Molecular Modeling 2020 Lecture 24

SARS-2 lifecycle. ACE2 in mammals. Searching for an animal model. MOE tools for MSAs

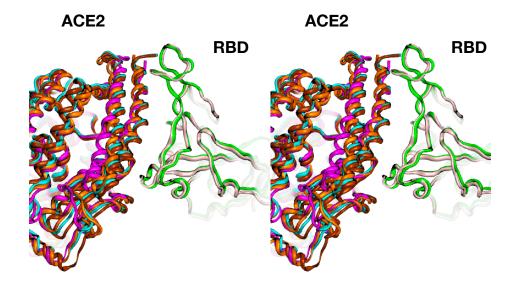
Youtube video on SARS-2

• Find link on course web page

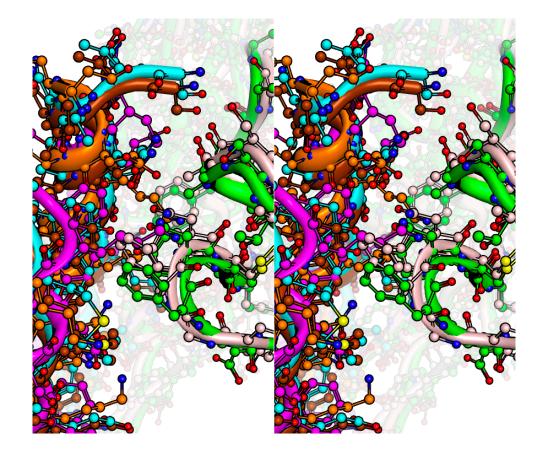
What mammalian species can be infected by SARS-2?

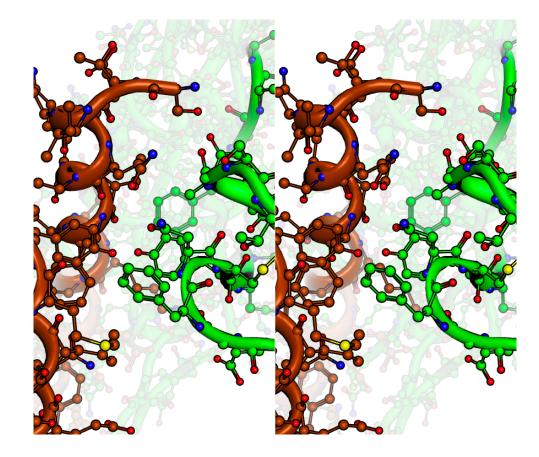
Looking at the ACE2·RBD interface

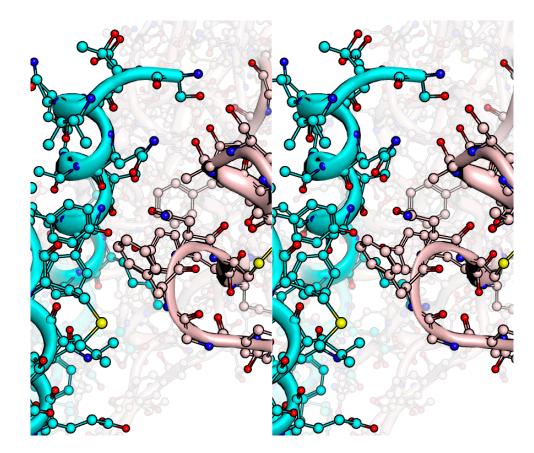
ACE2 = angiotensin converting enzyme 2, the cellular receptor for SARS-2 RBD = receptor binding domain of the SARS-2 spike glycoprotein



Multiple copies, structures of ACE2 and multiple copies of RBD are in basic agreement.

