKSKERTYPPELMOHPF-----
PYESMGTPFQQLKQVVEEPPQLP-----ADRFSPPEVDFTAQCLRKNPAERMSVLELMEHPF-----
PYEKQNSVFDQLTQVVKGDPPQLSN-----SEERFSPSFLNFVNLCLEKDESKRPKYKELLKHPF-----
PPYAT-SIFQLVSLILKDPVWRWPS-----TISPCFKNFLQGLLTKDPRQLRSLWPDLLYHPF-----
PFEDGAKLRIVNGKYS---IPPHD-----TQYTVFHSLIRAMLQVNPPEARLSIAEYVH-----
PFGE-SQTAICDGNFT---IPDNS-----RYSQDMHCLIRYMLEPDPKRFDIYC-----
AFEAGSMKNLVLKIISGFFPVVSL-----HYSYDLRSLVSQLEKRNPRDRPVSNSILEKG-----
PYPGMPVDSKFKYKMKKEGFRMLSPE-----HAPAEMYDIMKTCDADPLKRFTEKQIVQ-----
PYPGIFVDANFYKLIQNGFKMDQPF-----YATEEYIIMQSCMAEDSRKRPSEFNLTSE-----
PYPELPTNEQFYNAIKRGYRMAQPA-----HANSDEIYEIMQKWEKEEIRPPFSQLV-----
PYPGVEN-SEIVDYLRQGNRLKQPA-----DCLDGLYALMSRCWELNPDQRPSFTELRED-----
PYPGVQN-HEMYDYLLHGHRLLKQPE-----DCLDELYEIMYSCMRDPLDRPTFSVLR-----
PYPDVNT-FDITVYLLQGRRLLOPE-----YCPDPLYEVMLKQWRPKAEMRPSFSELVS-----
SLPFQGDKRKEIMTLILKAKLQMPQ-----FLSTEAQSLLRALKRNPNARLGGSPDGAEELKRN-----
DDYLERPTMDFGLNVRGFLDRYCEP-----NCPSEFEETVRCDDLDPEKRFPSVVKLE-----
-CDTAFETSKFFTLRDGIISDIFD-----KKEKTLQKLLSKKPEDRPNISEIL-----
IINQISKLHALIDPNHEIEFFDIPF-----KDLQDVLKCCLEKRDPKQRIISPELLAHPV-----
RTPDKYKGLTEKEREVYSRCRDPT-----SCKELADMTRCMNDENRDFEPRATMR-----
```

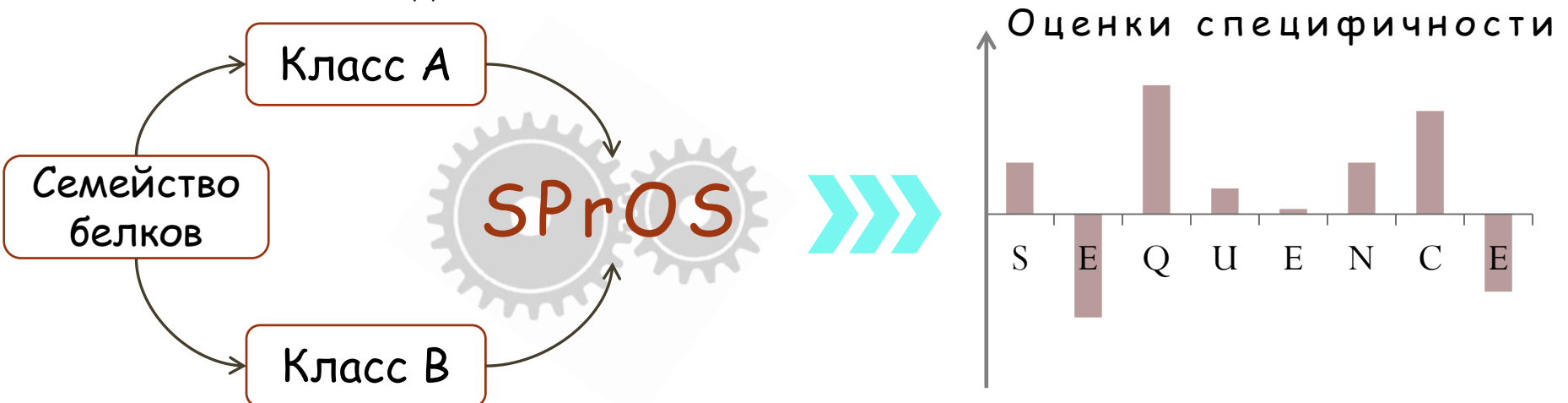
# Цель работы

Разработка метода идентификации аминокислотных остатков, определяющих взаимодействие белков с лигандами, и его апробация при прогнозе ингибиторной специфичности протеинкиназ.

