cresset®
software

innovative science • intuitive software

# Design in 2D, model in 3D: live 3D pose generation from 2D sketches

Paolo Tosco

# Outline

> Introduction to the project

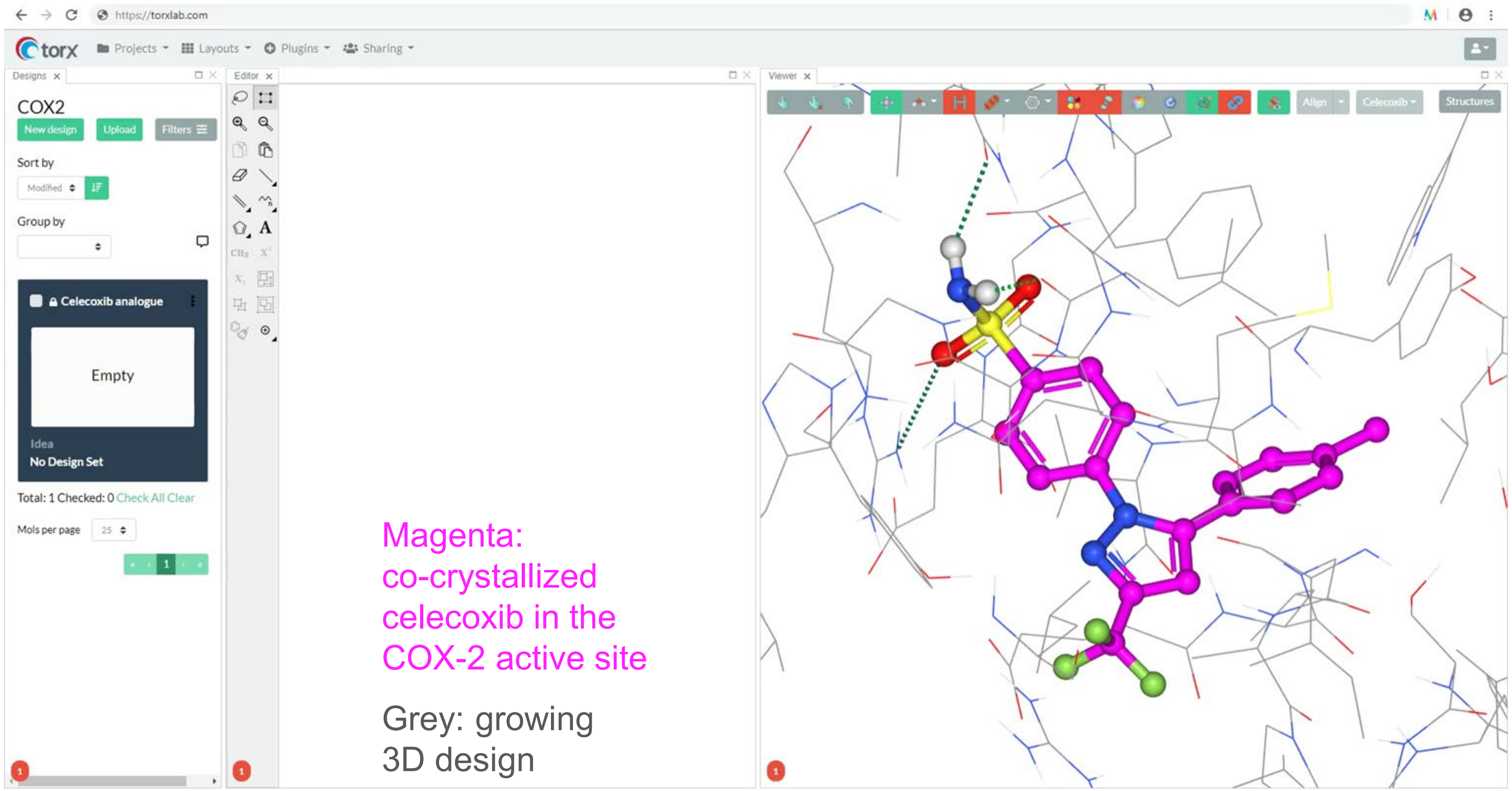
> Description of the *grow3D* algorithm

> The devil is in the details

> Conclusions and outlook

# How nice would it be…

…to design a molecule in the 2D sketcher and see it grow sensibly within the binding site in the 3D viewport?

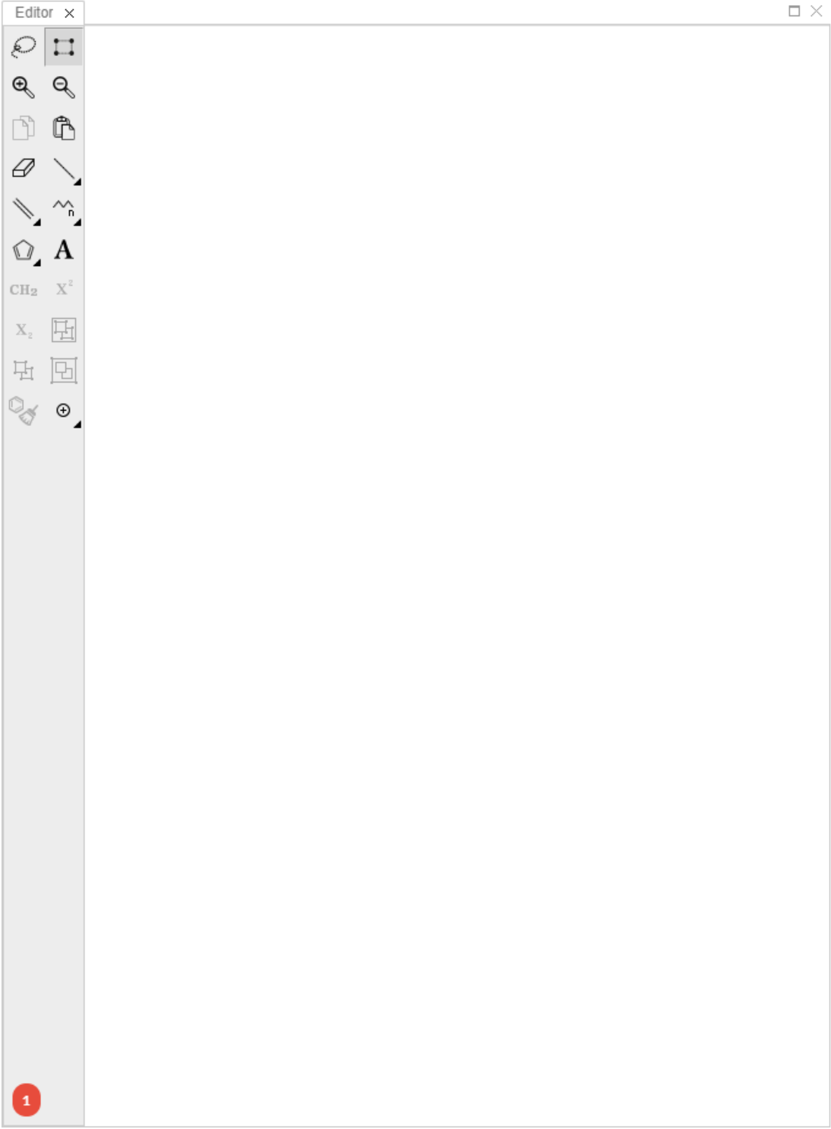


Magenta: co-crystallized celecoxib in the COX-2 active site

Grey: growing 3D design

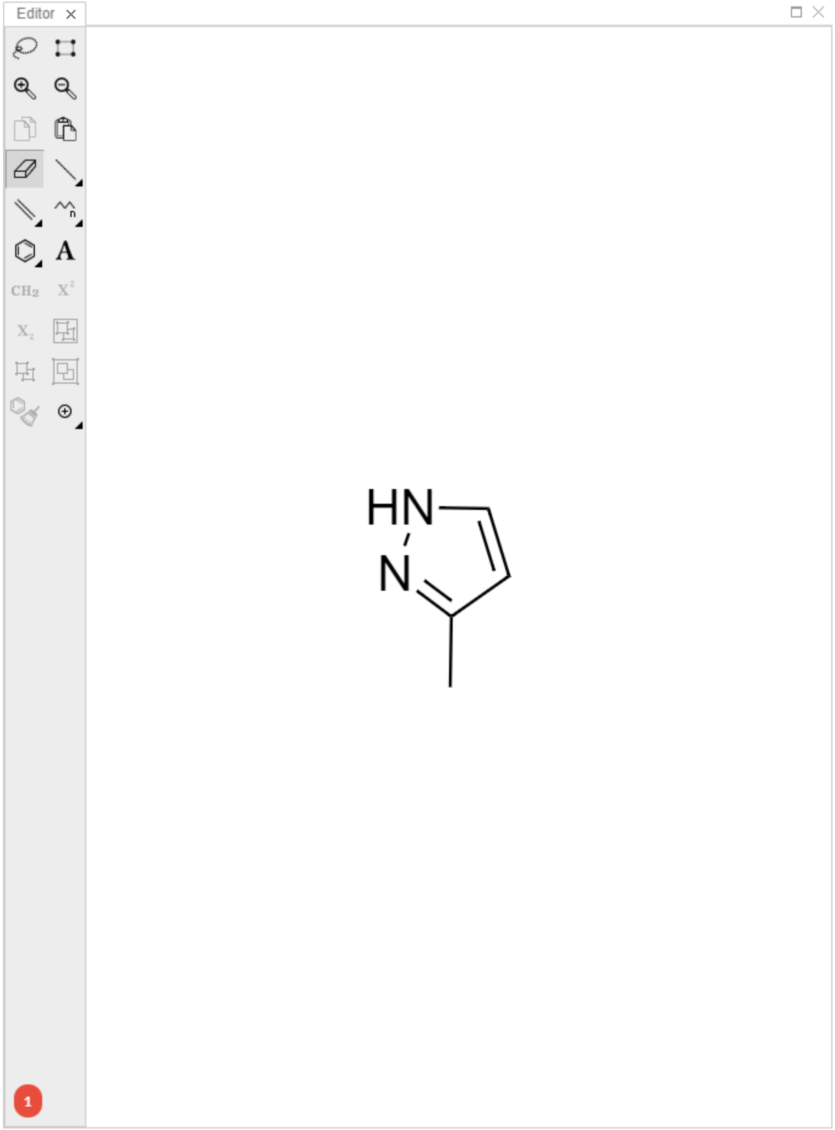
# How hard can it be?

- Start from a blank 2D sketcher canvas



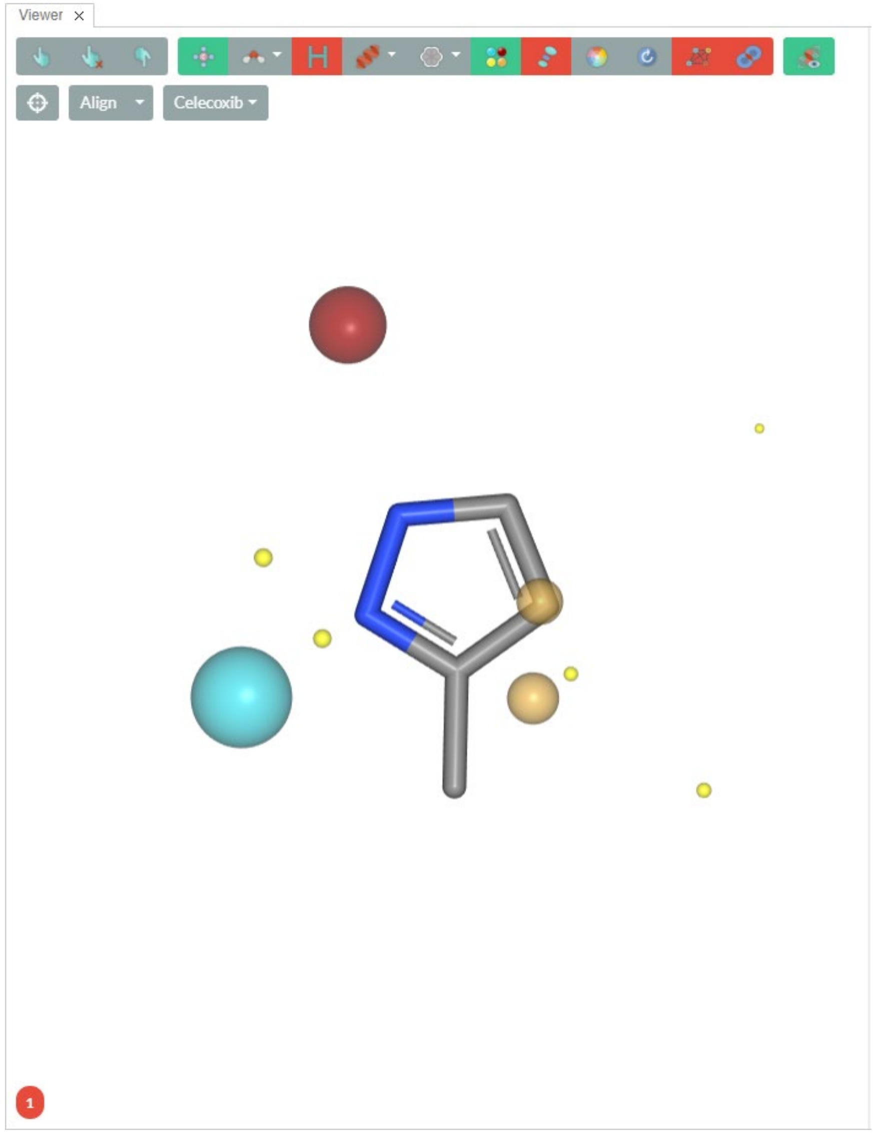
# How hard can it be?