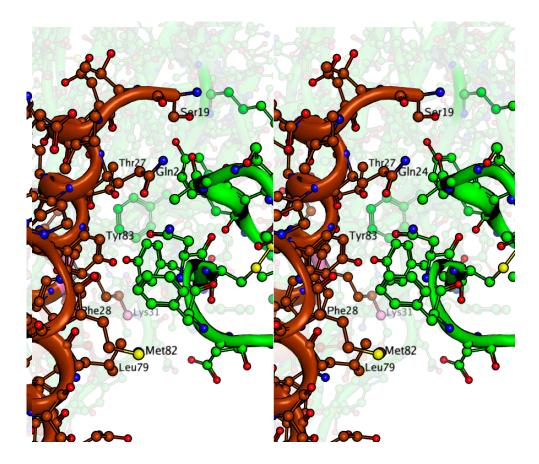




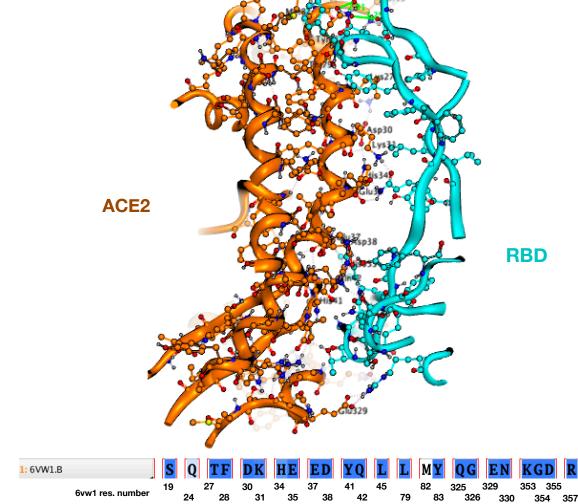


6VW1_BF





ACE2•RBD interface residues



Exercise 24.1 Phyogenetic analysis of a protein-protein interface

We are asking: Which species will favor this ACE2·RBD interface? Which species will not?

1. Search NCBI Proteins for "ACE2 human". See figure. Select Orthologs. Search mammals. Choose a subgroup of placentals, such as rodents or carnivores. Select individual species or all. Download RefSeq Proteins (fasta). Open in MOE.

2. Open 6vw1 (ACE2•RBD). Align all sequences, except RBD.

3. Select all interface residues on ACE2 by picking. SiteView it.

5.	With interface residues still selected, SEC	Ruler bar	right-mouse	hide	unselected.
(nc	w you see all interface residues only, all specie	es)			

- 3. SEQ: Select all chains of ACE2. Select all interface residues. SEQ: Alignment | similarity, tree
- 4. Unselect (don't delete) redundant species. Keep interesting ones. (The tree updates)

5. Select each residue in SEQ. Go to MOE to inspect it. As you do, ask: <u>Can the interface tolerate the mutations?</u> For which species?

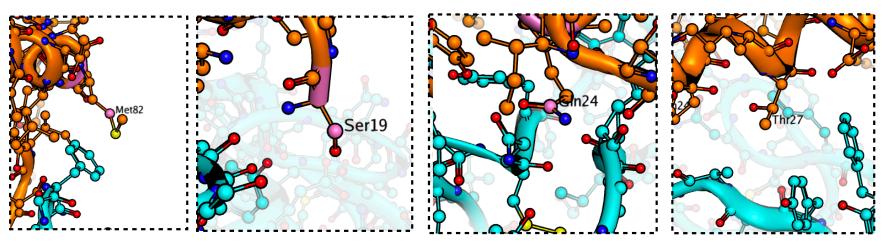
GENE	Was this helpful?
ACE2 - angiotensin I converting enzyme 2	
Homo sapiens (human)	
Processed peptides: angiotensin-converting enzyme 2	
Also known as: ACEH	
GeneID: 59272	
RefSeq transcripts (5) RefSeq proteins (5) RefSeqGe	ne (1) PubMed (271)
Orthologs Genome Browser BLAST	Download
RefSeq Sequences	+

