# Towards Formal Reasoning about Molecular Pathways in HOL

Sohaib Ahmad, Osman Hasan and Umair Siddique

System Analysis & Verification (SAVe) Lab,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan

WETICE/VSC 2014
Parma, Italy
June 25, 2014

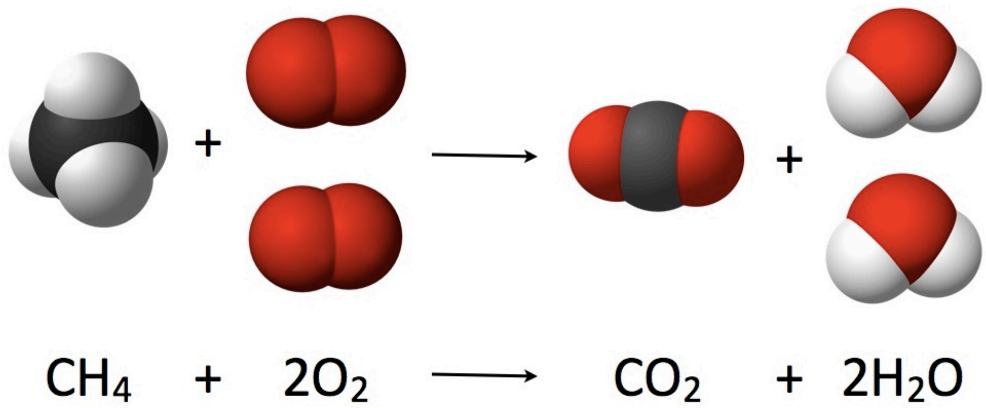
# Outline

- Introduction
- Proposed Methodology
- Formalization Details
- Case Studies
- Conclusions

# Chemical Reactions

- Transformation of one set of molecules to another

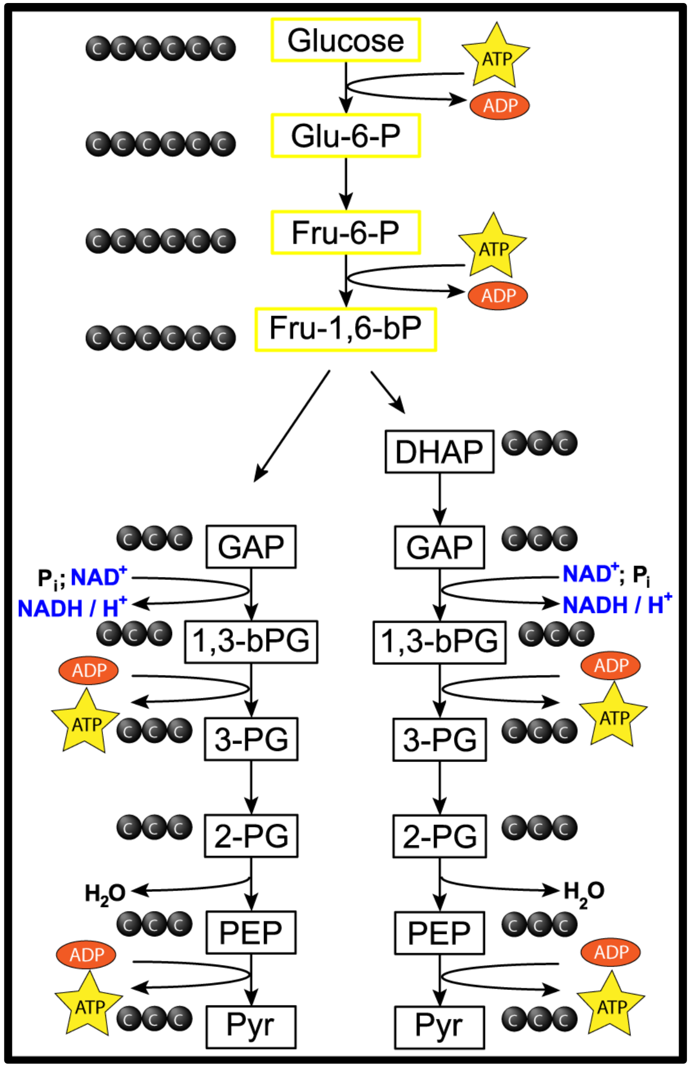
- Example: Burning of natural gas
  - Methane + Oxygen ➔ Carbon Dioxide + Water



