1 Kinetic constants estimation in biochemical networks

1.1 Problem definition

Definition 1 (Biochemical network)**.** A biochemical network G is a tuple (S, R, E, ν) where

- $S = U \cup X \cup Y$ is the set of **species** of the biochemical network where
  - U is the set of input species of the network
  - X is the set of internal species
  - Y is the set of output species of the network
  - U, X, Y are **non-disjoint** sets
- R is the set of **reactions** in the biochemical network
  - $R_{\text{reversible}} \subseteq R$ is the subset of reversible reactions
- $E \subseteq S \times R \times T$ is the set of **relationships** between the species and reactions, with T as the set of relationship types
- $\nu = \left(\nu_{\text{reactant}}, \nu_{\text{product}}\right)$
  - $\nu_t : E_t \to \mathbb{N}_1$ with $t \in \{\text{reactant}, \text{product}\}$ are the stoichiometry functions for the reactants and products of the reactions

with

- $T = \{\text{product}, \text{reactant}, \text{enzyme}, \text{activator}, \text{inhibitor}\}$
- $E_t = \{(s, r) \mid (s, r, t) \in E\}$ is the projection of E over $t \in T$
- $E_{\text{modifier}} = E_{\text{enzyme}} \cup E_{\text{inhibitor}}$

$$S \to P \tag{1}$$

$$\dot{v} = \frac{k_2 \cdot E \cdot S}{S + \frac{k_{-1}+k_2}{k_1}} \tag{2}$$

$$\text{sia } S_1 + S_2 \to P \text{ una reazione con enzima } E \tag{3}$$

$$\dot{v} = \frac{k_2 \cdot [E] \cdot [S_1 \cdot S_2]}{[S_1 \cdot S_2] + \frac{k_{-1}+k_2}{k_1}} \tag{4}$$

dove $[S]$ è la concentrazione della species S

```
MATCH (r:ReactionLikeEvent)
WHERE EXISTS {
    MATCH (r)--(c1:CatalystActivity), (r)--(c2:CatalystActivity)
    WHERE c1 <> c2
}
RETURN COUNT (DISTINCT r) // 17
```

```
MATCH path = ({dbId: 983702})--(:CatalystActivity)
RETURN path
```

Let $\mathbb{N}_1 = \mathbb{N} \setminus \{0\}$

One enzyme molecule can catalyze thousands of reactions per second (this so-called turn over number ranges from 102 to 107 s 1). Enzyme cataly sis leads to a rate acceleration of about 106 up to 1012-fold compared to the noncatalyzed, spontaneous reaction.

The turnover number (kcat) of an enzyme is the maximum number of substrate molecules one enzyme active site can convert into product per second, indicating its catalytic efficiency, calculated as Vmax (maximum reaction velocity) divided by the total enzyme concentration. It shows how fast an enzyme works, with values ranging from less than one to millions of molecules per second, and is key to understanding how effectively an enzyme processes substrates

10^2 to $10^7 \frac{\text{reactions}}{s}$

Chemical and biochemical kinetics rely on the assumption that the reaction rate v at a certain point in time and space can be expressed as a unique function of the concentrations of all substances at this point in time and space.

Classical enzyme kinetics assumes for sake of simplicity a spatial homogeneity (the "well-stirred" test tube) and no direct dependency of the rate on time:

time-invariant system

concentration: always based on a volume

- $\frac{n}{V}$ $\frac{\text{quantity}}{\text{litre}}$

1.2 Assumptions

1.2.1 Assumption 1 (quasi-equilibrium of E and ES)

Michaelis and Menten [5] considered a quasi-equilibrium between the free enzyme and the enzyme–substrate complex, meaning that the reversible conversion of E and S to ES is much faster than the decomposition of ES into E and P, or in terms of the kinetic constants, that is,

$k_1, k_{-1} \gg k_2$

- (possibile assunzione "quasi-equilibrium" di enzima-libero e complesso)

enzima-specie "la reazione regolata da un enzima, che produce il complesso "enzima - specie" è molto più lenta della reazione che prende il complesso e ci fa cose!"

1.2.2 Assumption 2 (concentration of ES is constant)

This works only if $S(t=0) \gg E$ (because the turnover rate of E is kinda big, thus $\frac{\mathrm{dES}}{\mathrm{d}t} = 0$

Given a biochemical network $G = (S, R, , \nu)$ let Equation 5 be the general form of a modular rate law that describes the kinetics of reaction $r \in R$.

1.2.3 Stuff

1. Draw a wiring diagram of all steps to consider (e.g., Eq. (4.11)). It contains all substrates and products (S and P) and n free or bound enzyme species

(E and ES).