Interface residues of representative mammals

																Interface
	6vw1 res. number	19	24	27 28	30 31	34 35	37 38	41 42	45	79	82 83	325 326	329 330	353 35) 354	5 357	tolerates
	6VW1.B	S	Q	ΤF	DK	ΗE	ED	YQ	L	\mathbf{L}	МY	QG	EN	KGD	R	changes?
	Bactrian camel	S	\mathbf{L}	T F	ΕE	ΗE	ΕD	ΥQ	L	Т	ΤY	QG	D N	KGD	R	yes
	Przewalski's horse	S	\mathbf{L}	T F	ΕK	S E	ΕE	ΗQ	L	\mathbf{L}	ΤY	QG	ΕN	K G D	R	yes
	Blue whale	S	Q	T F	Q <mark>K</mark>	ΗE	ΕD	<mark>Y</mark> R	\mathbf{L}	Ι	ΤY	Q <mark>E</mark>	<mark>V</mark> N	K G D	R	no, 2 hits against it
	Leopard	S	L E	T F	ΕK	ΗE	ΕE	YQ	L	\mathbf{L}	ТΥ	QG	ΕN	K G D	R	yes
	-Sunda pangolin	S	Ε	T F	ΕK	S E	ΕE	ΥQ	L	Ι	ΝY	Q T	ΕN	K <mark>H</mark> D	R	no, 3 hits against it
	-Coquerel's sifaka	S	Q	T F	D K	ΗE	E D	ΥQ	L	\mathbf{L}	ΤY	QG	ΕN	KGD	R	yes
	-human	S	Q	T F	D K	ΗE	E D	ΥQ	L	\mathbf{L}	М <mark>Ү</mark>	QG	ΕN	K G D	R	yes, known host
[American pika	S	\mathbf{L}	T F	D K	QE	E D	ΥQ	L	\mathbf{L}	ΤY	QG	ΕN	K D D	R	no
	European rabbit	S	\mathbf{L}	T F	ΕK	QE	E D	ΥQ	L	\mathbf{L}	ΤY	QG	ΕN	K G D	R	yes
	long-tailed chinchilla	S	Q	T F	<mark>D</mark> N	ΕK	E D	ΥQ	L	\mathbf{L}	ΑY	QG	QN	K D D	R	no
	-thirteen-lined ground squirrel	S	\mathbf{L}	T F	D K	QE	E D	ΥQ	L	\mathbf{L}	ΑY	QG	ΕN	KGD	R	yes
	prairie vole	S	D	ΑF	D K	QE	E D	ΥQ	L	\mathbf{L}	S <mark>Y</mark>	QG	E N	K D D	R	no, 2 hits against it
	house mouse	S	N M	T F	N N	QE	E D	YQ	L	T V	S F	QG	AN	HG D	R	no, probably not.
	lesser jerboa	S		$\mathbf{T} \mathbf{F}$	D K	QE	E D	YQ	L	V	ΤY	QG	EN	K N D	R	no
	-black rat	S	К	S F	N <mark>K</mark>	QE	E <u>D</u>	YQ	L	Ι	N F	P <mark>G</mark>	T N	H G D	R	yes, maybe
L	horseshoe bat	S	\mathbf{L}	ΚF	<mark>D</mark> D	S E	ΕN	ΗQ	L	\mathbf{L}	NF	EG	Ν <mark>Ν</mark>	K G D	R	yes, known host
s=surface (+/-)		S	-	pn	S	s	bç		pp		S	S	S-	bc b	-	
p=pocket (charged/pola	ar/nonpolar)		pn	pr	n S	S	bo	; pc		pn	bp arom	bn sma		pn small!	S	

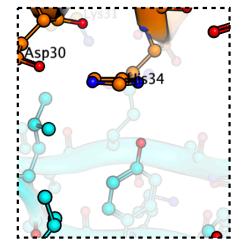
b=buried (charged/polar/nonpolar)

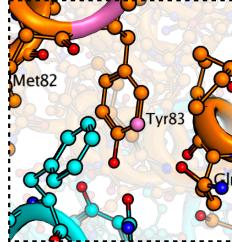
Interface residues on ACE2



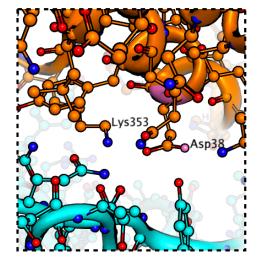
M82 is [MTAS] in mammals, is also not doing much in the structure. So anything goes, but M<A<T,S **S19** is highly exposed. Anything goes here. **Q24** is partially buried. [LDMQ] in mammals. Probably wouldn't like D (*prairie vole*) because of the absence of a cation nearby. **T27** is buried. Can't be bigger than a T. Only [TA] appear. Ala is fine here.