# Локальное сходство аминокислотных последовательностей для поиска специфичных остатков

Метод *SPrOS* - *Specificity Projections on Sequence*

- Аминокислотные последовательности разделяются на два класса в соответствии с исследуемым свойством.
- Позволяет оценить специфичность каждого остатка тестовой последовательности к определенному классу.
- Метод оценки специфичности аминокислотных позиций основан на пофрагментных сравнениях последовательностей обучающей выборки с тестовой последовательностью.



# Локальное сходство аминокислотных последовательностей для поиска специфичных остатков

Все остатки получают оценки в результате сравнений со всеми последовательностями обучающей выборки.

AANRDP SQFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVA	2
ANRDP SQFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVAL	1
NRDP SQFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALR	1
RDPSQFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRA	0
DPSQFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRAL	1
PSQFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALF	2
SQFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFG	1
QFPDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGR	1
FDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRF	2
PDPHRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFP	0
DPHRF DVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPA	1
PHRF DVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPAL	0
<b>HRFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALS</b>	<b>9</b>
RFDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSL	0
FDVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSLG	3
DVTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSLGI	1
VTRDTRGHLSFGQGIHFCMGRPLAKLEGEVALRALFGRFPALSLGID	1
GTAINKPLSEKMMLFGMGKRRCIGEV LAKWEIFLFLAILLQQLEFSV	9

Каждый остаток  $i$  тестовой последовательности получает максимальную оценку  $S_i$  полученную при всех смещениях. Совпавшие дополнительно умножаются на поправочный коэффициент  $w$ .

## Оценки позиционной специфичности к классам обучающей выборки

$$E_{ia} = \frac{a_i - b_i}{a_i + b_i}$$

$S_{ik}$  – Позиционная оценка полученная при сравнении с  $k^{th}$  последовательностью

$E_{ia}$  – оценка специфичности позиции  $i$  к классу  $A$

$a_k$  – коэффициент принадлежности  $k^{th}$  последовательности к классу  $A$

$b_k$  – коэффициент принадлежности  $k^{th}$  последовательности к классу  $B$

$$a_i = \frac{\sum_k S_{ik} a_k}{\sum_k a_k}$$

$$b_i = \frac{\sum_k S_{ik} b_k}{\sum_k b_k}$$

$$-1 \leq E_{ia} \leq 1$$

$$a_k, b_k = 0, 1$$

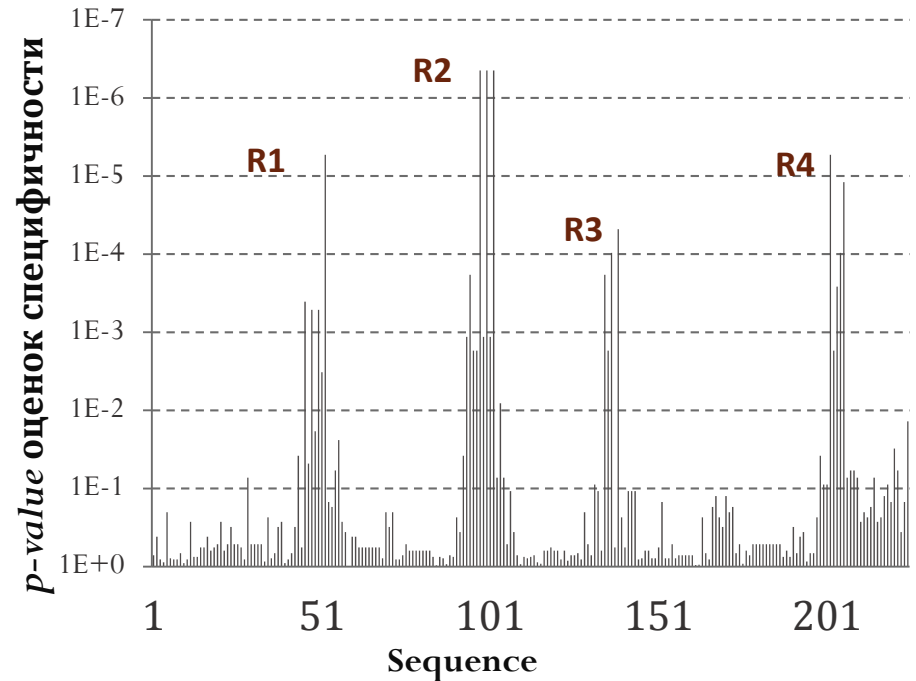
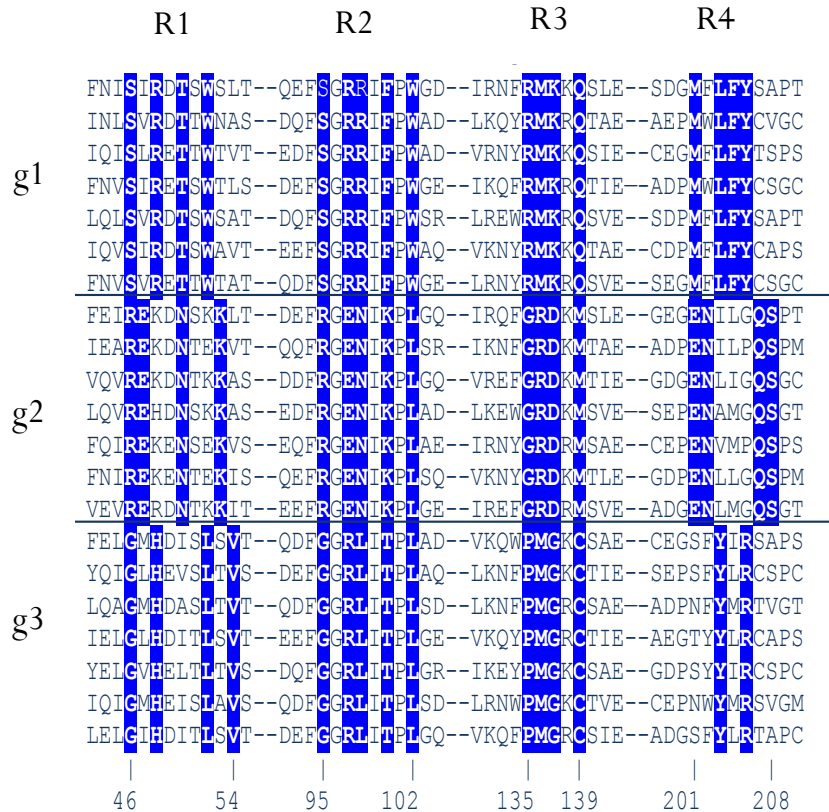
*P-values* рассчитываются в результате «перемешивания» классов обучающей выборки и перерасчете  $E_{ia}$  для рандомизированных классификаций



## Типы тестовых данных

- Модельные аминокислотные последовательности (21) разделенные на три равные группы с искусственно заданными классообразующими заменами.
- Последовательности белков семейства LacI/GalR (125) с четким разбиением на 10 групп в соответствии с распознаваемыми ДНК-мотивами и эффекторными соединениями (Mazin et al. 2010, Algorithms Mol Biol. 15;5:29).
- Последовательности протеинкиназ (184), сгруппированные в соответствии с показателями ингибирования низкомолекулярными лигандами (158) (Gao et al. 2013, Biochem. J., 451 313-328).

