

BIOINFORMATIKA

DATABÁZE PROTEINOVÝCH SEKVENCÍ

UniProtKB

- SWISS-PROT: kvalitní ruční anotace
- TrEMBL: automatická anotace (TrEMBL → SWISS-PROT)

DATABÁZE DNA SEKVENCÍ

- EMBL-Bank : Europe (EMBL-EBI),
přístup z ENA (European Nucleotide Archive)
- GenBank: USA, vyhledávač ENTREZ
- DDBJ Japan, vyhledávač ARSA, DBGet

STRUKTURNÍ DATABÁZE

- PDB
- PDBsum: shrnutí a analýzy
- SCOP: fold–superfamily–family
- CATH: class–architecture–topology–homology

PÁROVÉ PŘILOŽENÍ (PAIRWISE ALIGNMENT)

D A G T K V S A E Q I L

| | | | |

D A G T K E C H Q I L

score=5, gap=0

D A G T K V S A E Q I L

| | | | | | | |

D A G T K E C H - Q I L

score=8, gap=1

D A G T K V S A E - - Q I L

| | | | | | | | |

D A G T K - - - E C H Q I L

score=9, gap=5

BLOSUM62

	C	S	T	A	G	P	D	E	Q	N	H	R	K	M	I	L	V	W	Y	F	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
A	0	1	0	4																	A
G	-3	0	-2	0	6																G
P	-3	-1	-1	-1	-2	7															P
D	-3	0	-1	-2	-1	-1	6														D
E	-4	0	-1	-1	-2	-1	2	5													E
Q	-3	0	-1	-1	-2	-1	0	2	5												Q
N	-3	1	0	-2	0	-2	1	0	0	6											N
H	-3	-1	-2	-2	-2	-2	-1	0	0	1	8										H
R	-3	-1	-1	-1	-2	-2	-2	0	1	0	0	5									R
K	-3	0	-1	-1	-2	-1	-1	1	1	0	-1	2	5								K
M	-1	-1	-1	-1	-3	-2	-3	-2	0	-2	-2	-1	-1	5							M
I	-1	-2	-1	-1	-4	-3	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-1	-4	-3	-4	-3	-2	-3	-3	-2	-2	2	2	4					L
V	-1	-2	0	0	-3	-2	-3	-2	-2	-3	-3	-3	-2	1	3	1	4				V
W	-2	-3	-2	-3	-2	-4	-4	-3	-2	-4	-2	-3	-3	-1	-3	-2	-3	11			W
Y	-2	-2	-2	-2	-3	-3	-3	-2	-1	-2	2	-2	-2	-1	-1	-1	-1	2	7		Y
F	-2	-2	-2	-2	-3	-4	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	1	3	6	F
	C	S	T	A	G	P	D	E	Q	N	H	R	K	M	I	L	V	W	Y	F	

PÁROVÉ PROHLEDÁVÁNÍ DATABÁZÍ

”Fast local similarity algorithms”

- FastA
- BLAST

```
MGIKQYSQEELKEMALVEIAHELFEHKKPVPFQELLNEIASLLGVKKEELGDRIAQFYT
DLNIDGRFLALSDQTWGLRSWYPYDQLDEETQPTVKAKKKKAKKAVEEDLDLDEFEEIDE
DDLDEVEEELDLEADDFDEEDLDEDDDDLEIEEDIIDEDDEDYDDEEEEEIK
```

```
ttgggtatca aacaatattc acaggaagag ctaaaggaaa tggctttagt tgaatcgct
cacgaattat ttgaagaaca taaaaacca gttccttttc aggagctttt aatgaaatc
gcatctttgc tcggcgtgaa aaaagaagag cttggagacc gcattgctca atttataca
gatttaaaca ttgacggccg cttcctggcg ctttctgacc agacgtgggg gcttcgcagc
tggatcctt atgatcagct tgatgaagaa actcagccga cagtcaaggc gaaaagaaa
aaagcgaaga aagcagtcga agaagatctt gatcttgacg agtttgaaga gatcgacgaa
gacgaccttg atttggatga agttgaggaa gaactcgatc ttgaagccga cgattttgac
gaagaagatc ttgatgaaga cgacgatgat cttgagatcg aagaagatat tattgatgaa
gatgatgaag actatgatga tgaagaagag gaaattaaat ag
```