# Molecular Pathways

- Series of chemical reactions

- Study of Molecular Pathways (Pathway Analysis)
  - What can be obtained from a bunch of reacting molecules
  - Vital role in developing effective drugs for various human infectious diseases
    - Critical System

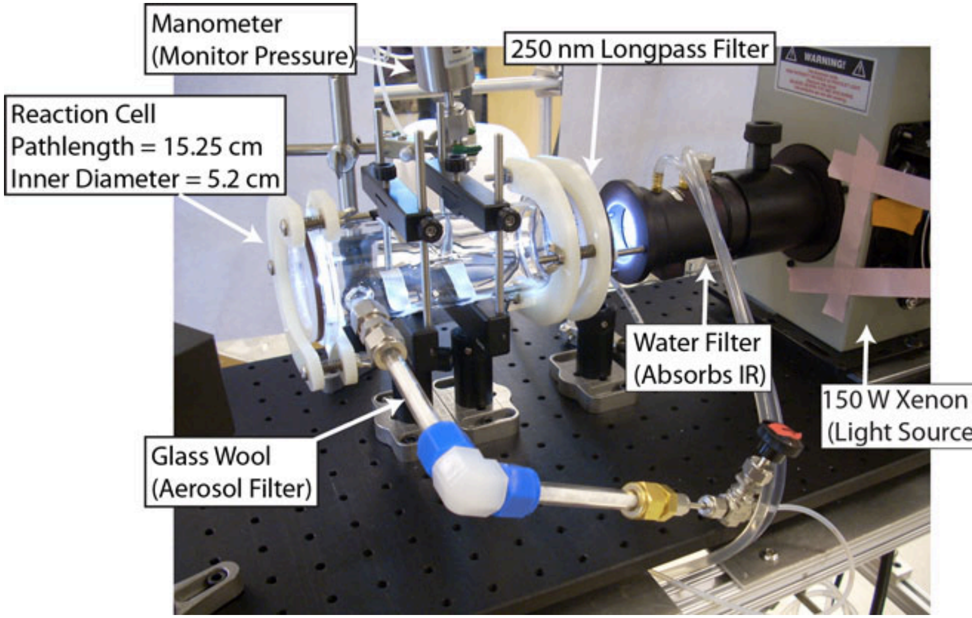


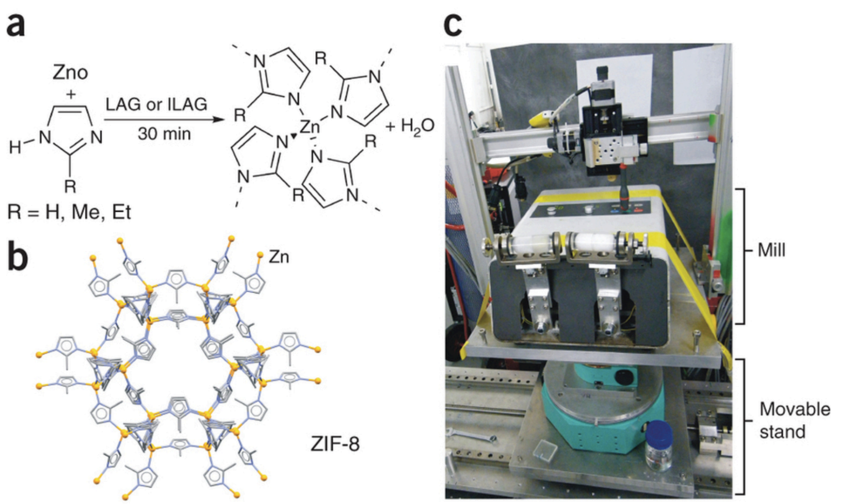

Glucose to Pyruvate

# Molecular Pathway Analysis

- Traditionally done in a wet-lab setting

- Slow
- Expensive
- Unpredictable Behavior?