- Start from a blank 2D sketcher canvas
  - sketch something



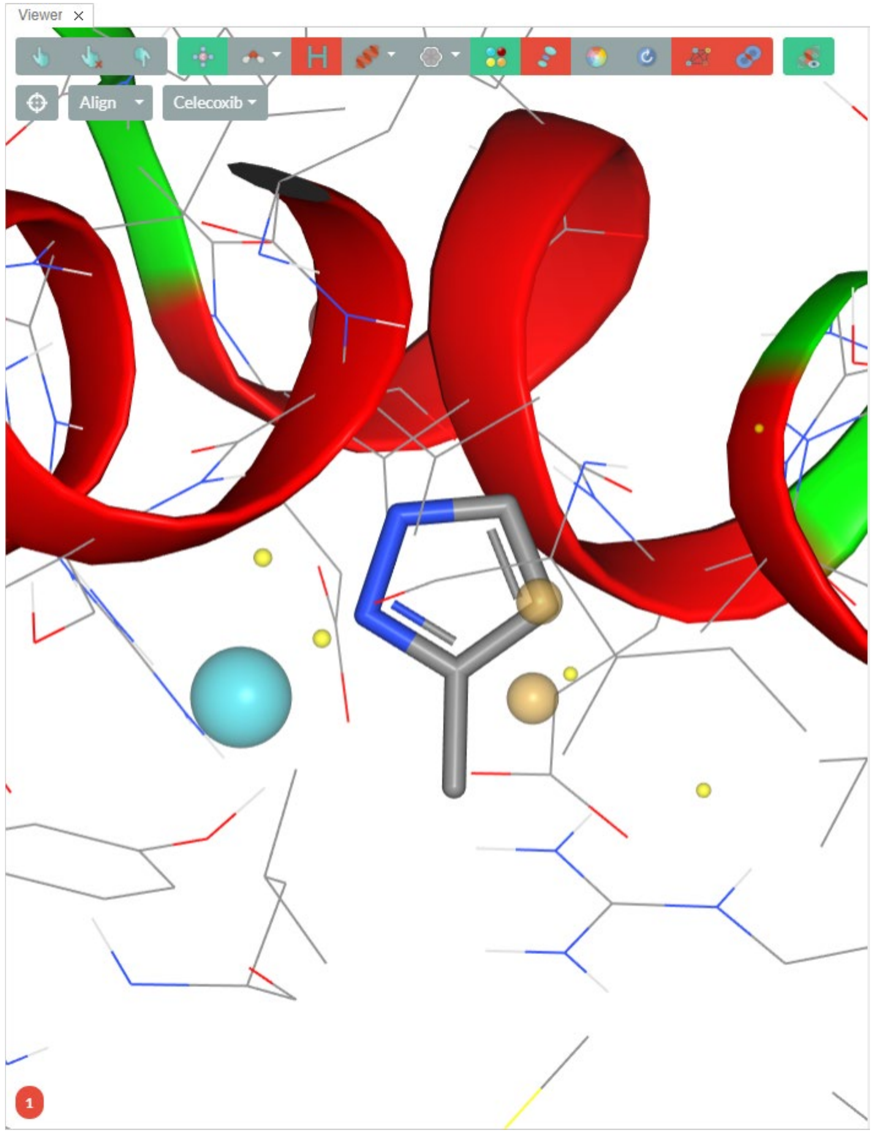
# How hard can it be?

- Start from a blank 2D sketcher canvas
  - sketch something
  - the largest 2D fragment is popped to a 3D conformation…



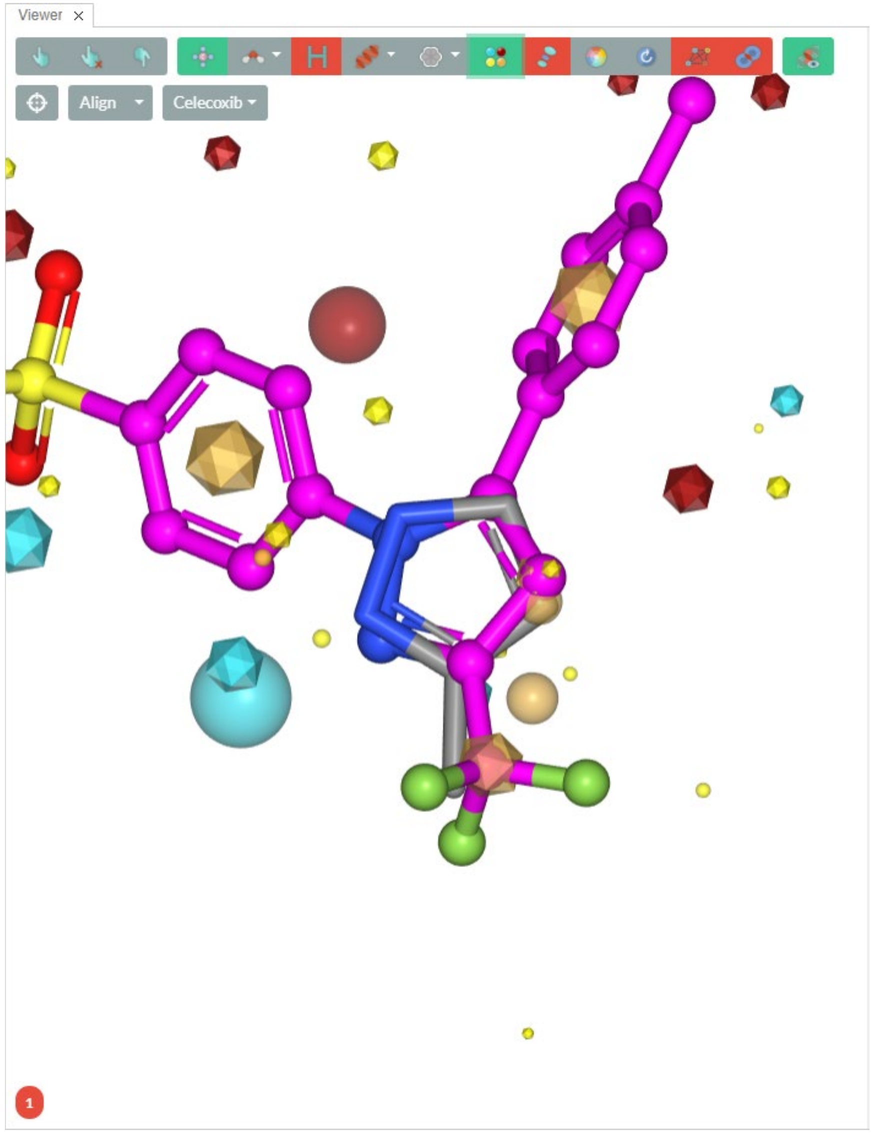
# How hard can it be?

- Start from a blank 2D sketcher canvas
  - sketch something
  - the largest 2D fragment is popped to a 3D conformation…
  - …and docked into the protein’s active site



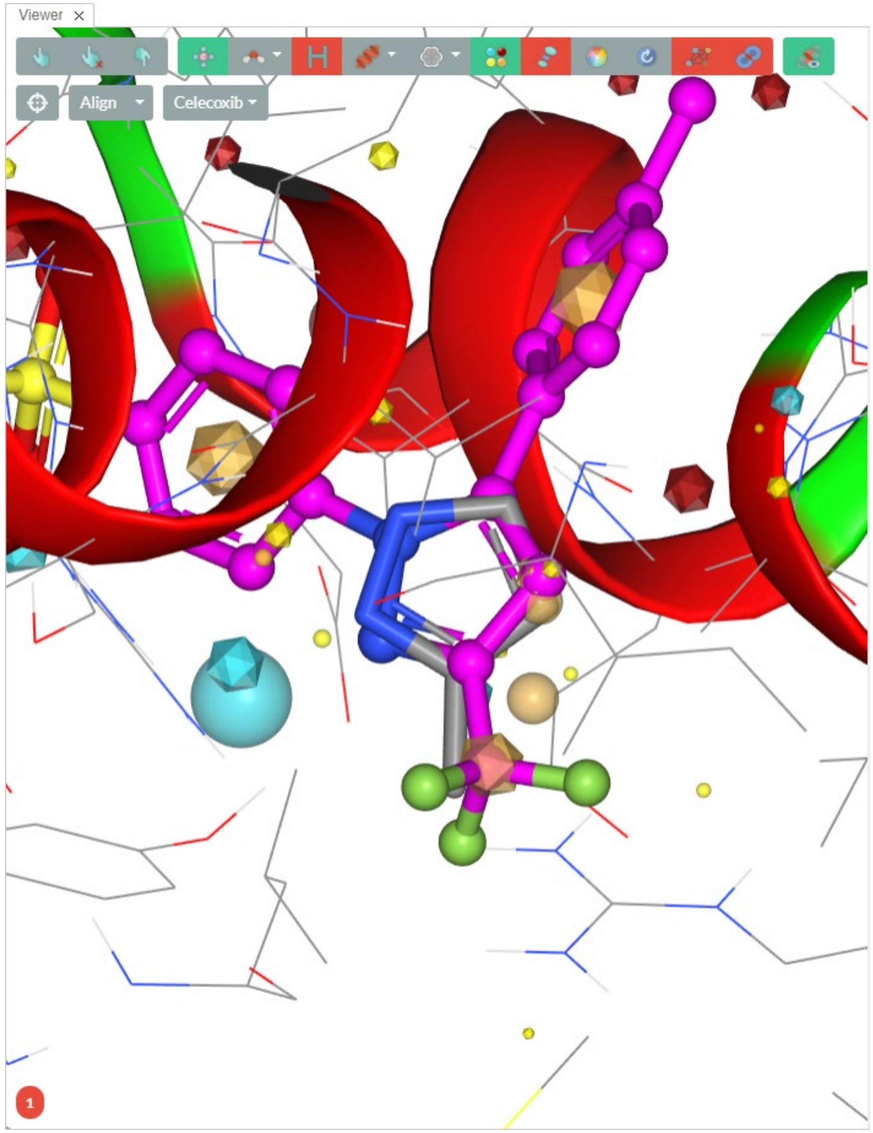
# How hard can it be?

- Start from a blank 2D sketcher canvas
  - sketch something
  - the largest 2D fragment is popped to a 3D conformation…
  - …and docked into the protein's active site
  - …or aligned against a reference using a combination of 3D fields and shape



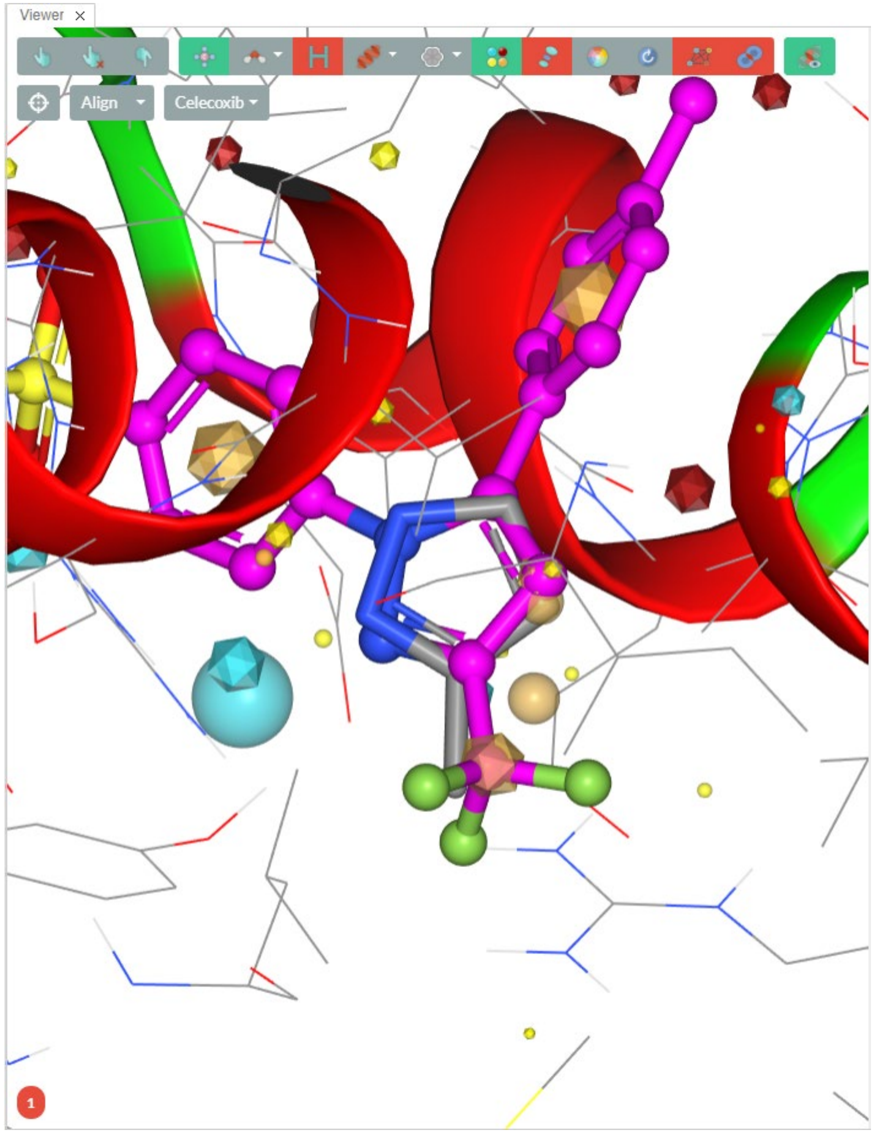
# How hard can it be?