VÍCENÁSOBNÉ PŘILOŽENÍ

(MSA = MULTIPLE SEQUENCE ALIGNMENT)

Postupné (progresivní) algoritmy

- CLUSTAL: párové přiložení + tvorba stromů příbuznosti

```
>1SEM1|Chains A, B|SEM-5|Caenorhabditis elegans (6239)
```

```
ETK FVQALFDFNPQESGELAFKRGDVITLINKDDPNWEGQLNNRRGIFPSNYVCPYN
```

```
>1GL51|Chain A|TYROSINE-PROTEIN KINASE TEC|MUS MUSCULUS (10090)
```

```
GSEIVVAMYDFQATEAHDRLRLERGQEYIILEKNDLHWRARDKYGSEGYIPSNYVTGKKSNNLDQY
```

```
>P42682SH3
```

```
DERIQVKALYDFLPREPGNLALKRAEEYLILERCDPHWWKARDRFGNEGLIPSNYVTENRL
```

PŘEDPOVÍDÁNÍ STRUKTURY ZE SEKVENCE

- Sekundární struktura: PSI-PRED
- Fold: threading
- Terciární struktura z homologní struktury: homologní modelování
- Terciární struktura z MSA: AlphaFold.2

testovací sekvence:

PIAQIHILEGRSDEQKETLIREVSEAIRSLDAPLTSVRVIITEMAKGHFGIGGELASK

ALPHAFOLD2

Article

Highly accurate protein structure prediction with AlphaFold

<https://doi.org/10.1038/s41586-021-03819-2>