Interface residues on ACE2





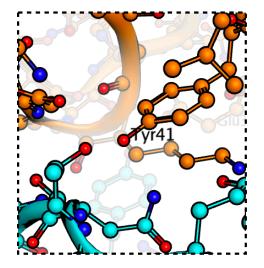
Glu37

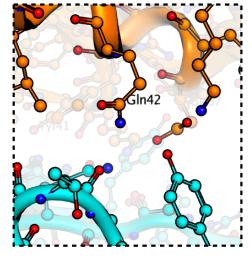


H34 is [HQSY] in mammals. All of those sidechains fit in the wide pocket. Y83 makes key interactions with RBD, but is strictly conserved in mammals. Does not help us.

E37 makes 1 Hbond with RBD, and is strictly conserved in mammals. Does not help us. D38 teams up with K353 to make a H-bond network with RBD. Mutations here could matter. E in *horse, cat* is too long, chooses a different rotamer. Binding lost?
K353 is N in *tarsier*. H-bonds lost.

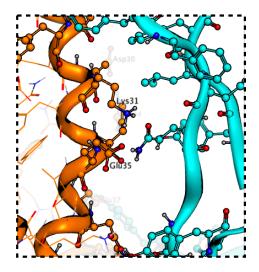
Interface residues on ACE2





Y41 is [H] in *marmoset, horse, tarsier.* Very tight, buried. **H** would fit.

Q42 is [E] in marmoset. Probably unfavorable due to local negative charges.



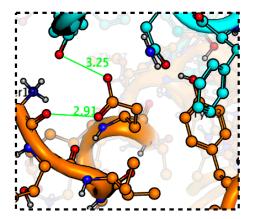
K31 and E35 combine to form a H-bond network. Both are strictly conserved in near-nonprimate mammals.

Interface residues of representative mammals

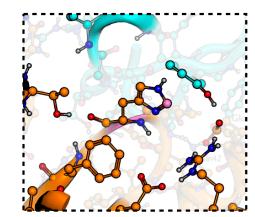
	6vw1 res. number	19	24	27 28	30 31	34 35	37 38	41 42	45	79	82 83	325 326	329 330	353 35 354	5 357	Interface tolerates
	6vw1 res. number 6VW1.B Bactrian camel Przewalski's horse Blue whale Blue whale Copard Coquerel's sifaka Coquerel's sifaka American pika Coquerel's sifaka Coquerel's sifaka C	19 S S S S S S S S S S S S S S S S S S S	24 Q L Q L Q Q L L Q L D N M						45 L L L L L L L L L L L L L	79 L T L L L L L L L V			E N E N E N E N E N E N E N E N E N E N		357 R R R R R R R R R R R R R R R	tolerates changes? yes yes no, 2 hits against it yes no, 3 hits against it yes yes, known host no yes no yes no, 2 hits against it no, probably not. no
s=surface (+/-) p=pocket (charged/pola	black rat	S S S	K L pn	S F K F pn pr	N <mark>K</mark> D D S	QE SE S	ED EN bc bc		L L pp	I L pn	N F N F S bp	PG EG s	TN NN S- S	HGD KGD bc bc pn small!	R R	yes, maybe yes, known host

b=buried (charged/polar/nonpolar)

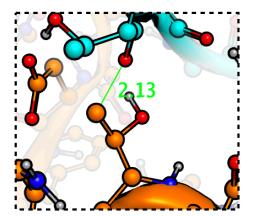
Pangolin, 3 hits against.



Q24E has too many Hbond acceptors, few Hbond donors.



G354H is too close to Arg.



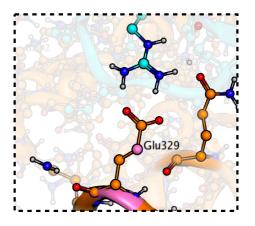
G326T creates unresolvable collisions in a tight place.