- Start from a blank 2D sketcher canvas
  - sketch something
  - the largest 2D fragment is popped to a 3D conformation…
  - …and docked into the protein’s active site
  - …or aligned against a reference using a combination of 3D fields and shape
  - using the protein as excluded volume (if available)



# How hard can it be?

- Start from a blank 2D sketcher canvas
  - sketch something
  - the largest 2D fragment is popped to a 3D conformation…
  - …and docked into the protein's active site
  - …or aligned against a reference using a combination of 3D fields and shape
  - using the protein as excluded volume (if available)
  - then the *grow3D* process begins



# Outline

> Introduction to the project

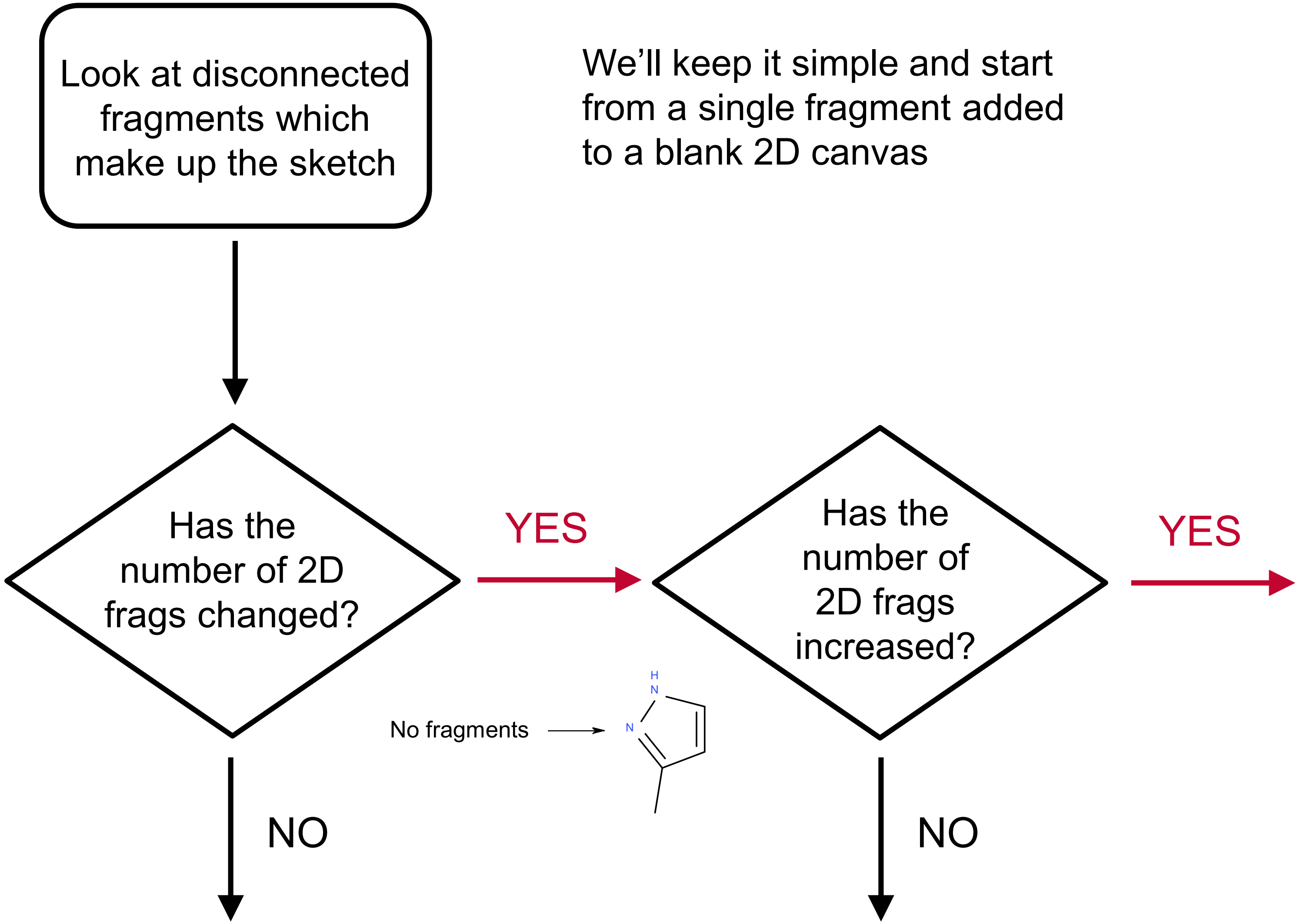
> Description of the *grow3D* algorithm

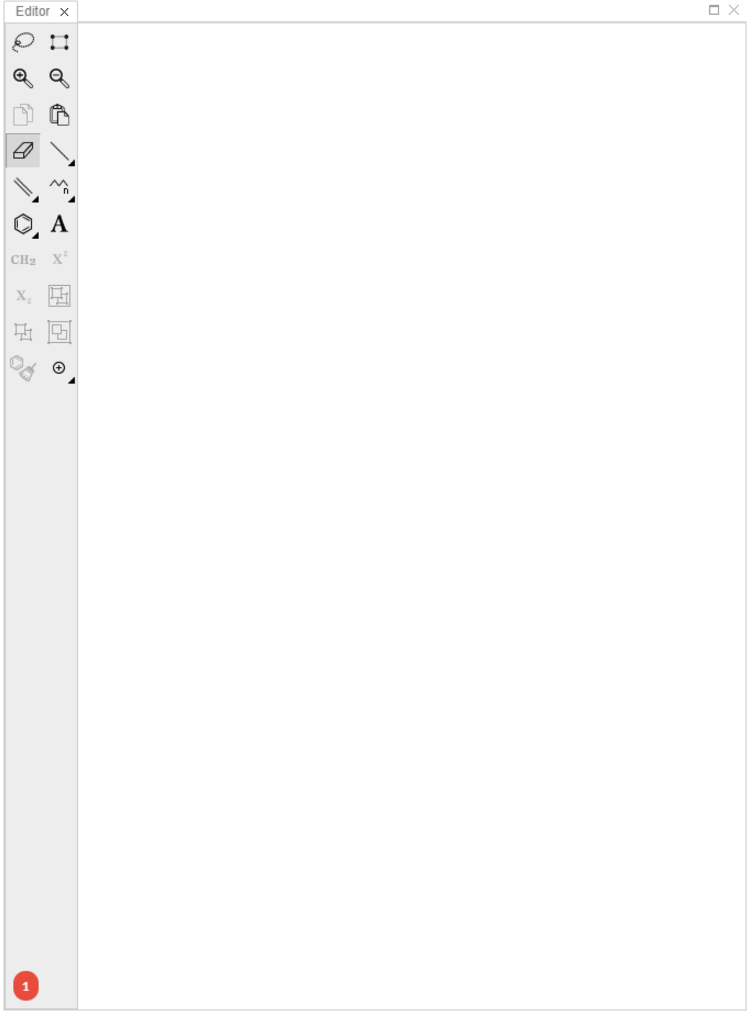
> The devil is in the details

> Conclusions and outlook

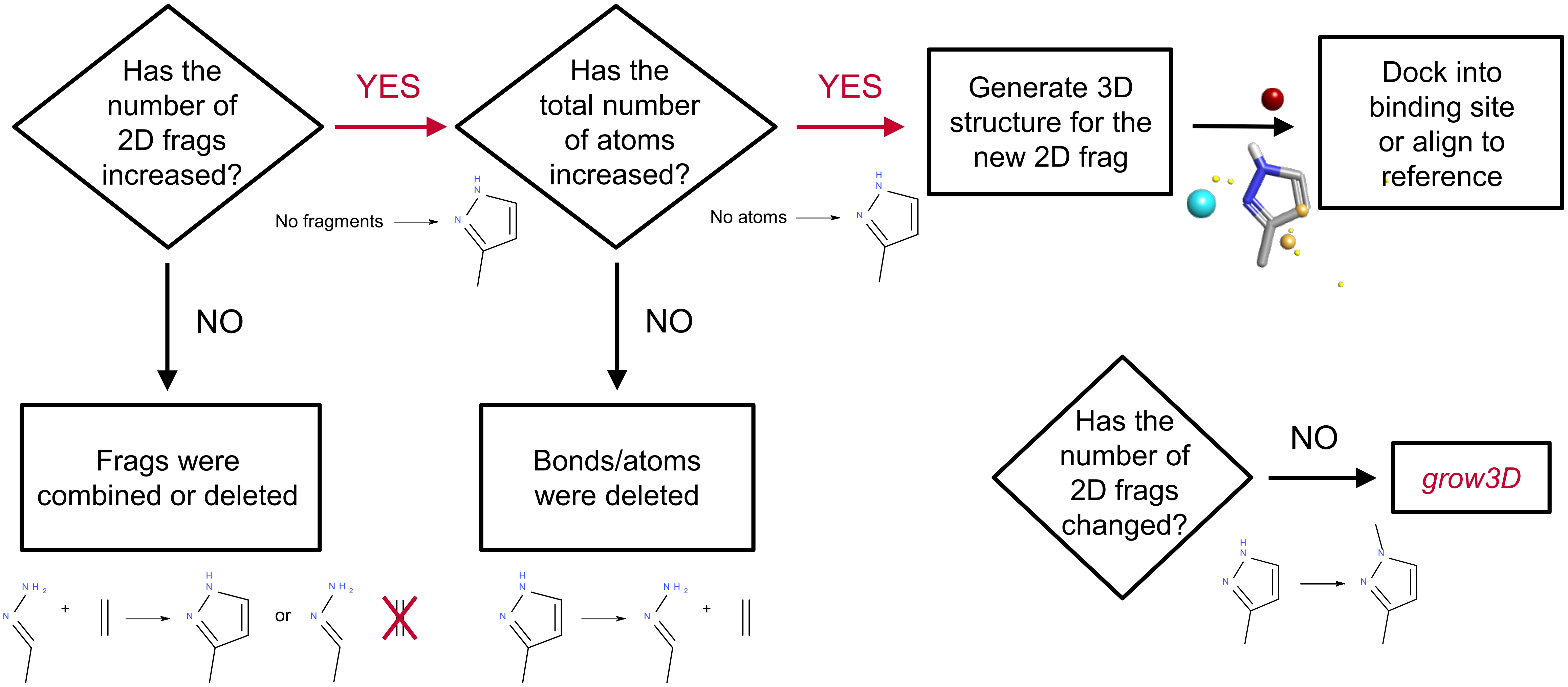
# A simple flow chart will help: start from blank

We'll keep it simple and start from a single fragment added to a blank 2D canvas


Look at disconnected fragments which make up the sketch
Has the number of 2D frags changed?
YES
Has the number of 2D frags increased?
YES
NO
NO
No fragments
H
N
N



# A simple flow chart will help: generate 3D or *grow3D*?



Has the number of 2D frags increased?

YES

Has the total number of atoms increased?

YES

Generate 3D structure for the new 2D frag

Dock into binding site or align to reference

NO

Frags were combined or deleted

NO

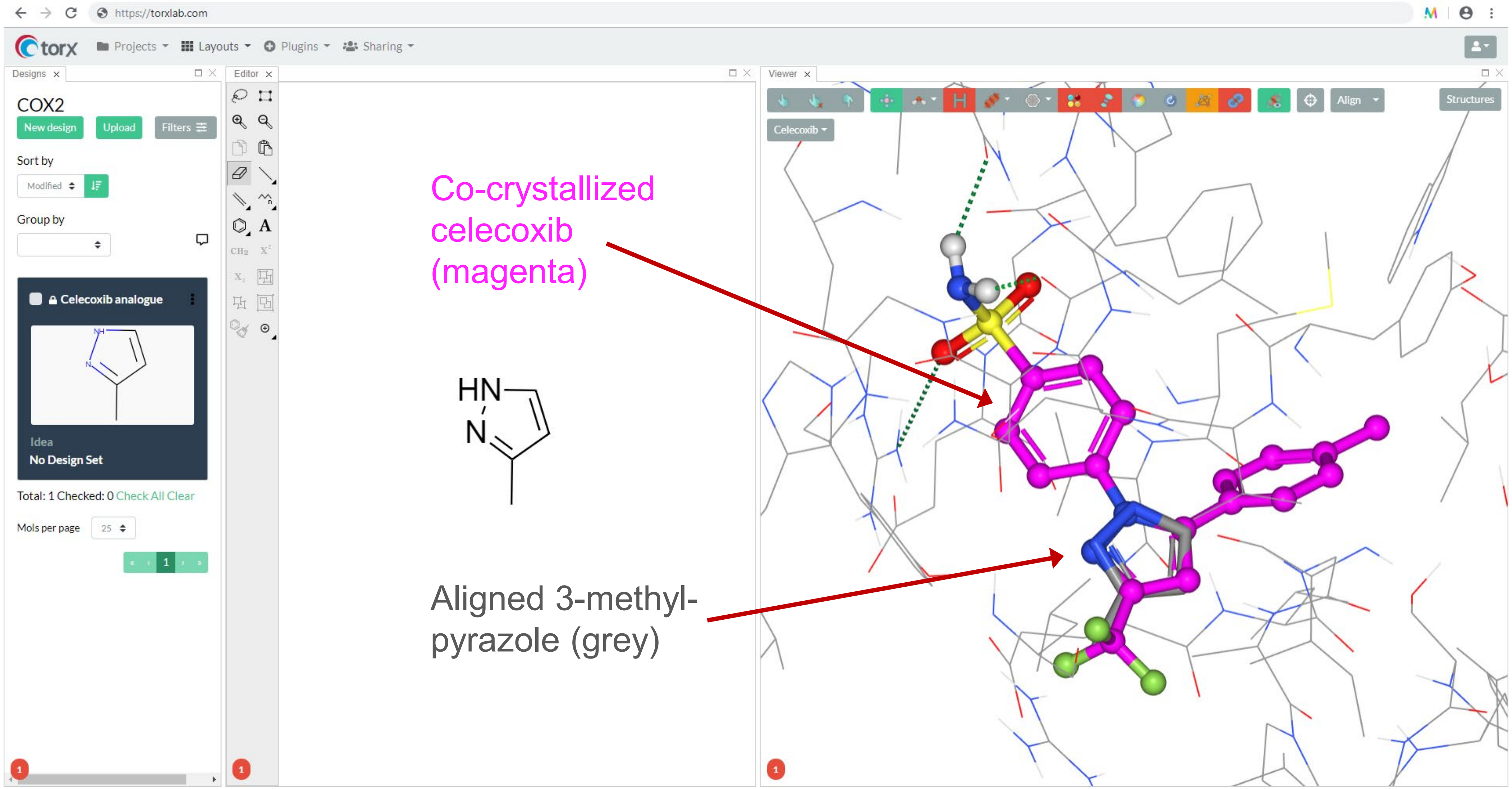
Bonds/atoms were deleted

Has the number of 2D frags changed?