# Тестирование на модельных последовательностях



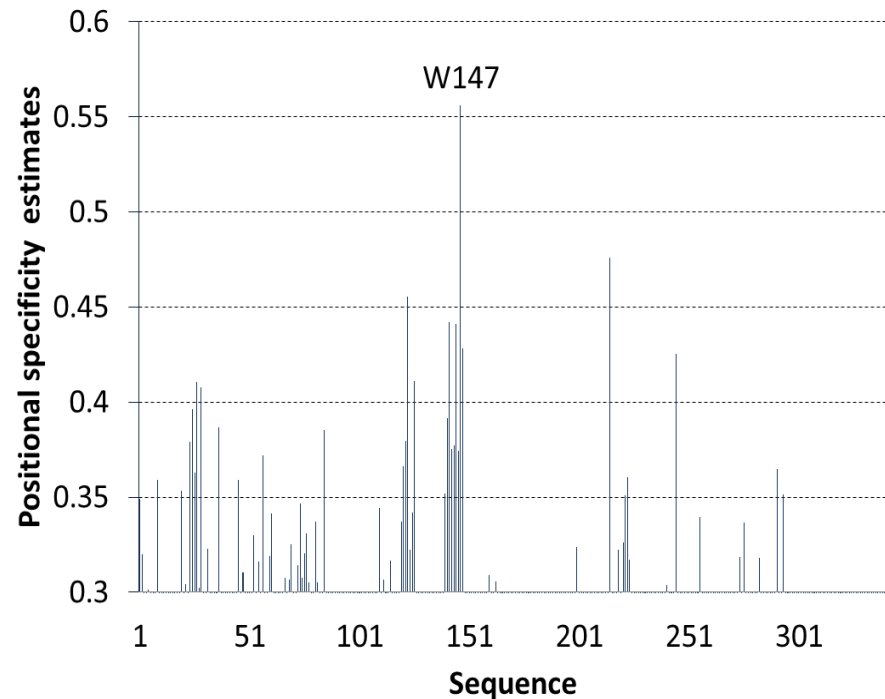
При сравнении последовательностей первой группы (g1) с остальными последовательностями (группы g2 и g3), все внесенные замены были идентифицированы со статистически значимыми оценками ( $p\text{-value} < 10^{-3}$ ) .

# Сравнение результатов прогноза SPrOS and SDPlight \*

Область	Позиция	SPrOS	SDPlight
R1	46	+	+
	48	+	-
	50	+	-
	52	+	+
R2	95	+	+
	97	+	+
	98	+	+
	100	+	+
	102	+	+
R3	135	+	+
	136	+	+
	137	+	+
	139	+	+
R4	202	+	+
	204	+	-
	205	+	-
	206	+	+
Результат прогноза		100%	76%

\*Mazin et al., 2010  
Algorithms Mol Biol.; 5:29

# Прогноз специфичных остатков для *Purine Repressor* (PurR) с $p\text{-value} < 10^{-6}$

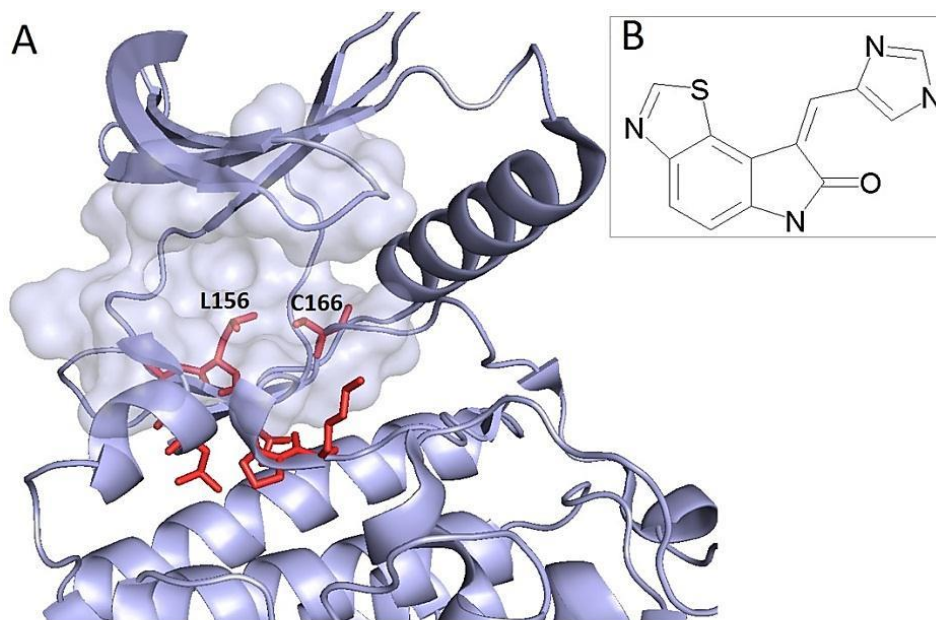


Предсказанные остатки отмечены лиловым.