Interface residues of representative mammals

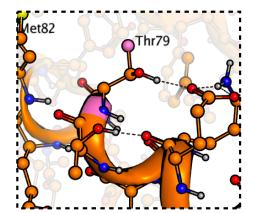
	6vw1 res. number	19	24	27 28	30 31	34 35	37 38	41 42	45	79	82 83	325 326	329 5 330	353 35) 354	5 357	Interface tolerates
	6VW1.B	S	Q	ΤF	DK	ΗE	ED	YQ	L	\mathbf{L}	МY	QG	ΕN	KGD	R	changes?
	Bactrian camel	S	\mathbf{L}	T F	ΕE	ΗE	ΕD	ΥQ	L	Т	ΤY	QG	D N	KGD	R	yes
	Przewalski's horse	S	\mathbf{L}	T F	ΕK	S E	ΕE	Η <mark>Q</mark>	L	\mathbf{L}	ΤY	QG	ΕN	KGD	R	yes
	Blue whale	S	Q	ΤF	QK	ΗE	ED	<mark>Y</mark> R	L	Ι	ΤY	QE	<mark>V</mark> N	KGD	R	no, 2 hits against it
	Leopard	S	\mathbf{L}	T F	ΕK	ΗE	ΕE	YQ	L	\mathbf{L}	ΤY	QG	ΕN	KGD	R	yes
	Sunda pangolin	S	Е	T F	ΕK	S E	ΕE	ΥQ	L	Ι	ΝY	QΤ	ΕN	K <mark>H</mark> D	R	no, 3 hits against it
	Coquerel's sifaka	S	Q	T F	D K	ΗE	ΕD	ΥQ	\mathbf{L}	\mathbf{L}	ТҮ	QG	ΕN	KGD	R	yes
	-human	S	Q	T F	D K	ΗE	E D	ΥQ	\mathbf{L}	\mathbf{L}	МY	QG	ΕN	KGD	R	yes, known host
	American pika	S	\mathbf{L}	T F	D K	QE	E D	YQ	\mathbf{L}	\mathbf{L}	ΤY	QG	ΕN	K D D	R	no
	European rabbit	S	\mathbf{L}	T F	ΕK	QE	ΕD	ΥQ	\mathbf{L}	\mathbf{L}	ΤY	QG	ΕN	KGD	R	yes
	long-tailed chinchilla	S	Q	T F	<mark>D</mark> N	ΕK	ΕD	ΥQ	\mathbf{L}	\mathbf{L}	ΑY	QG	QN	K D D	R	no
	thirteen-lined ground squirrel	S	\mathbf{L}	T F	DK	QE	ΕD	ΥQ	L	\mathbf{L}	Α <mark>Υ</mark>	QG	ΕN	KGD	R	yes
	prairie vole	S	D	ΑF	D K	QE	ΕD	ΥQ	L	\mathbf{L}	SY	QG	ΕN	K D D	R	no, 2 hits against it
	house mouse	S	Ν	$\mathbf{T} \mathbf{F}$	ΝN	QE	ΕD	ΥQ	\mathbf{L}	Т	S F	QG	Α <mark>Ν</mark>	H G D	R	no, probably not.
	lesser jerboa	S	М	T F	D K	QE	E D	ΥQ	\mathbf{L}	V	ТҮ	QG	ΕN	K N D	R	no
	black rat	S	К	S F	N <mark>K</mark>	QE	ΕD	ΥQ	\mathbf{L}	Ι	N F	P G	ΤN	H G D	R	yes, maybe
	horseshoe bat	S	\mathbf{L}	ΚF	<mark>D</mark> D	S E	E N	Η <mark>Q</mark>	\mathbf{L}	\mathbf{L}	N F	Ε <mark>G</mark>	N <mark>N</mark>	KGD	R	yes, known host
s=surface (+/-) p=pocket (charged/pola	ar/nonpolar)	S	pn	pn pr	S I S	s s	bç bç		pp	pn	s bp _{arom}	S br sma		bc bo pn small!	s	

b=buried (charged/polar/nonpolar)

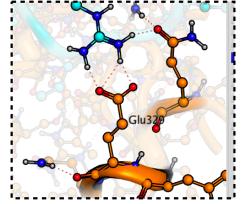
Mouse, probably not a host.