Received: 11 May 2021

Accepted: 12 July 2021

Published online: 15 July 2021

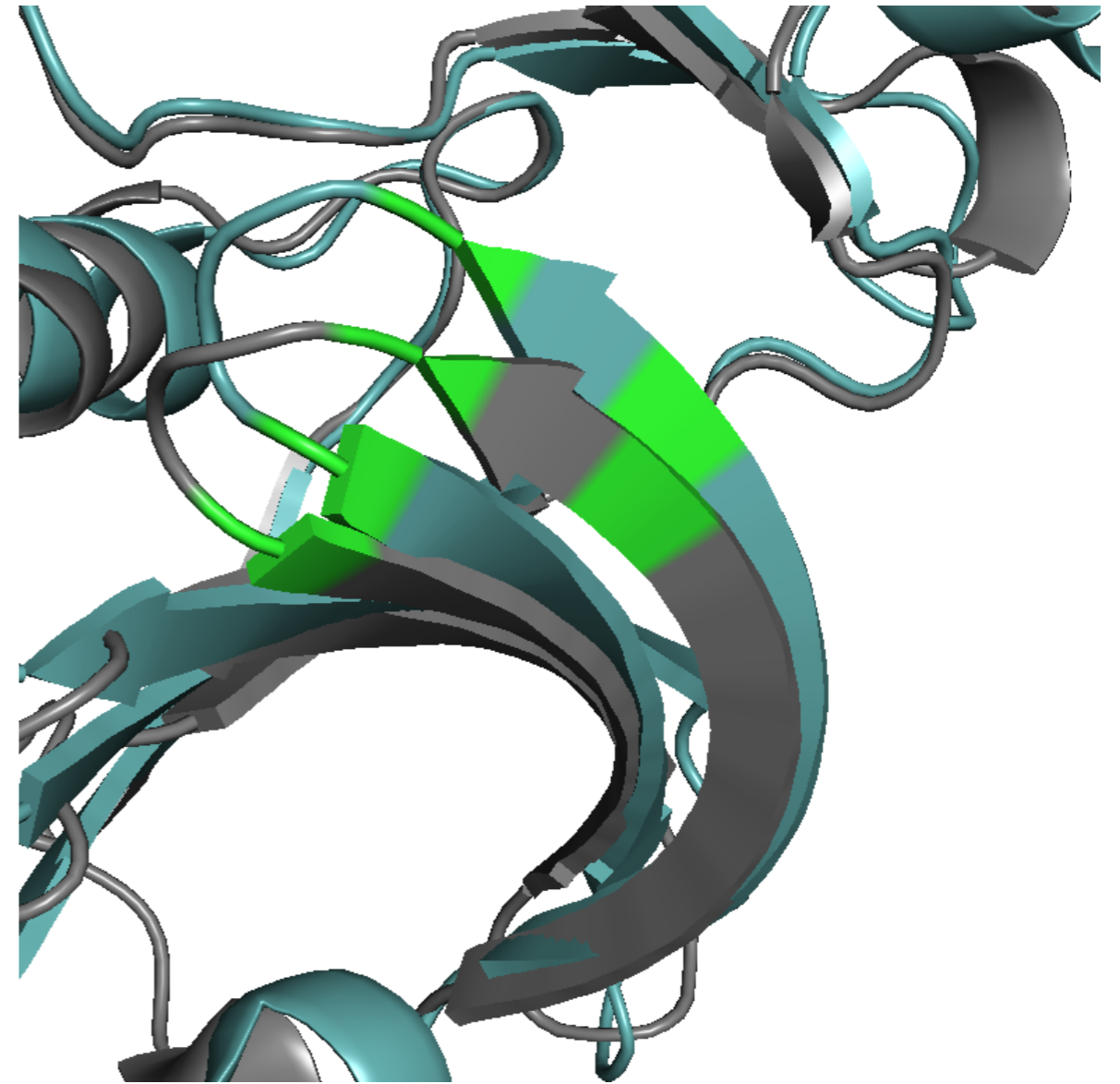
Open access



Check for updates

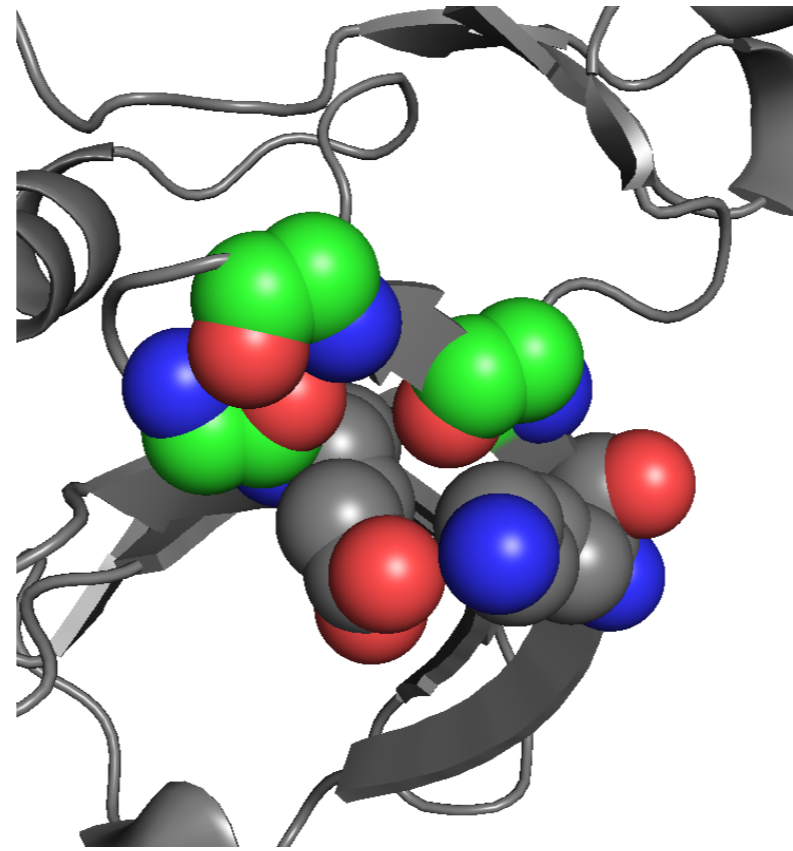
John Jumper^{1,4}✉, Richard Evans^{1,4}, Alexander Pritzel^{1,4}, Tim Green^{1,4}, Michael Figurnov^{1,4}, Olaf Ronneberger^{1,4}, Kathryn Tunyasuvunakool^{1,4}, Russ Bates^{1,4}, Augustin Židek^{1,4}, Anna Potapenko^{1,4}, Alex Bridgland^{1,4}, Clemens Meyer^{1,4}, Simon A. A. Kohl^{1,4}, Andrew J. Ballard^{1,4}, Andrew Cowie^{1,4}, Bernardino Romera-Paredes^{1,4}, Stanislav Nikolov^{1,4}, Rishub Jain^{1,4}, Jonas Adler¹, Trevor Back¹, Stig Petersen¹, David Reiman¹, Ellen Clancy¹, Michal Zielinski¹, Martin Steinegger^{2,3}, Michalina Pacholska¹, Tamas Berghammer¹, Sebastian Bodenstern¹, David Silver¹, Oriol Vinyals¹, Andrew W. Senior¹, Koray Kavukcuoglu¹, Pushmeet Kohli¹ & Demis Hassabis^{1,4}✉