# Off-Lab Molecular Pathway Analysis Requirements

- Deduction System
  - To reason about the reaction outcomes

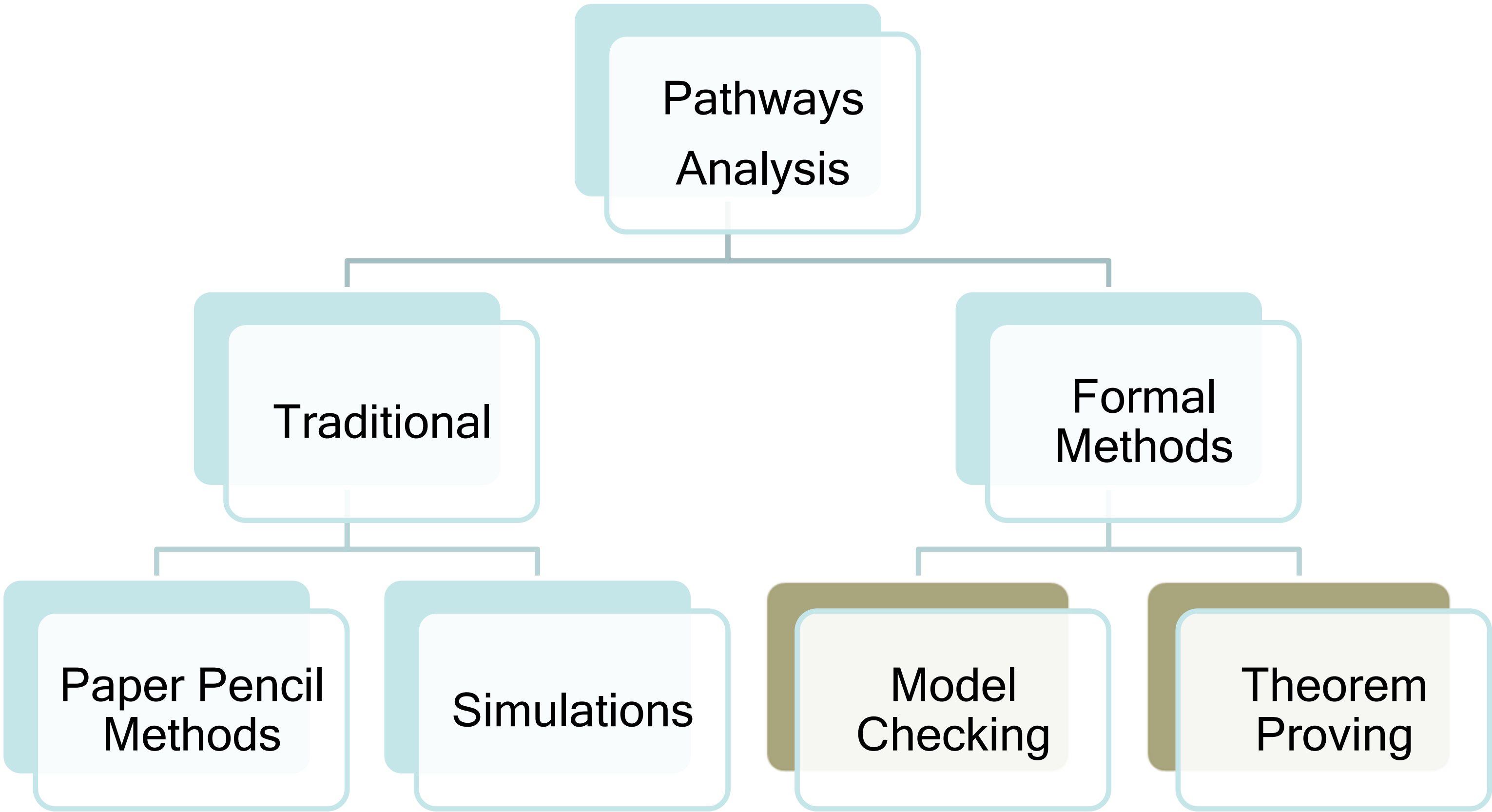
- Reaction Kinetics
  - Concentration of molecules at any time
  - Modeled by Differential Equations