NO

*grow3D*

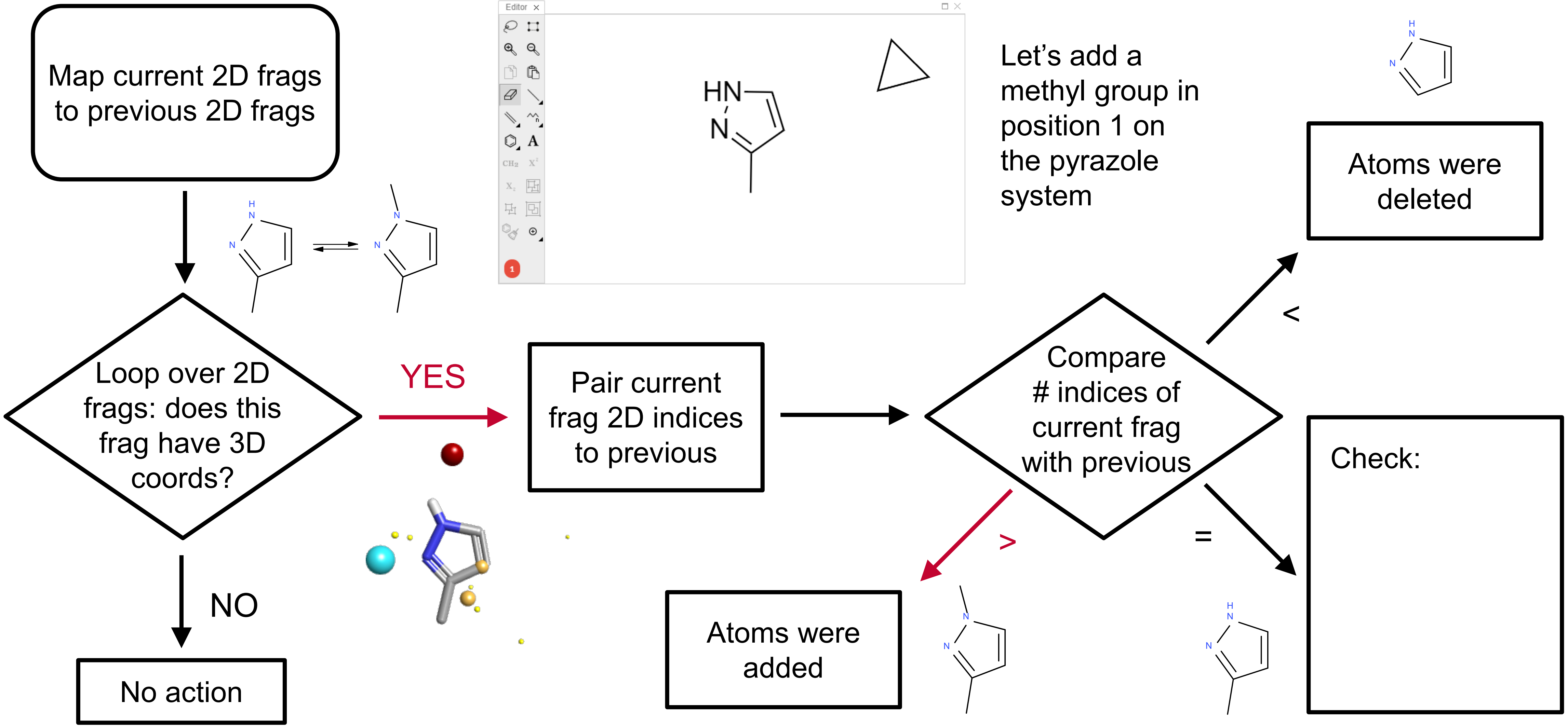
# Initial placement of the 3D design in the 3LN1 pocket



Co-crystallized celecoxib (magenta)

Aligned 3-methyl-pyrazole (grey)

# A simple flow chart will help: growing the 2D design



Map current 2D frags to previous 2D frags

Loop over 2D frags: does this frag have 3D coords?

YES

NO

No action

Pair current frag 2D indices to previous

Compare # indices of current frag with previous

<

>

=

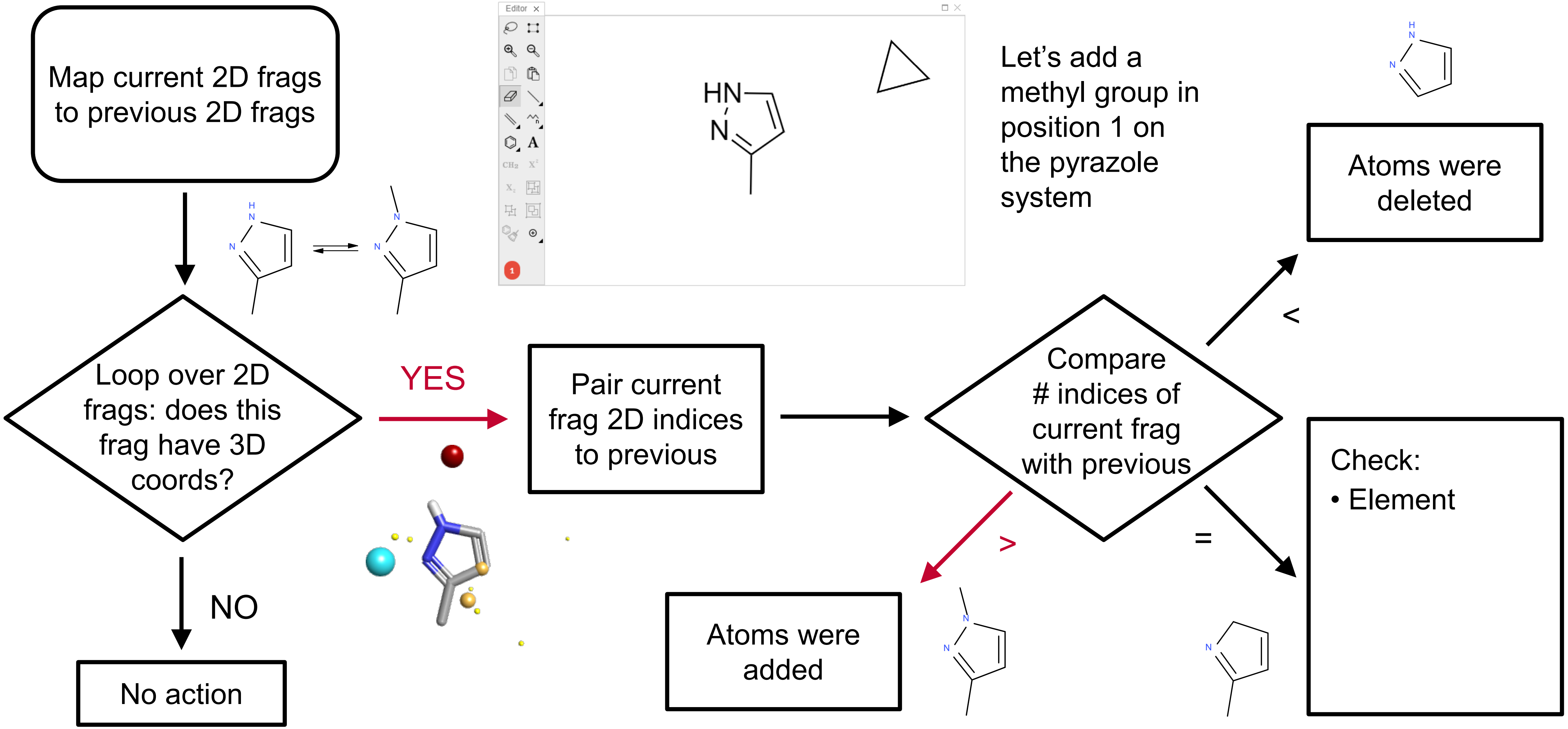
Atoms were deleted

Atoms were added

Check:

Let's add a methyl group in position 1 on the pyrazole system

# A simple flow chart will help: growing the 2D design



Map current 2D frags to previous 2D frags

Loop over 2D frags: does this frag have 3D coords?

YES

NO

No action

Pair current frag 2D indices to previous

Compare # indices of current frag with previous

<

>

=

Atoms were deleted

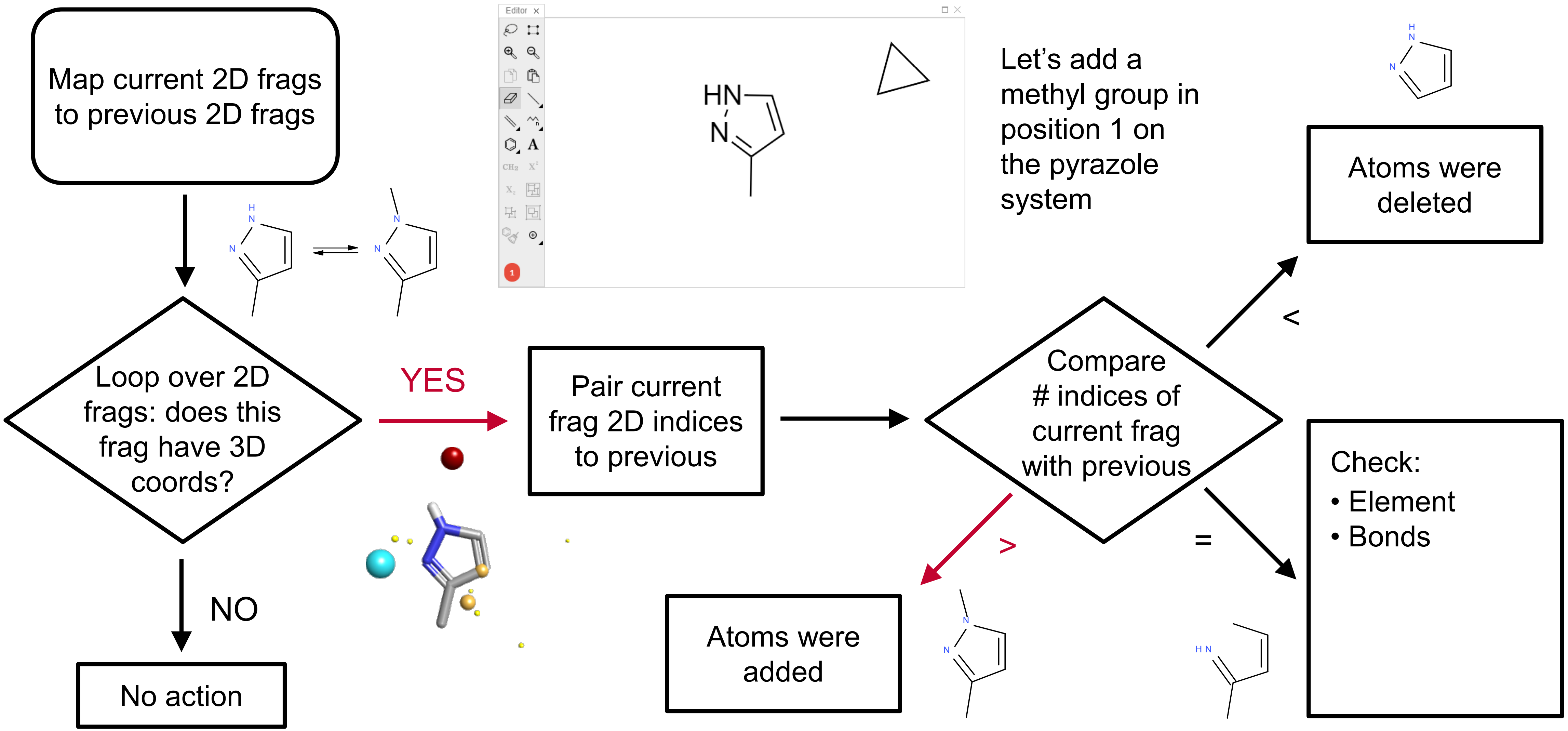
Atoms were added

Check:

- Element

Let’s add a methyl group in position 1 on the pyrazole system

# A simple flow chart will help: growing the 2D design



Map current 2D frags to previous 2D frags

Loop over 2D frags: does this frag have 3D coords?

YES

Pair current frag 2D indices to previous

Compare # indices of current frag with previous

<

Atoms were deleted

>

Atoms were added

=

Check:

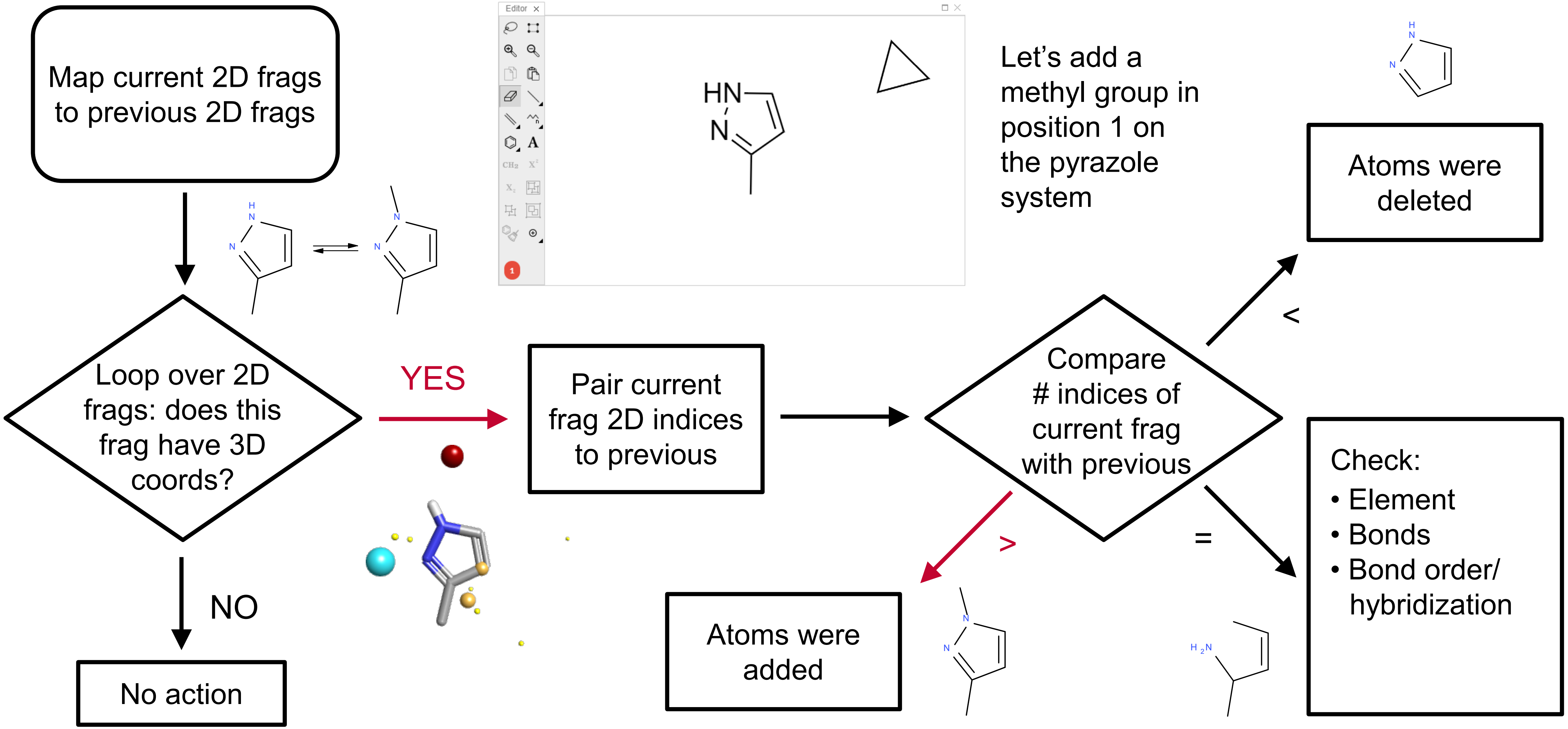
- Element
- Bonds

NO

No action

Let’s add a methyl group in position 1 on the pyrazole system

# A simple flow chart will help: growing the 2D design



Map current 2D frags to previous 2D frags

Loop over 2D frags: does this frag have 3D coords?