	-4	+4
SRC	V	K
	L	G
	Q	G
	C	F
	G	E
	V	
HCK	K	L
	G	A
	G	Q
	F	G
	E	V
ABL	H	K
	L	G
	G	G
	Q	Y
	G	E
	V	
LCK	E	R
	L	G
	A	G
	Q	F
	G	E
	V	
SLK	G	E
	L	G
	D	A
	F	G
	K	V
SBK	R	E
	L	G
	K	G
	T	Y
	G	K
	V	



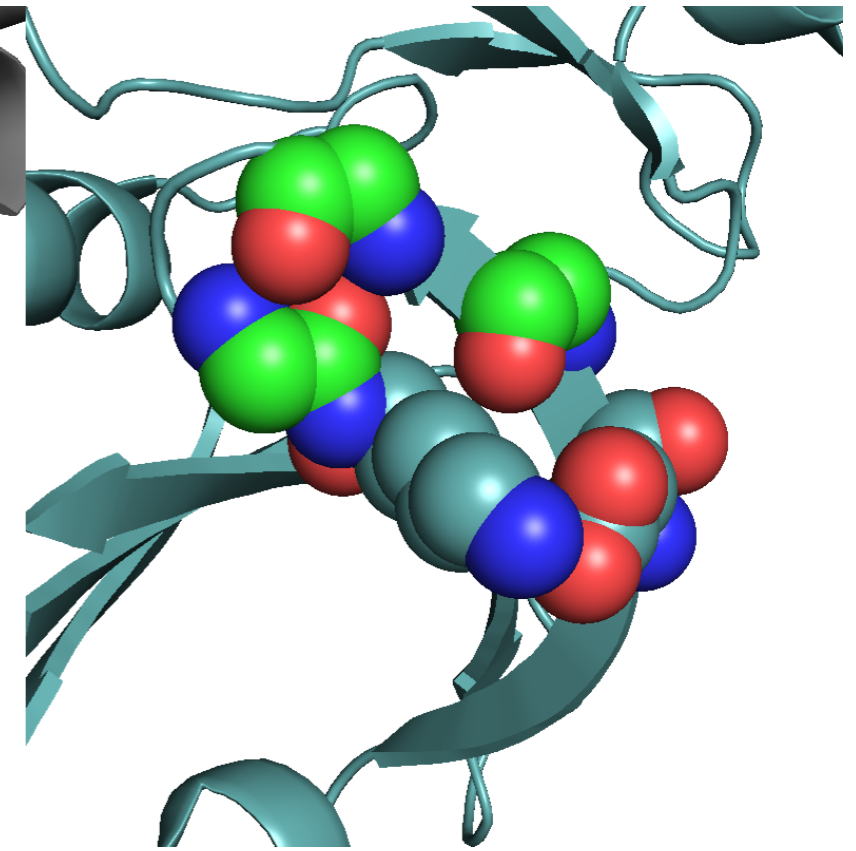
kinasa SRC, kinasa SLK
 GXGXXG - smyčka

	-4	+4									
SRC	V	K	L	G	Q	G	C	F	G	E	V
HCK	K	K	L	G	A	G	Q	F	G	E	V
ABL	H	K	L	G	G	G	Q	Y	G	E	V
LCK	E	R	L	G	A	G	Q	F	G	E	V
SLK	G	E	L	G	D	G	A	F	G	K	V
SBK	R	E	L	G	K	G	T	Y	G	K	V