35 из 47 позиций описаны в литературе  
(Suckow et al. 1996, Mol Biol., 261, 4, 509-523).

Наиболее значимая оценка получена для W147, который участвует в аллостерической регуляции  
(Huffman et al., 2002, Biochemistry, 41, 511-520).

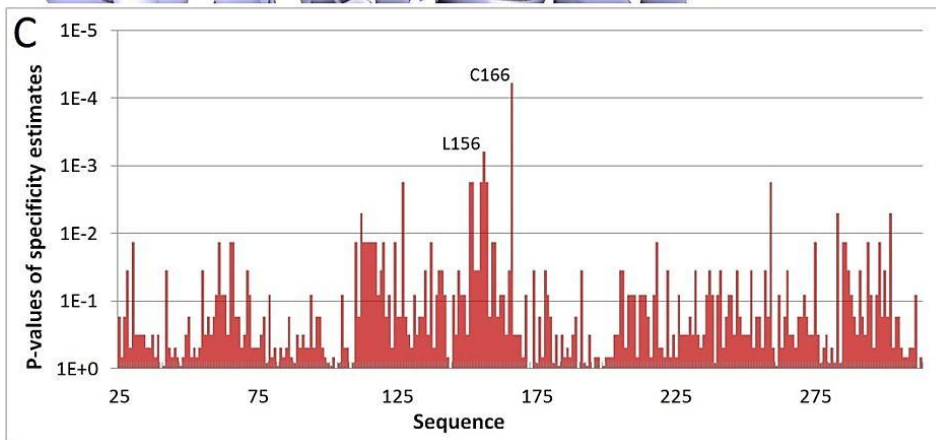
# Прогноз специфичных позиций для ингибиторов протеинкиназы ERK1



А. Трехмерная структура ERK1

В. Структура ингибитора

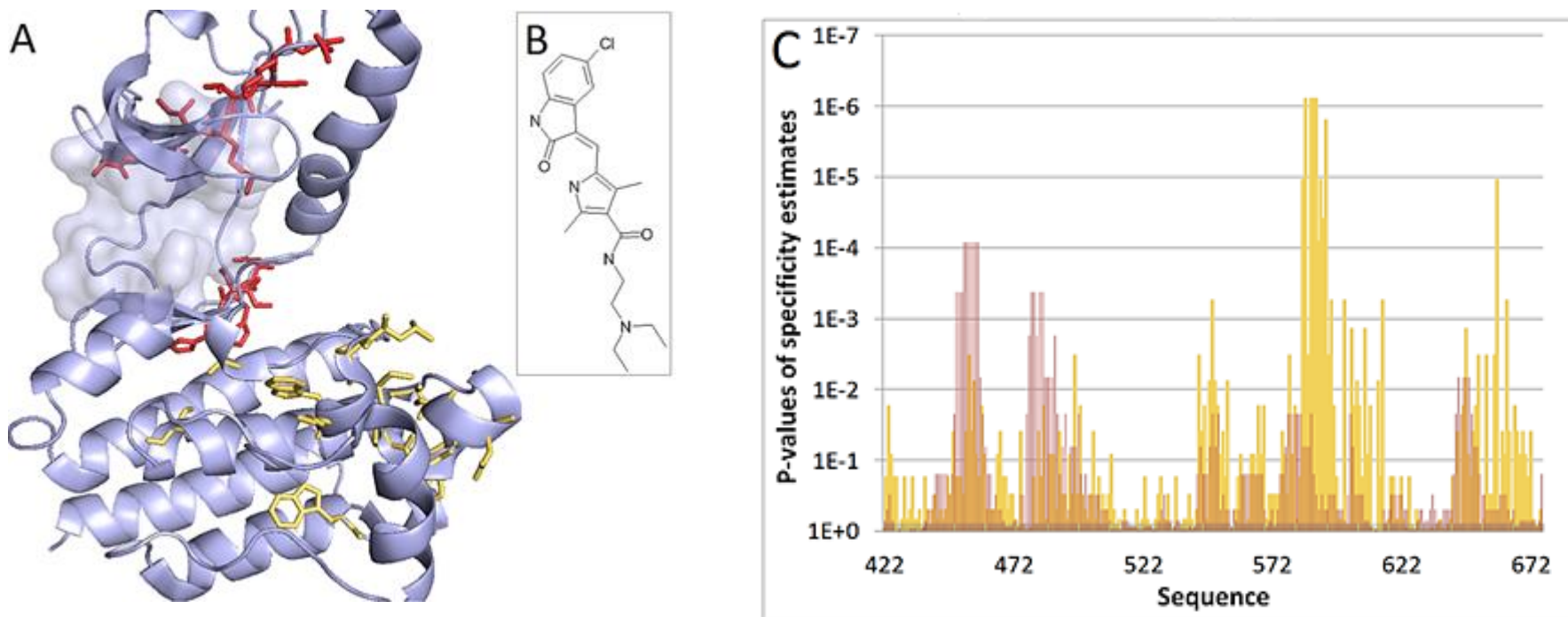
С. Статистическая оценка специфичности аминокислотных позиций



«Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (ERK) using conformational control.» Aronov A.M. et al. (2009)

"Role of a cysteine residue in the active site of ERK and the MAPKK family." Ohori M. et al. (2007) ... etc.

# Прогноз специфичных позиций для ингибиторов протеинкиназы FAK1



A. Трехмерная структура FAK1

B. Структура ингибитора

C. Статистическая оценка специфичности аминокислотных позиций

# Выводы

Метод SPrOS позволяет идентифицировать аминокислотные позиции, связанные с ингибиторной специфичностью протеинкиназ.

Предсказанные позиции могут определять групповую специфичность, участвуя в непосредственном взаимодействии с лигандом, либо опосредованно.

Наличие близких гомологов в обучающих данных может существенно влиять на результат прогноза. Исключение ближайших гомологов позволило точнее локализовать аминокислотные остатки, формирующие лиганд-связывающую область.

## Prediction of amino acid positions specific for protein functional classes

The method SPrOS (Specificity Projection On Sequence) is developed to analyze the amino acid sequences related to the same protein family in order to recognize the amino acid residues associated with separated subclasses within this family.