YES

NO

No action

Pair current frag 2D indices to previous

Compare # indices of current frag with previous

<

>

=

Atoms were deleted

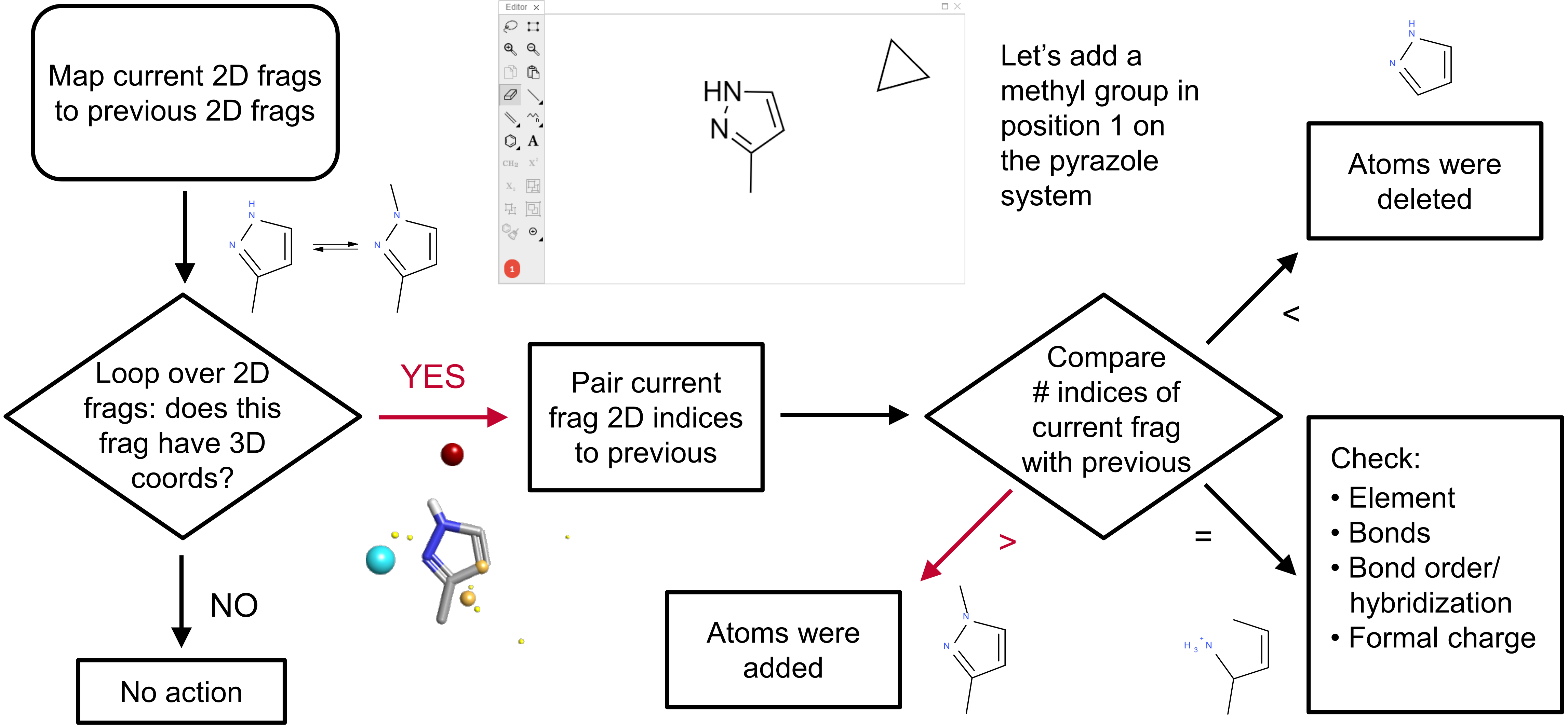
Atoms were added

Check:

- Element
- Bonds
- Bond order/ hybridization

Let’s add a methyl group in position 1 on the pyrazole system

# A simple flow chart will help: growing the 2D design



Map current 2D frags to previous 2D frags

Loop over 2D frags: does this frag have 3D coords?

YES

NO

No action

Pair current frag 2D indices to previous

Let's add a methyl group in position 1 on the pyrazole system

Compare # indices of current frag with previous

<

Atoms were deleted

>

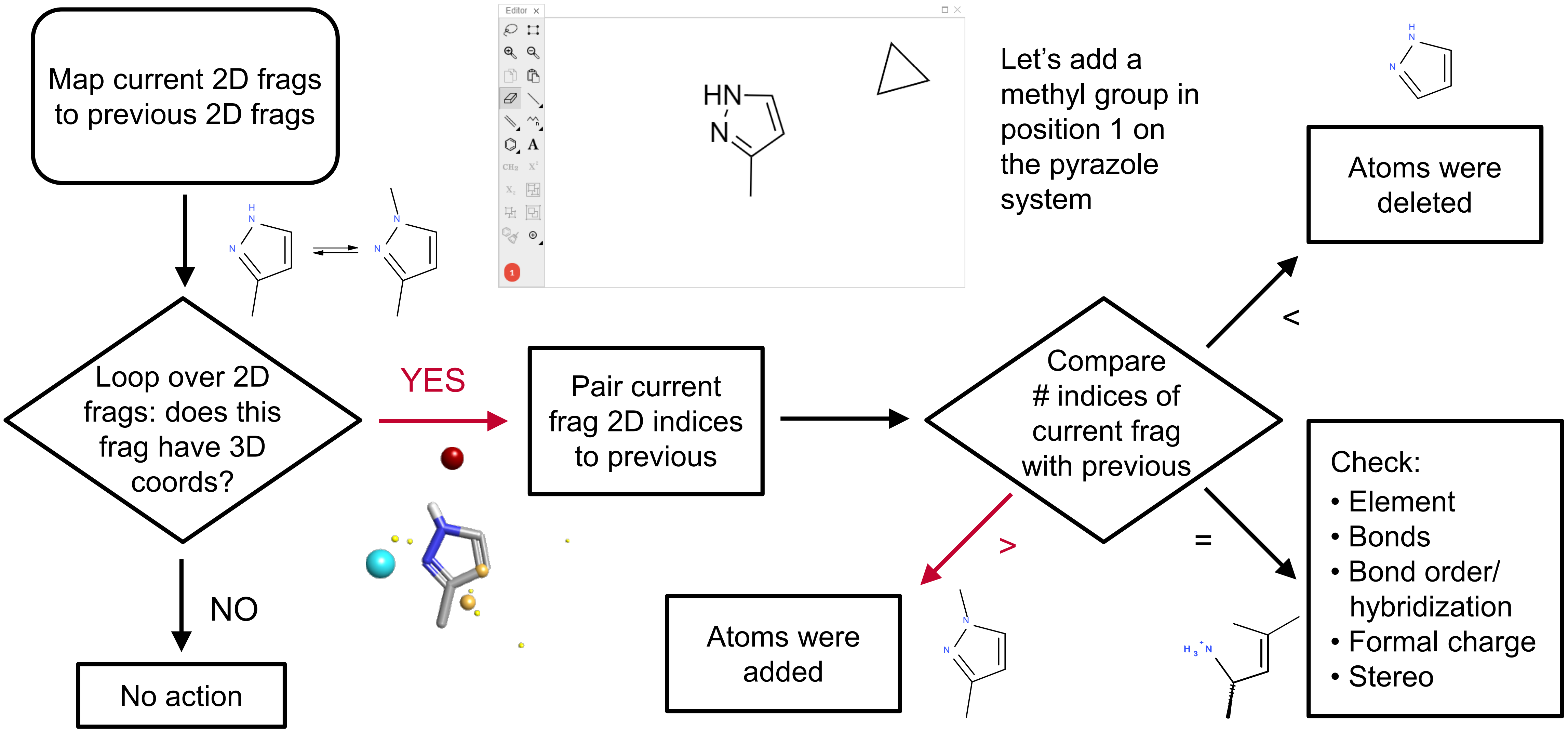
Atoms were added

=

Check:

- Element
- Bonds
- Bond order/ hybridization
- Formal charge

# A simple flow chart will help: growing the 2D design



Map current 2D frags to previous 2D frags

Loop over 2D frags: does this frag have 3D coords?

YES

Pair current frag 2D indices to previous

Compare # indices of current frag with previous

<

Atoms were deleted

>

Atoms were added

=

Check:

- Element
- Bonds
- Bond order/ hybridization
- Formal charge
- Stereo

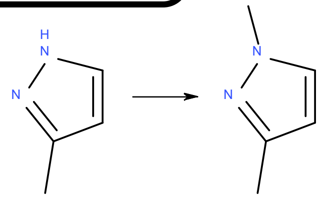
NO

No action

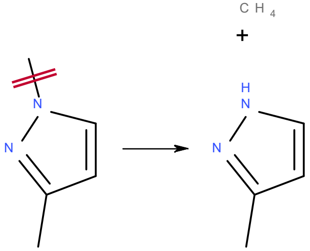
Let's add a methyl group in position 1 on the pyrazole system

# A simple flow chart will help: growing the 3D design

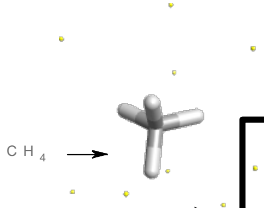
A methyl group was added



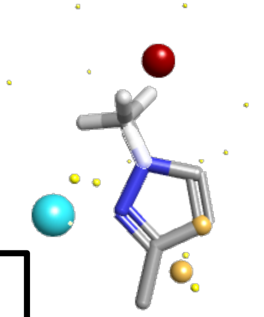
- Find bonds to added atoms
- Delete those bonds and cap with hydrogens



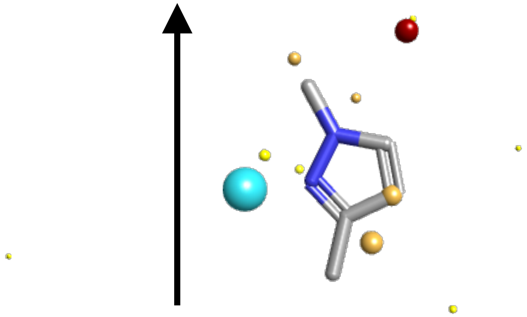
Generate 3D coords for newly added 2D frag



Align frag Hs with parent's heavy atoms and vice versa



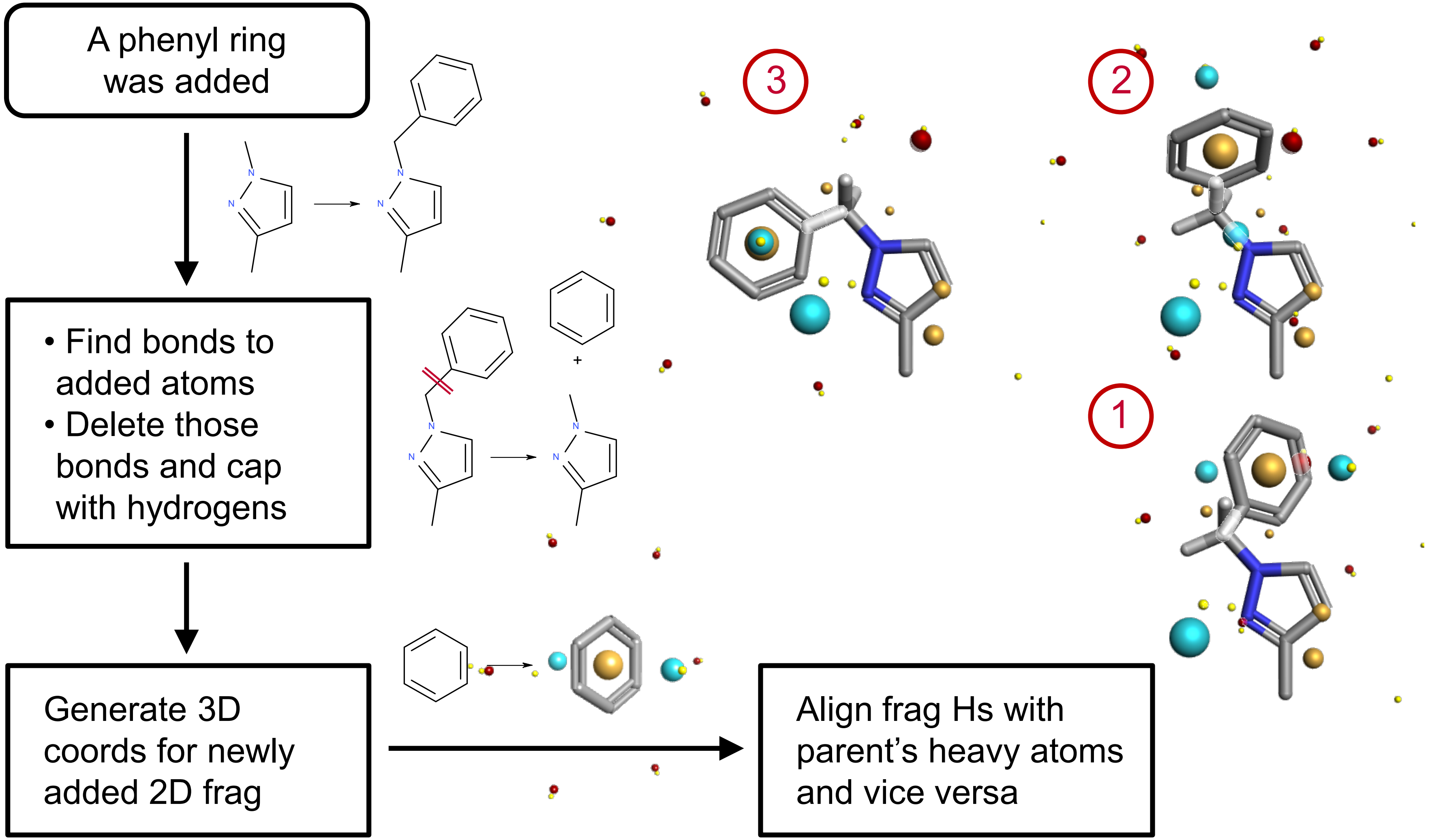
Attach to parent with appropriate bond order



Ready for the next *grow3D* iteration!

Methyl was easy, can I have more of those, please?