kinasa SRC

E...K

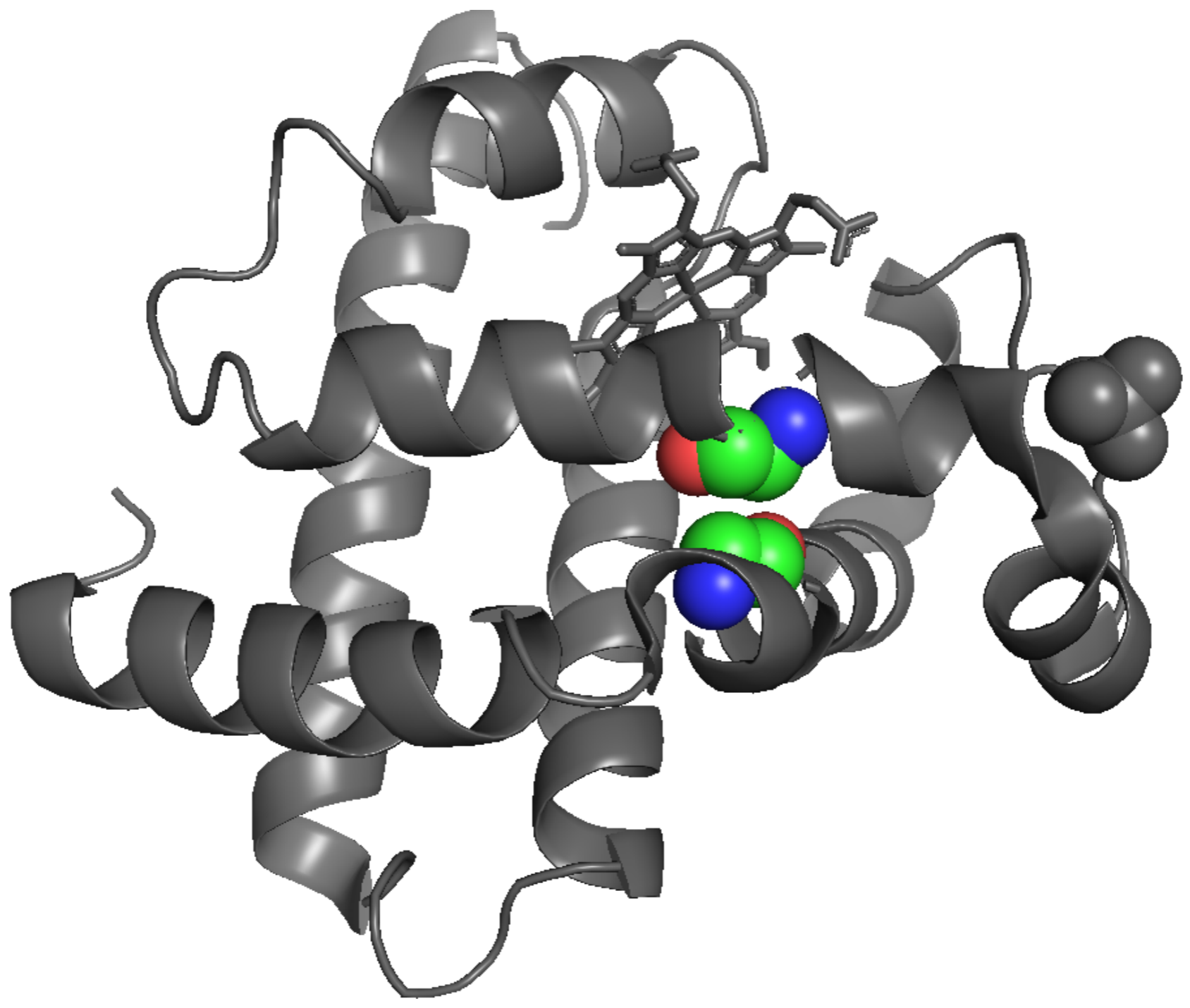


kinasa SLK

K...E

		24	64	116
<i>Homo sapiens</i>	člověk	HGQEV	HGAT	QSKH
<i>Sus scrofa</i>	prase	HGQEV	HGNT	QSKH
<i>Equus caballus</i>	kůň	HGQEV	HGTV	HSKH
<i>Dugong dugon</i>	moroň indický	HGLEV	HGTT	QSKH
<i>Balaena mysticetus</i>	velryba grónská	HGQDV	HGNT	HSRH
<i>Physeter macrocephalus</i>	vorvaň obrovský	HGQDI	HGVT	HSRH