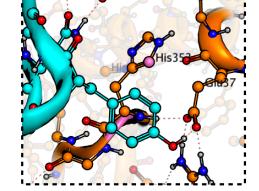
E329A loses a surface salt bridge.



L79T loses hydrophobic interaction, but packs well.



E329A loses a salt bridge.



K353H packs well, but Hbonds are not satisfied.

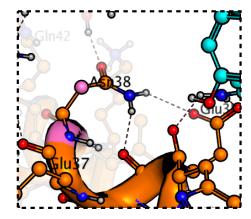
Interface residues of representative mammals

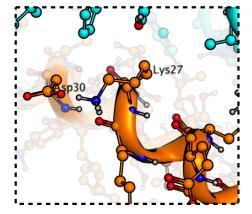
6vw1 res.	number	19	24	27 28	30 31	34 35	37 38	41 42	45	79	82 83	325 326	329 5 330	353 35) 354	5 357	Good fit?
6VW1.B		S	Q	ΤF	DK	ΗE	ED	YQ	L	\mathbf{L}	МY	QG	ΕN	KGD	R	
Bactrian camel		S	\mathbf{L}	T F	ΕE	ΗE	E D	YQ	\mathbf{L}	т	ΤY	QG	D <mark>N</mark>	KGD	R	yes
Przewalski's horse		S	\mathbf{L}	ΤF	E K	S E	ΕE	H <mark>Q</mark>	\mathbf{L}	\mathbf{L}	ΤY	QG	ΕN	K G D	R	yes
Blue whale		S	Q	ΤF	Q <mark>K</mark>	ΗE	E D	YR	\mathbf{L}	Ι	ΤY	QE	<mark>V</mark> N	K G D	R	no, 2 hits against it
Leopard		S	\mathbf{L}	TF	ΕK	ΗE	E E	ΥQ	\mathbf{L}	\mathbf{L}	ΤY	QG	ΕN	K G D	R	yes
Sunda pangolin		S	Е	TF	ΕK	S E	E E	ΥQ	\mathbf{L}	Ι	ΝY	QT	ΕN	K <mark>H</mark> D	R	no, 3 hits against it
Coquerel's sifaka		S	Q	ΤF	D K	ΗE	E D	ΥQ	\mathbf{L}	\mathbf{L}	ΤY	QG	ΕN	K G D	R	yes
human		S	Q	ΤF	D K	ΗE	E D	ΥQ	\mathbf{L}	\mathbf{L}	М <mark>Ү</mark>	QG	ΕN	K G D	R	yes
American pika		S	\mathbf{L}	TF	D K	QE	E D	ΥQ	\mathbf{L}	\mathbf{L}	ΤY	QG	ΕN	K D D	R	no
European rabbit		S	\mathbf{L}	TF	ΕK	QE	E D	ΥQ	\mathbf{L}	\mathbf{L}	ΤY	QG	ΕN	KGD	R	yes
long-tailed chinchilla		S	Q	T F	<mark>D</mark> N	ΕK	ΕD	ΥQ	\mathbf{L}	\mathbf{L}	ΑY	QG	Q <mark>N</mark>	K D D	R	no
thirteen-lined ground	squirrel	S	\mathbf{L}	$\mathbf{T} \mathbf{F}$	D K	QE	ΕD	ΥQ	\mathbf{L}	\mathbf{L}	ΑY	QG	ΕN	KGD	R	yes
prairie vole		S	D	ΑF	D K	QE	ΕD	ΥQ	\mathbf{L}	\mathbf{L}	S <mark>Y</mark>	QG	ΕN	K D D	R	no, 2 hits against it
house mouse		S	Ν	T F	ΝN	QE	ΕD	YQ	\mathbf{L}	т	S F	QG	A <mark>N</mark>	H G D	R	
lesser jerboa		S	М	T F	D K	QE	ΕD	YQ	\mathbf{L}	V	ΤY	QG	ΕN	KND	R	no
black rat		S	К	SF	N <mark>K</mark>	QE	E D	YQ	\mathbf{L}	Ι	ΝF	P G	Τ <mark>Ν</mark>	H G D	R	no, 1 hit
horseshoe bat		S	\mathbf{L}	ΚF	<mark>D</mark> D	S E	ΕN	ΗQ	\mathbf{L}	\mathbf{L}	N F	E <mark>G</mark>	Ν <mark>Ν</mark>	KGD	R	yes
s=surface (+/-)		S	pn	pn pr	S S	S S	bc bc		pp	pn	-		S- S	bc bo pn	s	
p=pocket (charged/polar/nonpolar)											arom	sma	ıll	small!		