# A simple flow chart will help: scoring 3D designs



A phenyl ring was added

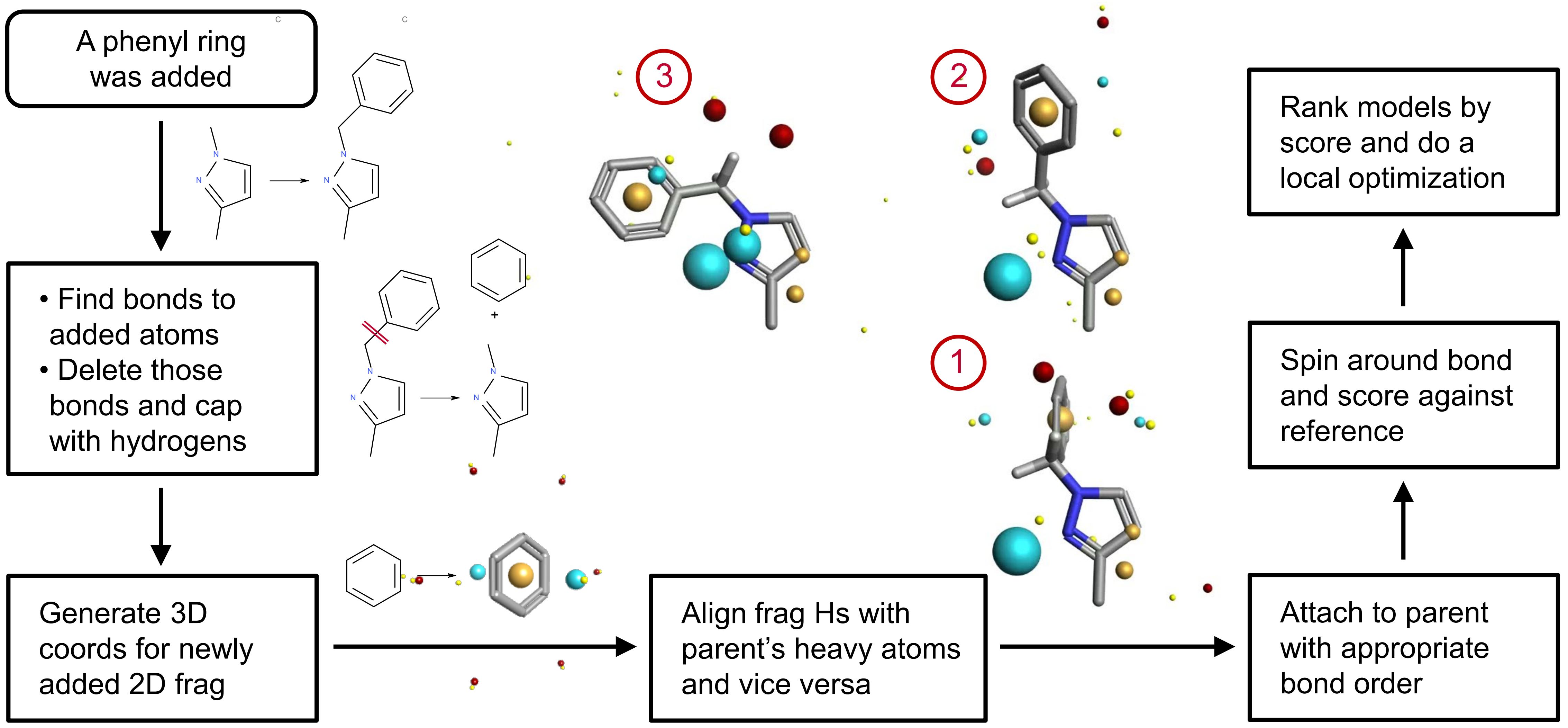
- Find bonds to added atoms
- Delete those bonds and cap with hydrogens

Generate 3D coords for newly added 2D frag

Align frag Hs with parent’s heavy atoms and vice versa

# A simple flow chart will help: scoring 3D designs

RDKit Open-Source Cheminformatics and Machine Learning



A phenyl ring was added

- Find bonds to added atoms
- Delete those bonds and cap with hydrogens

Generate 3D coords for newly added 2D frag

Align frag Hs with parent's heavy atoms and vice versa

Attach to parent with appropriate bond order

Spin around bond and score against reference

Rank models by score and do a local optimization

cresset software

# Outline

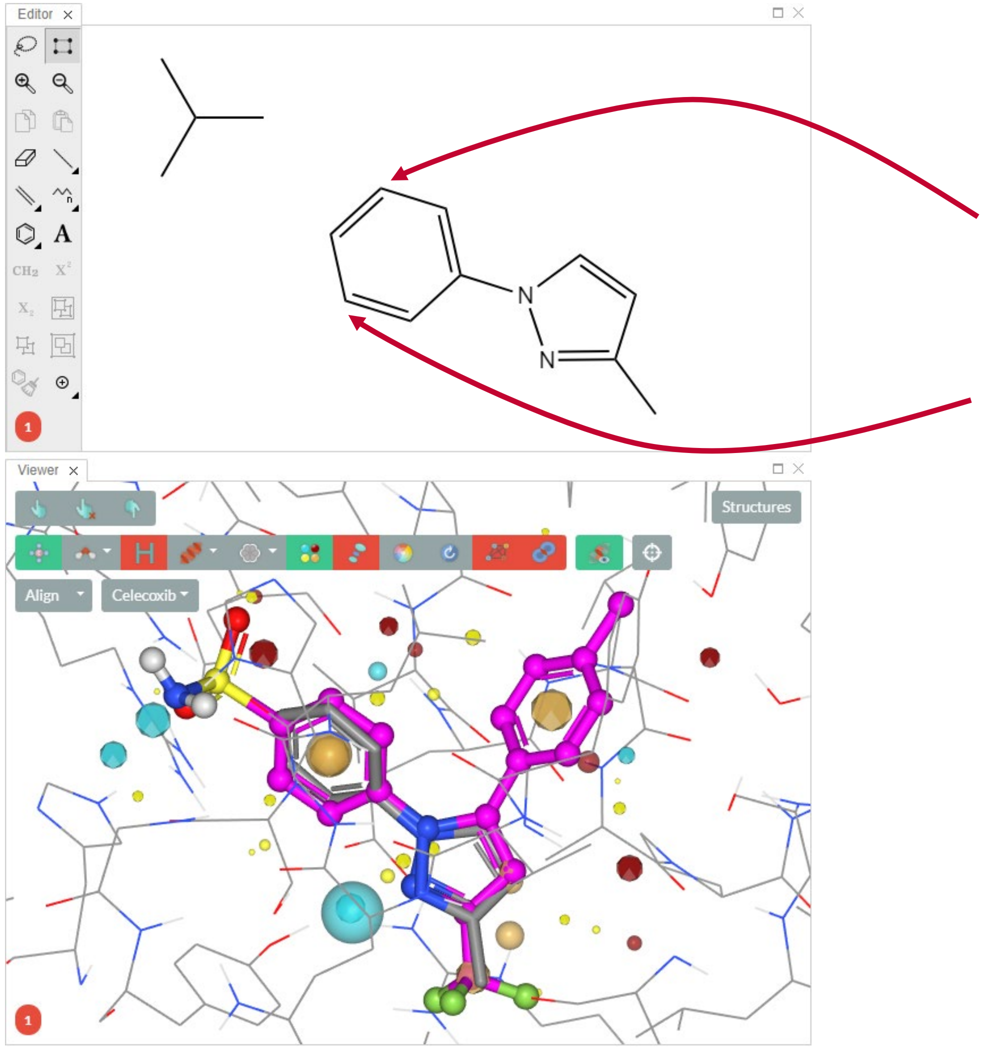
> Introduction to the project

> Description of the *grow3D* algorithm

> The devil is in the details

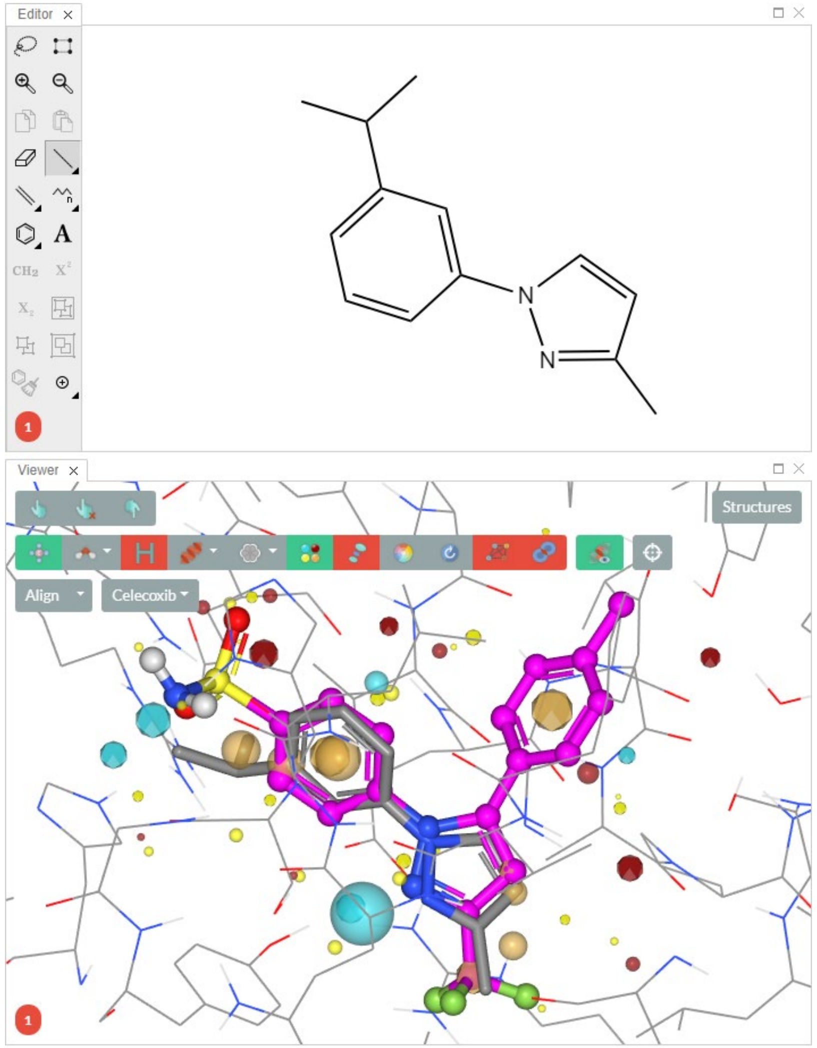
> Conclusions and outlook

# The devil is in the details (1): symmetries

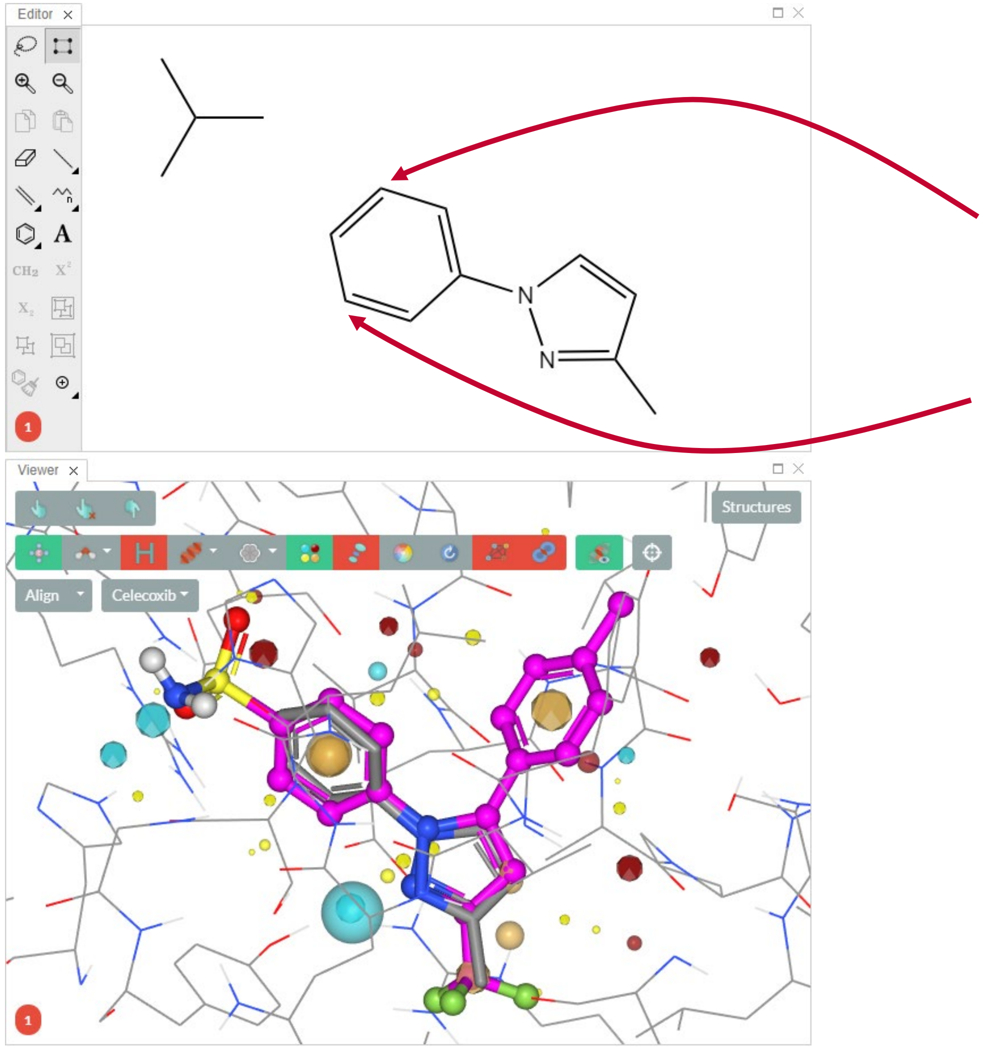


When adding a *m*-isopropyl substituent in 2D, there are actually two symmetry-equivalent positions it might fit in 3D