$$\frac{dn(t)}{dt} = \lambda * n(t)$$

- Stochastics
  - Probability that a molecule decays within a small time interval

$$p_1 = \lambda \Delta t$$

# Types of Analysis



# Pathways Analysis

| Criteria | Paper-and-Pencil Proof | Simulation | Model Checking | Higher-order-logic Proof Assistants |
|---|---|---|---|---|
| Expressiveness | ☑ | ☑ | ☒ | ☑ |
| Accuracy | ☑ ? | ☒ | ☑ | ☑ |
| Automation | ☒ | ☑ | ☑ | ☒ |

# Contribution of this paper

- **Deduction System**
  - To reason about the reaction outcomes
- **Reaction Kinetics**
  - Concentration of molecules at any time
  - Modeled by Differential Equations

$$\frac{dn(t)}{dt} = \lambda * n(t)$$

- **Stochastics**
  - Probability that a molecule decays within a small time interval

$$p_1 = \lambda \Delta t$$

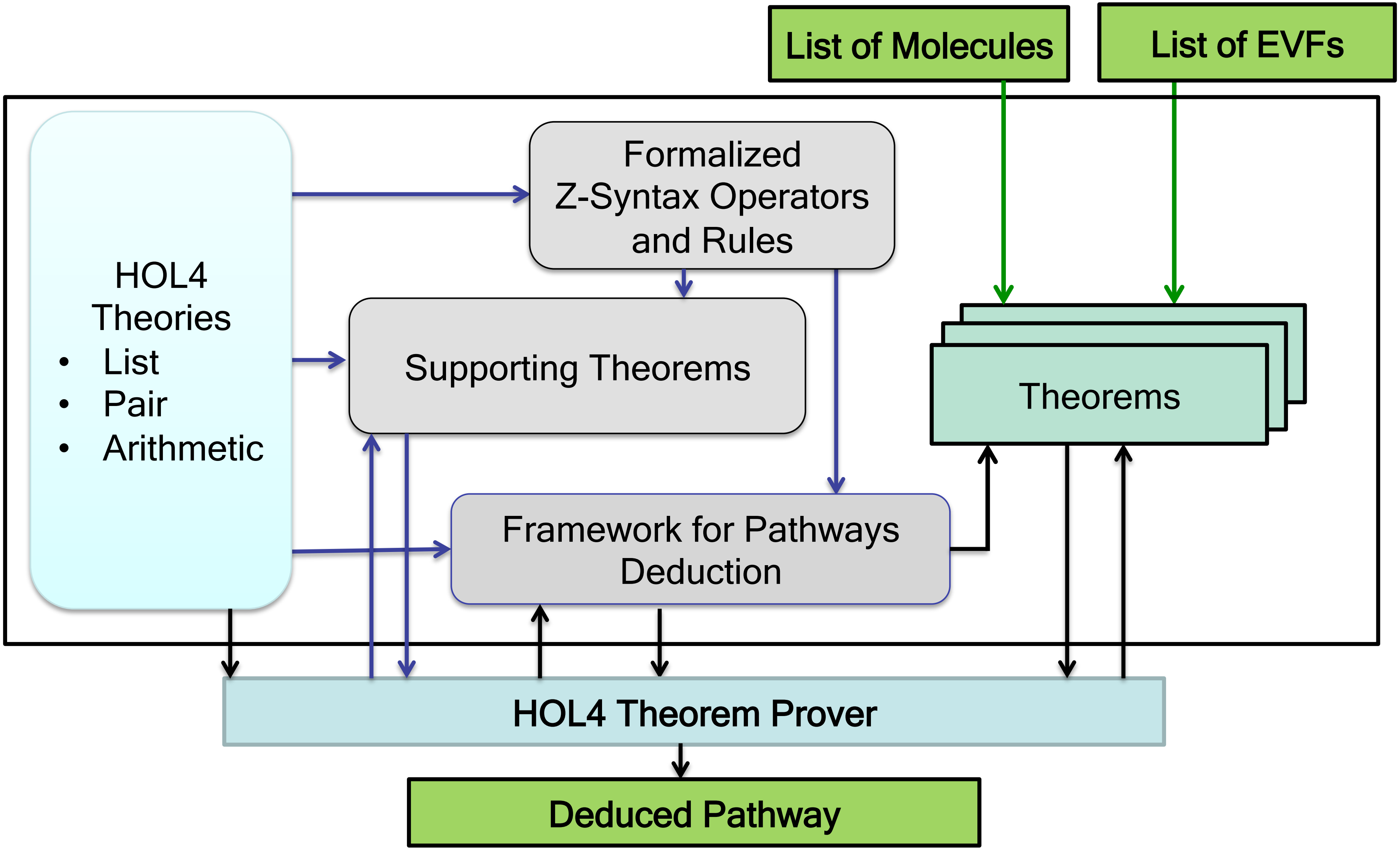
# HOL4 Theorem Prover

- Higher-order-logic Proof Assistant
  - Notation: ML
  - Small Core:
    - 5 basic axioms
    - 8 primitive inference rules
- Numerous automatic proof procedures are available
- Supports Reasoning about
  - Real Numbers
  - Calculus
  - Probability

# Molecular Pathways in HOL4

- Z-Syntax provides formal semantics to represent molecular pathways
  - Computer Implementable
  - Based on 3 operators and 4 inference rules

# Proposed Approach for Analysis of Molecular Pathways



# Z-Syntax Operators in HOL4

- Z-Interaction: (*)
  - Represents reaction between molecules
  - Single List represents interacting molecules
- Z-Conjunction: (&)
  - Represents non-reacting molecules
  - List of Lists represents a bag of molecules
- Z-Conditional: ( $\rightarrow$ )
  - Represents the occurrence of a net forward reaction under a given condition
    - Used to express *Empirically Valid Formulae* (EVFs)

# Formalizing Z-Syntax in HOL4

- Consumption of Molecules.
  - Reactants are always consumed in a reaction
  - Deleting a list element shows molecule consumption

- Elimination of Z-Conjunction allows deducing a single molecule from a bag (list) of molecules

# Pathway Deduction

- Automatic deduction of Pathways
  - Based on Z-Syntax Definitions
  - Recursively checks for reaction possibilities between all the available molecules and deduces new molecules based on given EVFs.
- Based on 5 major functions

# Case Studies

- Pathway Leading to TP53 Degradation
  - TP53, a tumor suppressing gene
    - regulates cell division, inactivated by MDM2 in normal cells
  - Repairs the cell or stops the damage from spreading