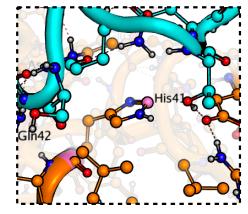
b=buried (charged/polar/nonpolar)

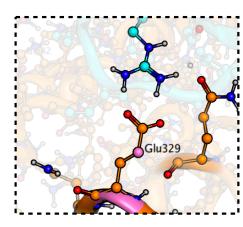
no

Horseshoe bat, true host for SARS-2









D38N makes better Hbonds with a better rotamer.

T27K actually adds a salt bridge.

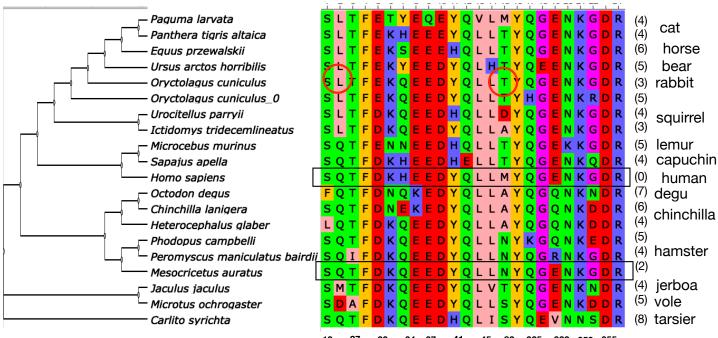
Y41H accomodates the smaller side chain and leaves room for waters.

E329N loses a surface salt bridge.

Bat ACE2 has lots of mutations relative to human, but many of these are acceptable or favorable to RBD binding.

Other mammals

animal model search



19 27 30 34 37 41 45 82 325 329 353 355 24 28 31 35 38 42 79 83 326 330 354 357

Maybe golden hamster is a good model.

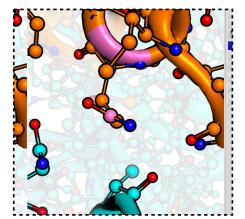


UGENE alignment

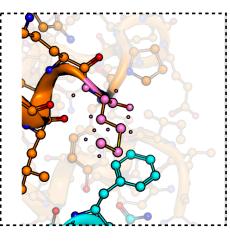
Prediction: Civit ACE2 may bind RBD

homo sapiens SQTFDKHEEDYQLLMYQGENKGDR paguma larvata SLTFETYEQEYQVLMYQGENKGDR

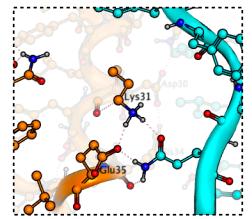




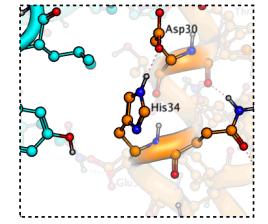
Q24L is in a hydrophobic pocket.



M82T is on the surface.



K31T is in a deep pocket. Smaller side chain leaves space for waters.



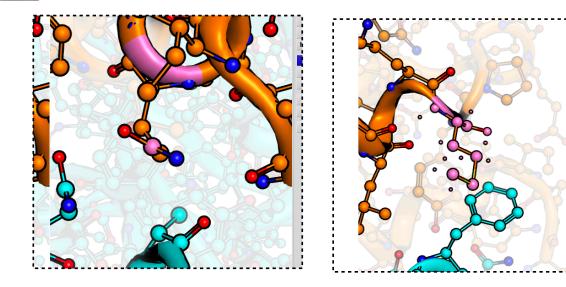
H34Y breaks one Hbond.