Let’s add it on one side first…

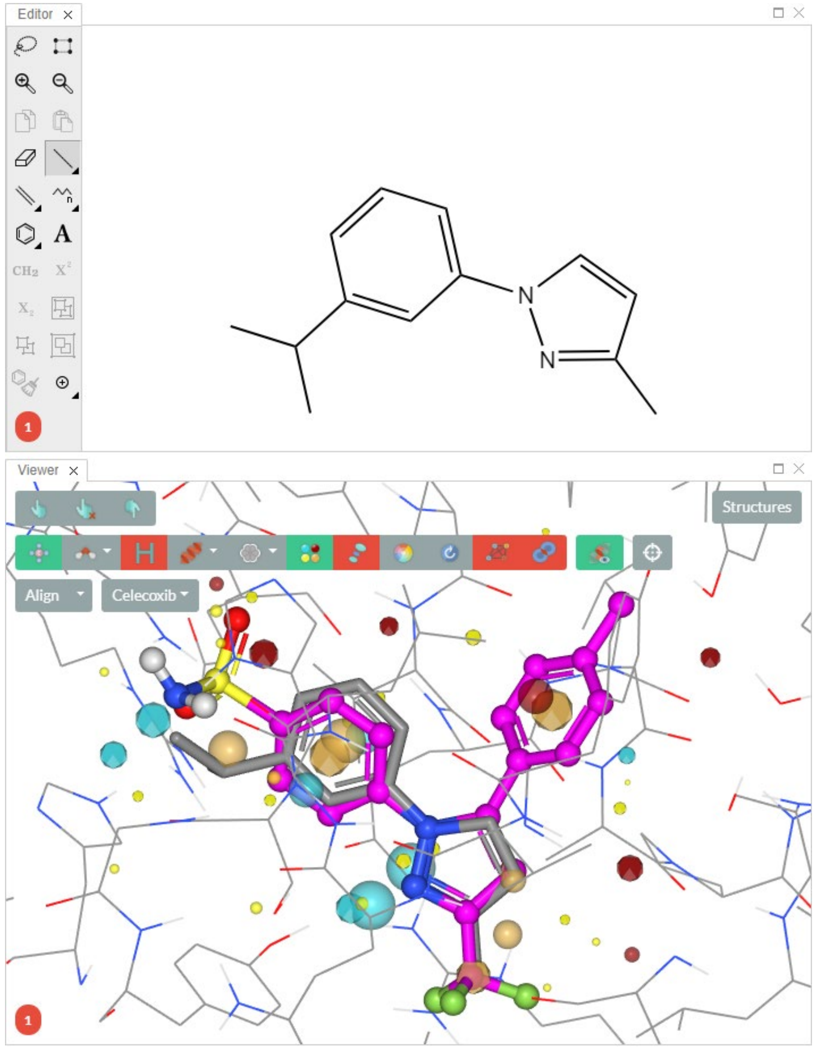


# The devil is in the details (1): symmetries

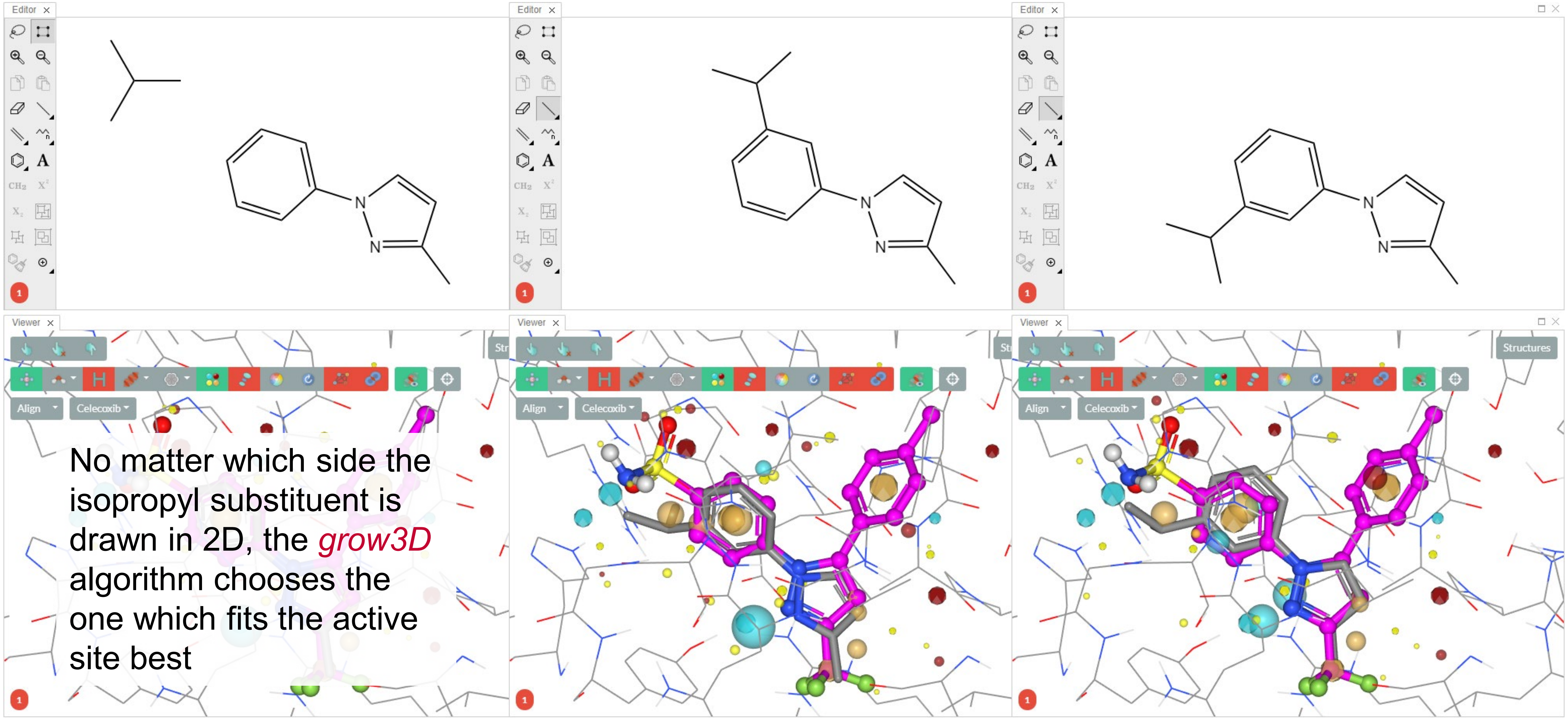


When adding a *m*-isopropyl substituent in 2D, there are actually two symmetry-equivalent positions it might fit in 3D

…and then on the other side

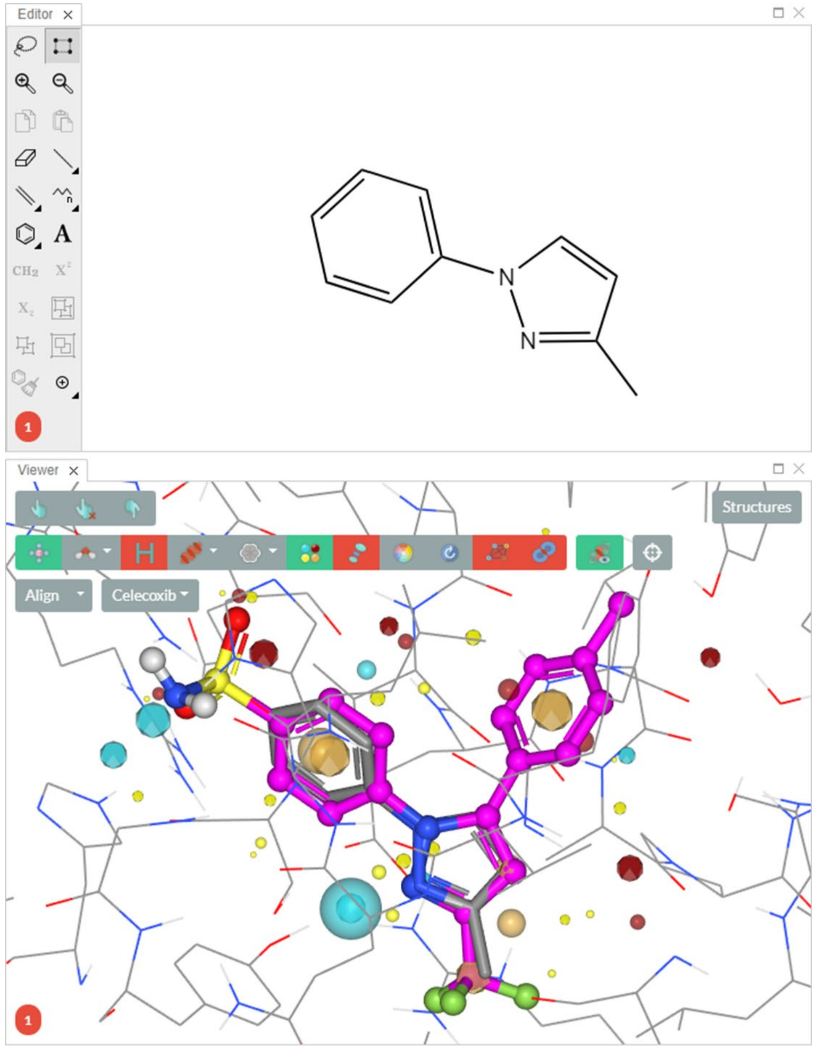


# The devil is in the details (1): symmetries



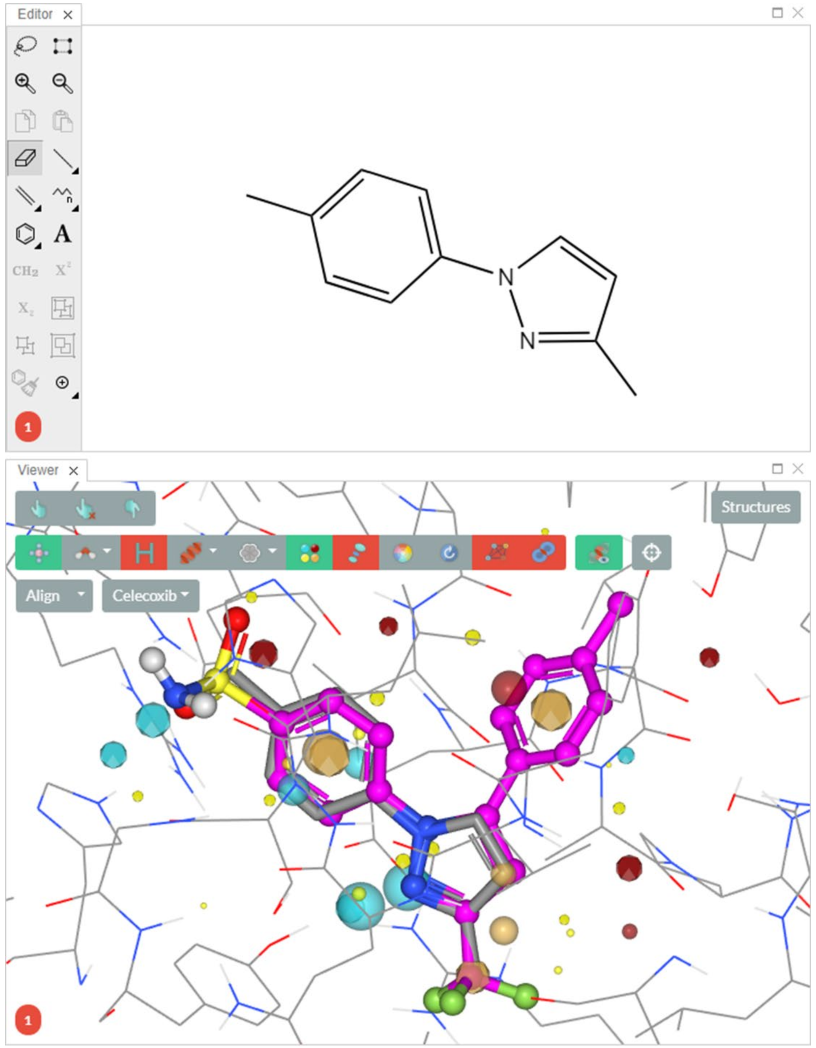
No matter which side the isopropyl substituent is drawn in 2D, the *grow3D* algorithm chooses the one which fits the active site best

# The devil is in the details (2): chemically invalid states



When we draw a sulfone-containing functional group in 2D, we tend to go through a number of chemically invalid states

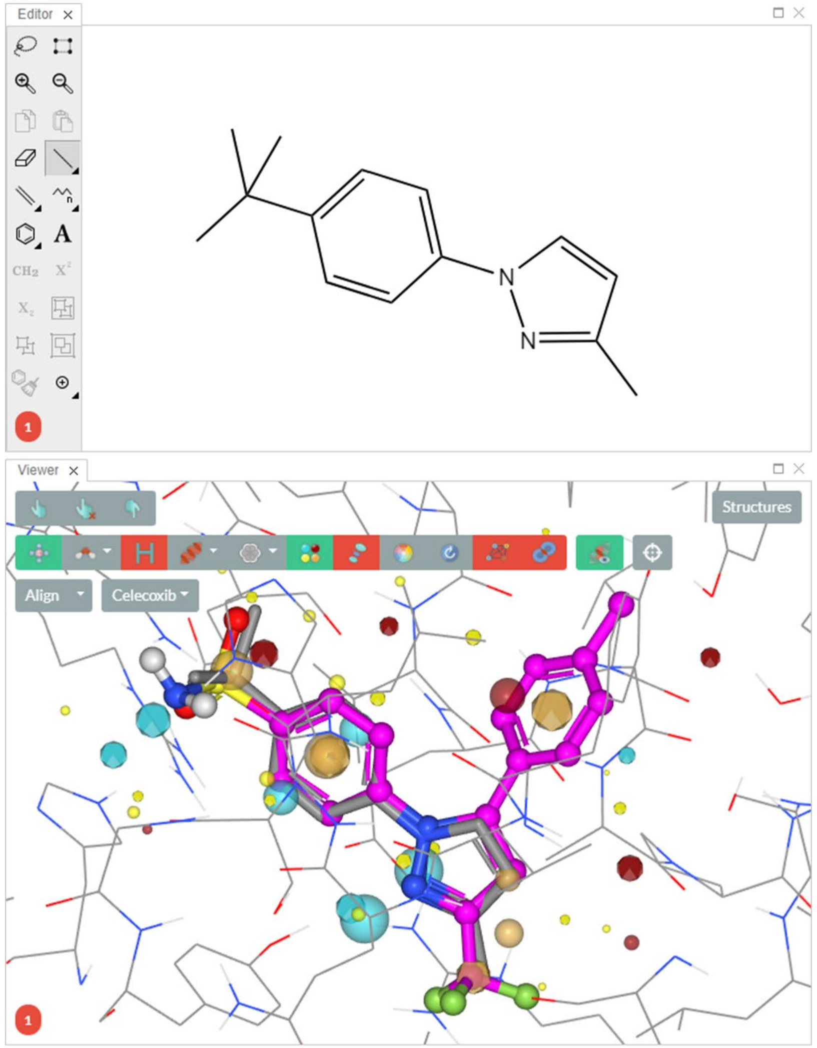
# The devil is in the details (2): chemically invalid states