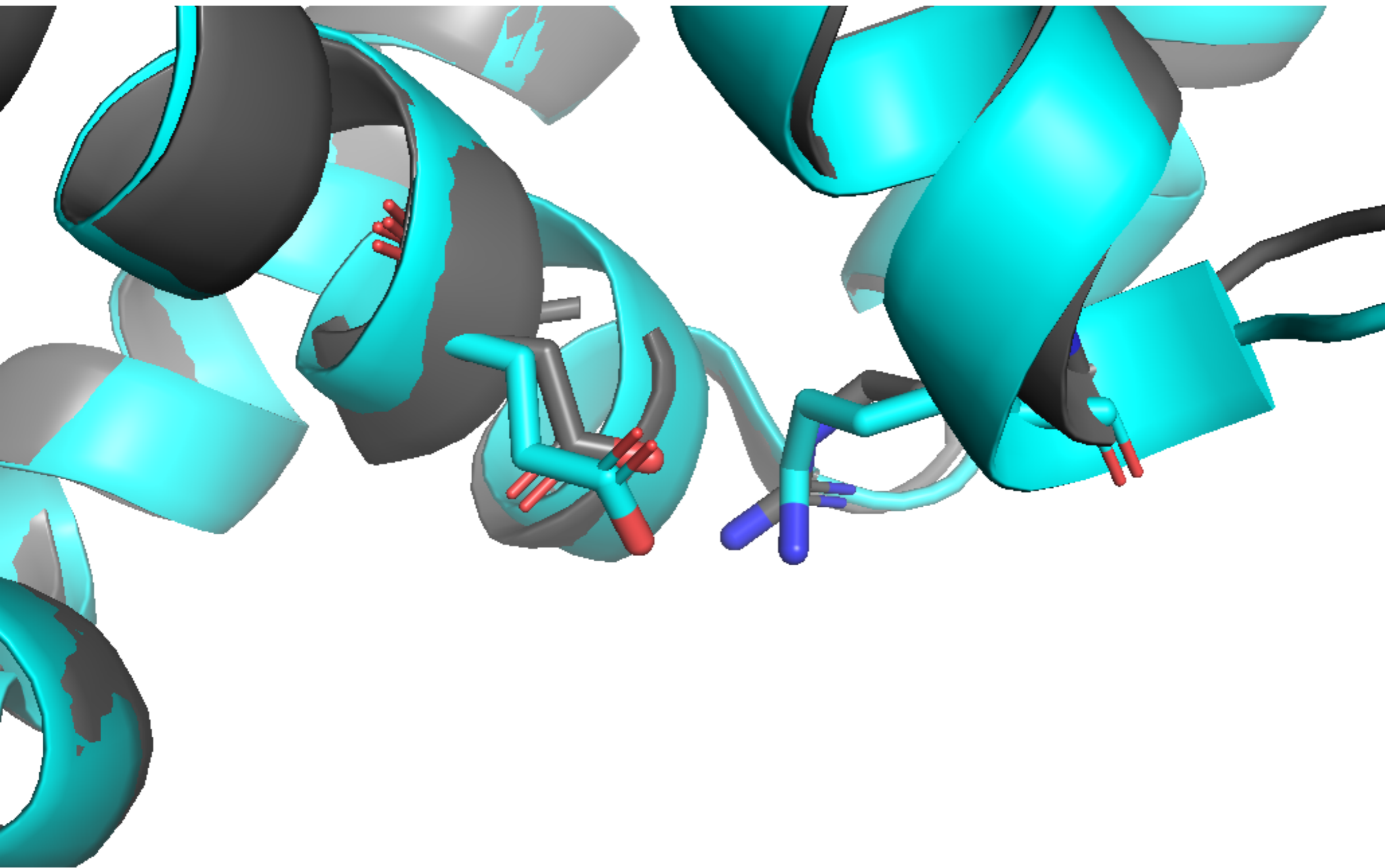
G: malá



		24	64	116
<i>Homo sapiens</i>	člověk	HGQEV	HGAT	QSKH
<i>Sus scrofa</i>	prase	HGQEV	HGNT	QSKH
<i>Equus caballus</i>	kůň	HGQEV	HGTV	HSKH
<i>Dugong dugon</i>	moroň indický	HGLEV	HGTT	QSKH
<i>Balaena mysticetus</i>	velryba grónská	HGQDV	HGNT	HSRH
<i>Physeter macrocephalus</i>	vorvaň obrovský	HGQDI	HGVT	HSRH