The algorithm SPrOS requires the training set of preliminary classified amino acid sequences. A user should choose the test sequence(s), in which class-specific positions to be predicted. The test sequence is excluded from the training set and compared with all the rest sequences. The obtained positional scores are used as input to the procedure, which estimates the specificity of each query sequence position to each given class

The method provides the specificity estimates  $E_{ia}$ , evaluating the specificity of the position  $i$  of tested sequence to the class  $A$ . The more the  $E_{ia}$  value the more the specificity evaluation. The **p-values** are used to obtain the statistical significance of the  $E_{ia}$  estimates. The lower the **p-value** the more significant  $E_{ia}$ .

For more details see [SPrOS.pdf](#)

Sequence\*:

Sequence classification\*:

Test Sequence\*:

Cutoff p-value:

Frame:

Mode:

\* indicates required fields

Karasev D.A., Veselovsky A.V., Oparina N.Y., Filimonov D.A., Sobolev B.N.  
Prediction of amino acid positions specific for functional groups in a  
protein family based on local sequence similarity. J Mol Recognit. 2016  
Apr;29(4):159-69.



Работа поддержана грантом РФФИ  
№ 16-04-00491-а

«Роль единичных аминокислотных остатков  
в формировании функциональной  
специфичности белков»

В работе участвовали:

Веселовский А.В.

Опарина Н.Ю.

Рудик А.В.

Филимонов Д.А.

Соболев Б.Н.

Спасибо за внимание!