When we draw a sulfone-containing functional group in 2D, we tend to go through a number of chemically invalid states

We draw a methyl…

# The devil is in the details (2): chemically invalid states

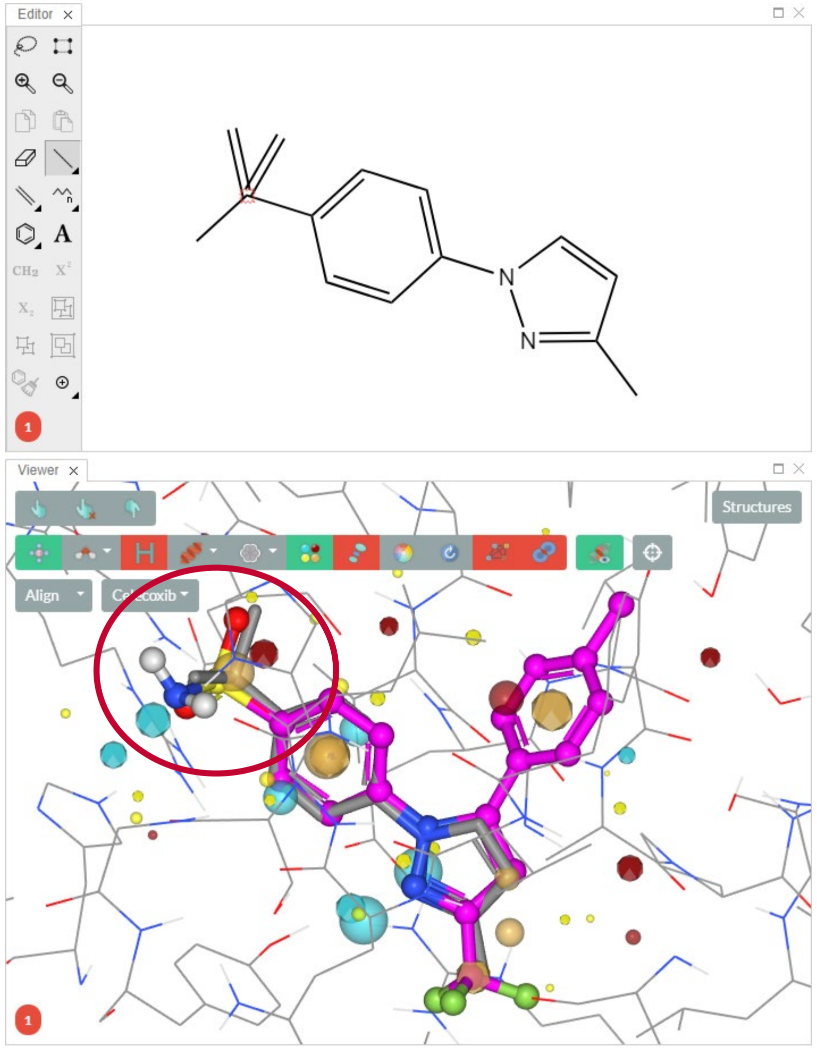


When we draw a sulfone-containing functional group in 2D, we tend to go through a number of chemically invalid states

We draw a methyl…

…then we keep adding methyl groups until we get to a *t*-butyl…

# The devil is in the details (2): chemically invalid states



When we draw a sulfone-containing functional group in 2D, we tend to go through a number of chemically invalid states

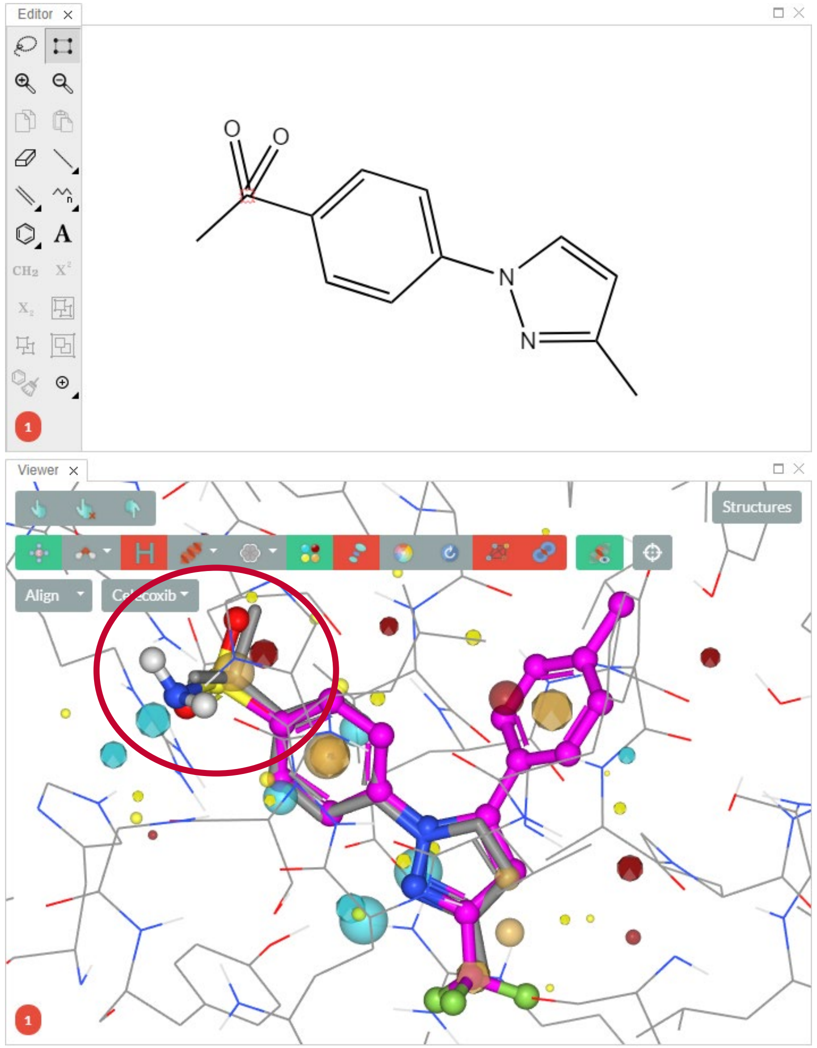
We draw a methyl…

…then we keep adding methyl groups until we get to a *t*-butyl…

…then we add the double bonds…

*grow3D* pauses,
waiting for our next move

# The devil is in the details (2): chemically invalid states



When we draw a sulfone-containing functional group in 2D, we tend to go through a number of chemically invalid states

We draw a methyl…

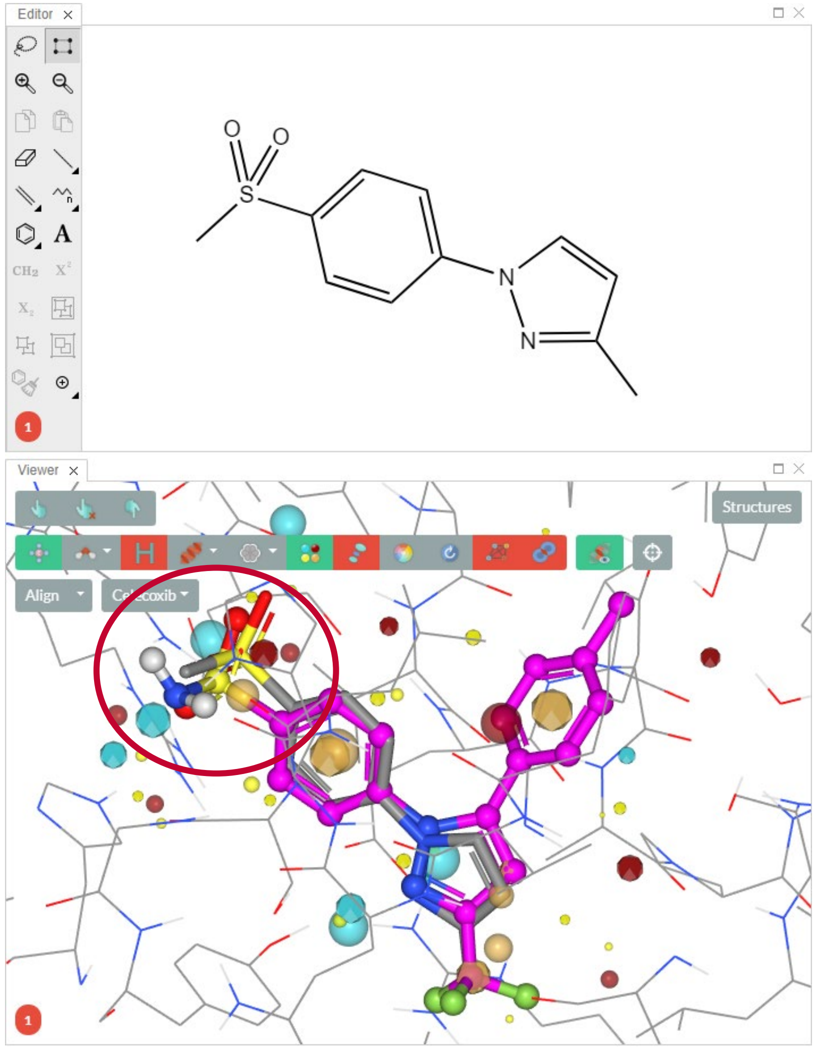
…then we keep adding methyl groups until we get to a *t*-butyl…

…then we add the double bonds…

…then we turn terminal methylenes into oxygens…

*grow3D* still patiently waiting for some decent chemistry

# The devil is in the details (2): chemically invalid states



When we draw a sulfone-containing functional group in 2D, we tend to go through a number of chemically invalid states

We draw a methyl…

…then we keep adding methyl groups until we get to a *t*-butyl…

…then we add the double bonds…

…then we turn terminal methylenes into oxygens…

…and only in the end we turn the central carbon into a sulfur

*grow3D* applies the change

# Outline

- > Introduction to the project
- > Description of the workflow
- > The devil is in the details
- > Conclusions and outlook

# Conclusions

- *grow3D* is an algorithm which aims at generating sensible 3D poses in response to edits to a 2D sketch

- Poses are scored against their ability to fit a binding site and/or mimic electrostatics and shape of a reference ligand

- The algorithm is enough quick and accurate to enable real time assessment of design ideas

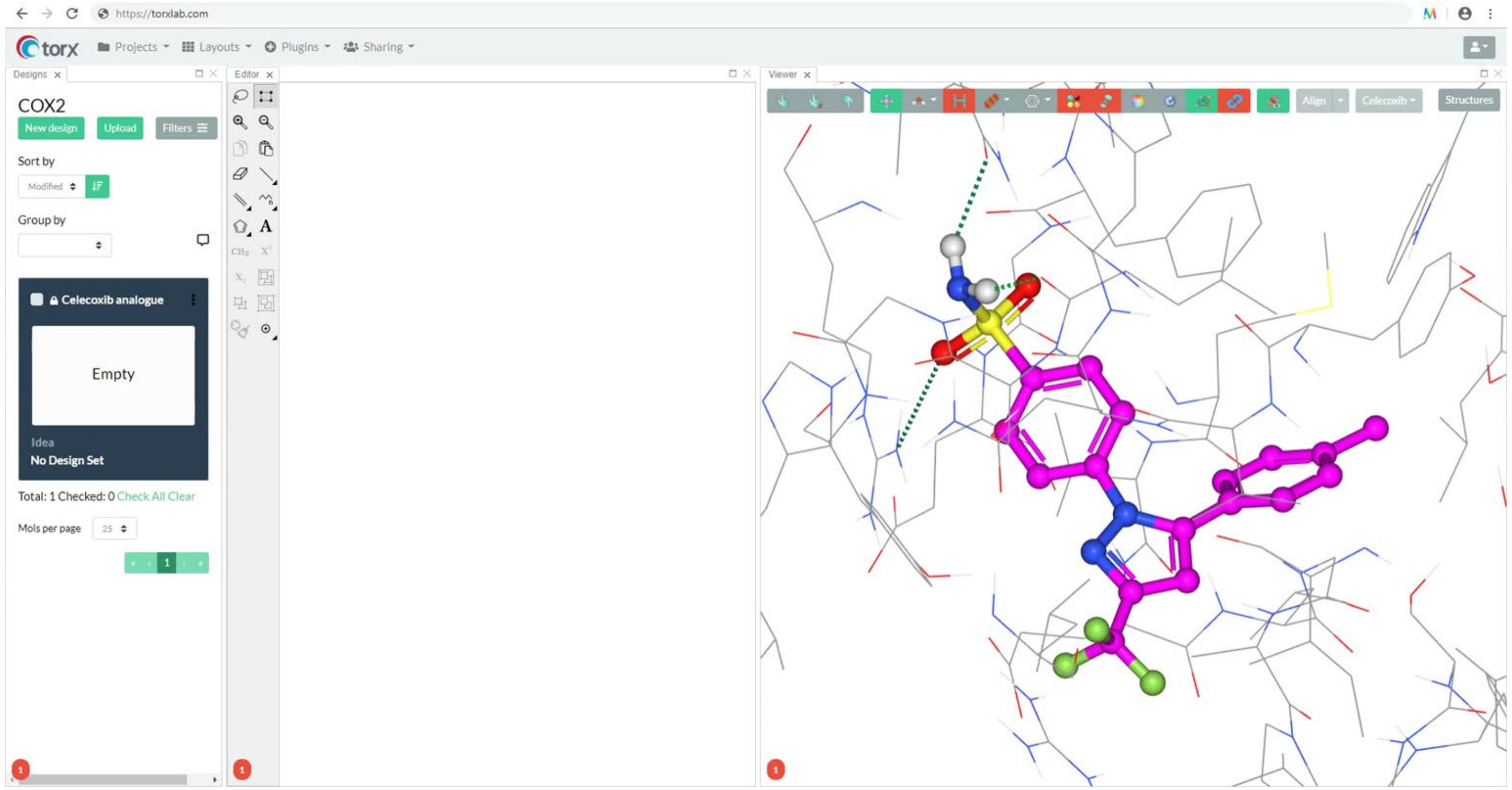
# The simple flow chart (to be continued)

cresset®
software

innovative science • intuitive software

# Thank you for your attention

paolo@cresset-group.com



COX2

Sort by: Modified

Group by:

Celecoxib analogue

Empty

Idea: No Design Set

Total: 1 Checked: 0 Check All Clear

Mols per page: 25

## Acknowledgments

@Cresset: James Foley, Harry Jubb, Mark Mackey, Tim Cheeseright

The RDKit

cressetgroup