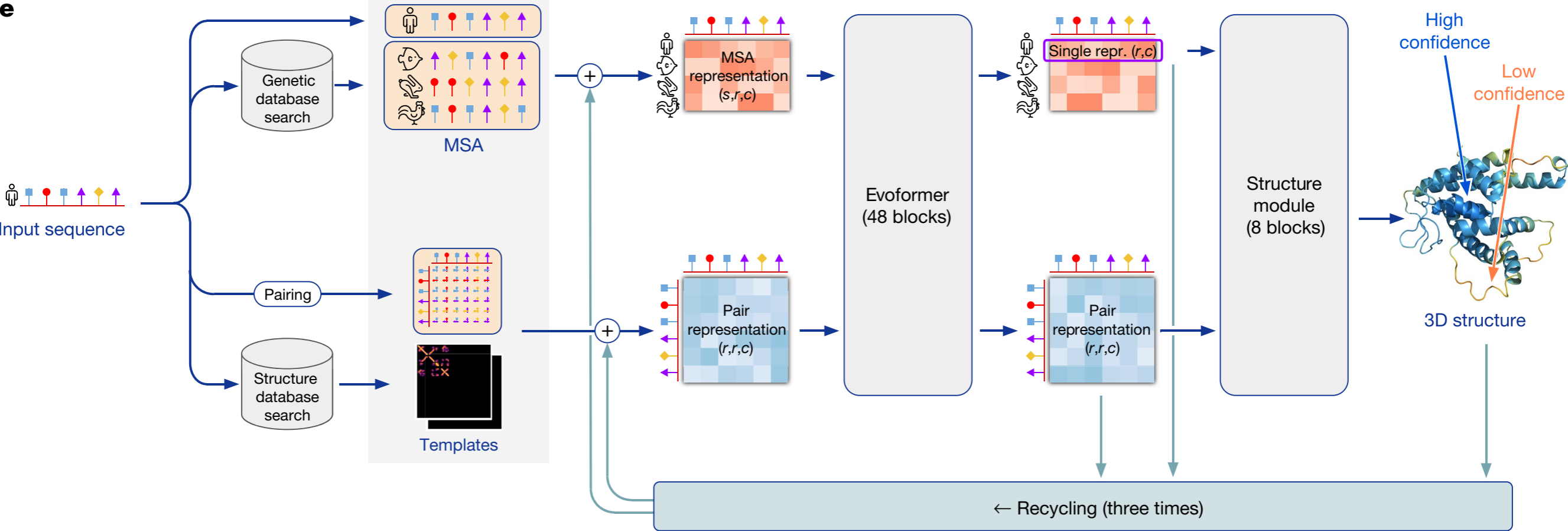
E: dlouhá \ominus ... \oplus K: krátká

D: krátká \ominus ... \oplus R: dlouhá



krátká-D $\ominus \dots \oplus$ R – dlouhá
 dlouhá – E $\ominus \dots \oplus$ K-krátká

e



190 198
SRC K L G Q G C F G E

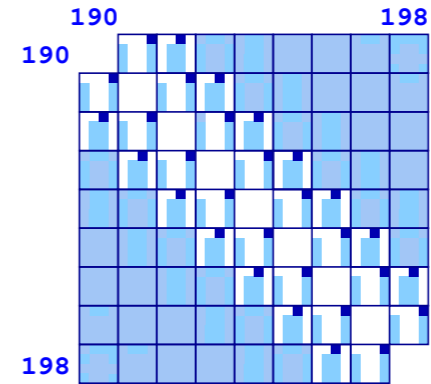
UniProt KB

Embedder

representace Z

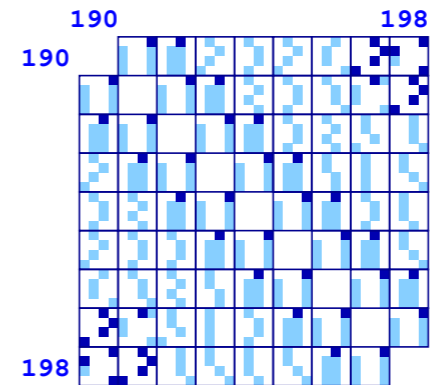
representace M

190 198
SRC
HCK
ABL
LCK
SLK



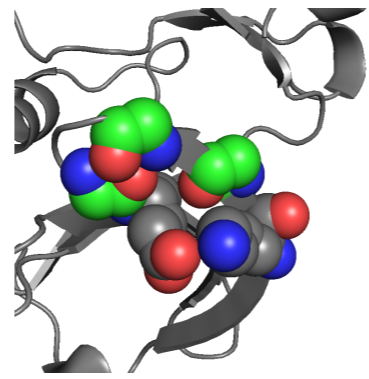
Evoformer

SRC
HCK
ABL
LCK
SLK



SRC
representace S

Structure Module



a