$$TP53 \ \& \ TP53 \ \& \ MDM2 \ \& \ U \ \& \ P \vdash d(TP53)$$

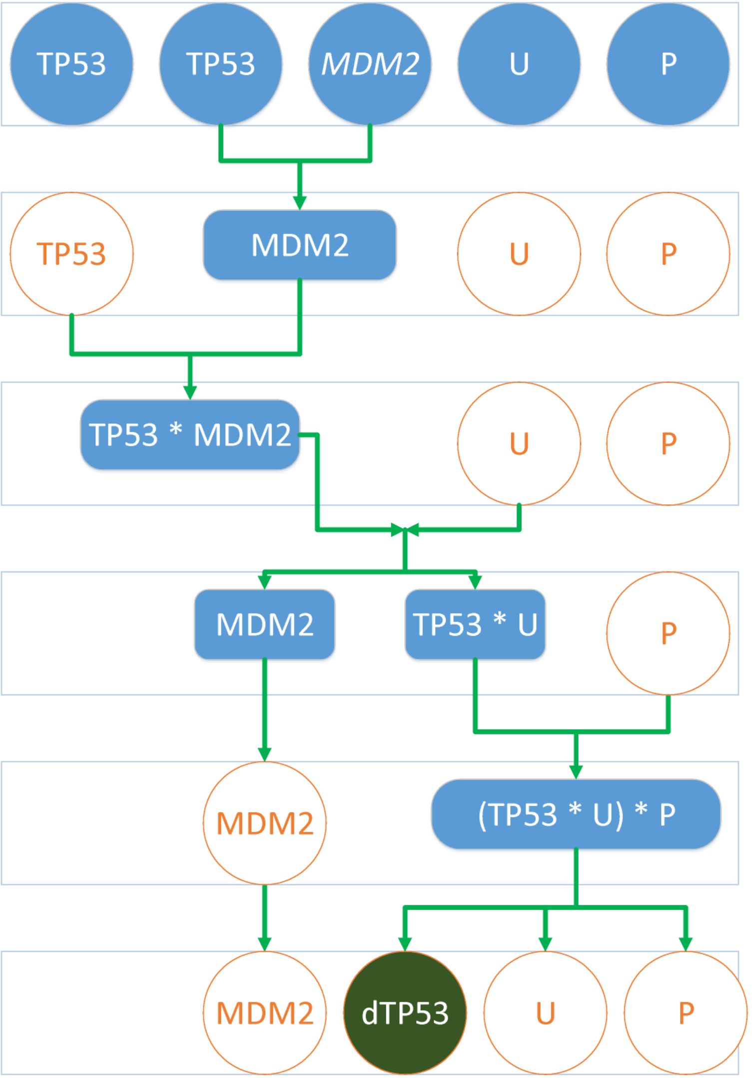
- F1,6P formation
  - Compound formed during Glycolysis
    - Research being done on its usefulness in Iron Chelation

$$Glc \ \& \ HK \ \& \ GPI \ \& \ PFK \ \& \ ATP \ \& \ ATP \vdash F1,6P$$

# TP53 Degradation

## Theorem : TP53 Degradation

```
⊢ DISTINCT [TP53;dTP53;P;U;iMDM2;MDM2] ⟹
z_conj_elim (z_deduction
[[TP53];[TP53];[iMDM2];[U];[P]]
[([TP53;MDM2],[[MDM2]]);
 ([MDM2;TP53],[[TP53;MDM2]]);
 ([TP53;MDM2;U],[[TP53;U];[MDM2]]);
 ([TP53;U;P],[[dTP53];[U];[P]])]) [dTP53]
                                   =[[dTP53]]
```




## Coding Statistics:

- ❑ Z-Syntax & Framework - 1000 lines
- ❑ Deducing TP53 Pathway - <10 lines
  - ❑ Automatically done using a TACTIC developed as part of this work

# Conclusions

- Formalization of Z-Syntax in HOL4
- Framework for Pathways deduction

- Advantages
  - Accurate Results
  - Reduction in user-effort while formally analyzing the molecular pathways
  - Helpful in discovery of different pathways

- Case Studies: Pathway analysis of TP53, F1,6P

# Future Directions

- Implementation of an interactive GUI
- Formal Verification of Deduction Framework
  - Generic Properties
- Include Reaction Kinetics, Stochastics and Diffusion

# Thanks!

- Questions

www.save.seecs.nust.edu.pk