Some loss of hydrophobic effect. But no strong negatives. Probably binds RBD.

Prediction: Tiger ACE2 will bind RBD







Q24L fits a hydrophobic pocket well.

M82T is on the surface.

Only 2 meaningful mutations in the interface region!

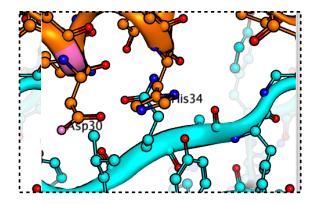
Both of the mutations are energetically acceptable.

A tiger has tested positive.

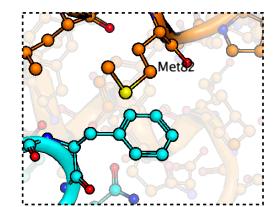
Prediction: <u>Hamster</u> ACE2 will bind RBD



homo sapiens SQTFDKHEEDYQLLMYQGENKGDR miscricetus auratus SQTFDKQEEDYQLLNYQGENKGDR



H34Q fits in a polar pocket.



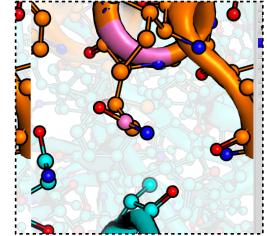
M82N is exposed to solvent and inconsequential.

Both of the mutations are energetically acceptable

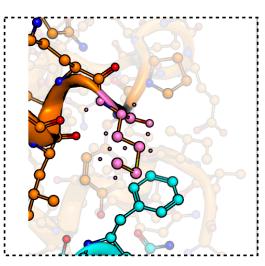
Prediction: Squirrel ACE2 will bind RBD



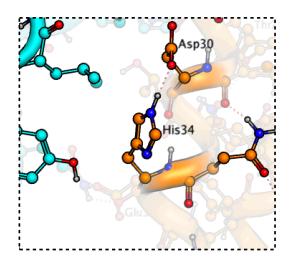
homo sapiens SQTFDKHEEDYQLLMYQGENKGDR ictidomys tridecemliniatus SLTFDKQEEDYQLLAYQGENKGDR



Q24L is in a hydrophobic pocket



M82A is on the surface



H34Q is acceptable.

All three mutations are energetically acceptable when modeled on 6VW1

What coronaviral species can infect humans?

If you want to know, repeat the process from Exercise 24.1 Then isolate the RBD residues at the interface. Which viruses have mutations that the interface will tolerate?

Docking and Design 1

- Dock using ribbons. Make sure 3-fold symmetry is maintained.
- Refine docked pose with backbone atoms only. Make sure CA-CA distances are at least 4Å, 3.5Å if one of them is a glycine.
- Identify interface residues on one monomer of the ligand by picking. Optionally hide other columns or color those residues in SEQ window.
- Extend selection to near residues twice. Unfix. Select | invert. Fix. Now backbone and sidechains are unfixed only in the interface region.
- Turn on gizmin. For each interface residue, mutate (Protein | Protein builder). Try to remove clashes, maintain good geometry, fill space loosely, satisfy charges and H-bonds.
- Add waters where space and H-bond partners allow.

Docking and Design 2

- When you can't find any more mutations, stop mutating. Turn off gizmin.
- Group the designed monomer and all of its waters under one tag.
- Copy the tag twice. Name the new tags "copy 2" and "copy 3". Align sequences.
- Superpose "copy 2" on the second monomer in the template trimer (Unfix "copy 2". Fix template second monomer. Set all other chains to "i", ignore.Superpose, moving all atoms in tag together.). Superpose "copy 3" on the third monomer. Delete the original 2nd and 3rd monomers. Now you have made the compete trimer of the designed monomer.
- Go through the structure and energy minimize locally, using EPUSEIPF. When you have finshed, select the whole trimer with its waters and extend the selection to near residues (the RBD). Unfix. Fix everything else. Perform a 100 ps molecular dynamics simulation at low temperature (t=50). Browse the results. Change to higher or lower T and repeat if necessary. Check geometry again. Use a surface to look for any new locations for waters. Finally, energy minimize.