2. The right sides of the ODEs for the concentrations changes sum up the rates of all steps leading to or away from a certain substance (e.g., Eqs. (4.12)–(4.15)). The rates follow mass action kinetics (Eq. (4.3)).
3. The sum of all enzyme-containing species is equal to the total enzyme concentration E_{total} (the right side of all differential equations for enzyme species sums up to zero). This constitutes one equation (ok, but what if enzymes are produced?)
4. The assumption of quasi-steady state for $n - 1$ enzyme species (i.e., setting the right sides of the respective ODEs equal to zero) together with (3) result in n algebraic equations for the concentrations of the n enzyme species.
5. The reaction rate is equal to the rate of product formation (e.g., Eq. (4.16)). Insert the respective concen trations of enzyme species resulting from (4).

$$v_r = E_{\text{total}} \cdot f_{\text{reg}} \cdot \frac{T}{D + D_{\text{reg}}} \tag{5}$$

where

$$E_{\text{total}} = \sum_{\substack{(s,r) \\ \in \\ E_{\text{enzyme}}}} [s] + \text{complexes that contain the enzyme??} \tag{6}$$

$$f_{\text{reg}} = \prod_{(s,r) \in E_{\text{enzyme}}} \frac{[s]^{n_r^s}}{K_r^s + [s]^{n_r^s}} \prod_{(s,r) \in E_{\text{inhibitor}}} \frac{K_r^s}{K_r^s + [s]^{n_r^s}} \tag{7}$$

$$T = k_{\text{cat}}^{\text{for}} \prod_{\substack{(s,r) \\ \in \\ E_{\text{reactant}}}} \left(\frac{[s]}{K_{m,s}^r} \right)^{\nu_{\text{reactant}}(s,r)} - k_{\text{cat}}^{\text{back}} \prod_{\substack{(s,r) \\ \in \\ E_{\text{product}}}} \left(\frac{[s]}{K_{m,s}^r} \right)^{\nu_{\text{product}}(s,r)} \tag{8}$$

Power-law modular rate law: D = 1 (such as mass action kinetics)

$$D_1 = 1 \tag{9}$$

Common modular rate law

$$\begin{aligned} D_2 = & \prod_{\substack{(s,r) \\ \in \\ E_{\text{reactant}}}} \left(1 + \sum_{n=1}^{\nu_{\text{reactant}}(s,r)} \left(\frac{[s]}{K_{m,s}} \right)^n \right) \\ & + \prod_{\substack{(s,r) \\ \in \\ E_{\text{product}}}} \left(1 + \sum_{n=1}^{\nu_{\text{product}}(s,r)} \left(\frac{[s]}{K_{m,s}} \right)^n \right) \end{aligned} \tag{10}$$

Simultaneous binding modular rate law

$$D_3 = \prod_{\substack{(s,r) \\ \in \\ E_{\text{reactant}}}} \left(1 + \frac{[s]}{K_{m,s}} \right)^{\nu_{\text{reactant}}(s,r)} \prod_{\substack{(s,r) \\ \in \\ E_{\text{product}}}} \left(1 + \frac{[s]}{K_{m,s}} \right)^{\nu_{\text{product}}(s,r)} \tag{11}$$

Direct binding modular rate law:

$$D_4 = 1 + \prod_{(s,r) \in E_{\text{reactant}}} \left(\frac{[s]}{K_{m,s}^r} \right)^{\nu_{\text{reactant}}(s,r)} + \prod_{(s,r) \in E_{\text{product}}} \left(\frac{[s]}{K_{m,s}^r} \right)^{\nu_{\text{product}}(s,r)} \tag{12}$$

Force-dependent modular rate law:

$$D_5 = \sqrt{\prod_{\substack{(s,r) \\ \in \\ E_{\text{reactant}}}} \left(1 + \frac{[s]}{K_{m,s}}\right)^{\nu_{\text{reactant}}(s,r)} \prod_{\substack{(s,r) \\ \in \\ E_{\text{product}}}} \left(1 + \frac{[s]}{K_{m,s}}\right)^{\nu_{\text{product}}(s,r)}} \tag{13}$$

Definition 2 (Dynamic biological model)**.** Given a biochemical network $G = (S, R, E, \nu)$ let $B = (G', \mathcal{K})$ be the biological model derived from G with added modular law kinetics, with $G' = (S, R', E', \nu')$ where

- $R' = R \cup R_{\text{input}} \cup R_{\text{output}}$ with
  - $R_{\text{input}} = \{r_s \mid s \in U - Y\}$
  - $R_{\text{output}} = \{r_s \mid s \in Y - U\}$
- $E' = E \cup \{(s, r, \text{product}) \mid r_s \in R_{\text{input}}\} \cup \{(s, r, \text{reactant}) \mid r_s \in R_{\text{output}}\}$
- $\nu' = \left(\nu'_{\text{reactant}}, \nu'_{\text{product}}\right)$
  - $$\nu'_{\text{reactant}(s,r)} = \begin{cases} 1 & \text{if } (s,r) \in R_{\text{input}} \\ \nu_{\text{reactant}}(s,r) & \text{otherwise} \end{cases} \tag{14}$$
  - $$\nu'_{\text{product}(s,r)} = \begin{cases} 1 & \text{if } (s,r) \in R_{\text{output}} \\ \nu_{\text{product}}(s,r) & \text{otherwise} \end{cases} \tag{15}$$

Then $\mathcal{K}$, the set of constants, can be defined on G' as

$$\begin{aligned} \mathcal{K} = &\left\{k^{\text{for}}_{\text{cat},r} \mid r \in R'\right\} \cup \\ &\left\{k^{\text{back}}_{\text{cat},r} \mid r \in R'\right\} \cup \\ &\left\{K^r_{m,s} \mid (s,r) \in E'_{\text{reactant}} \cup E'_{\text{product}}\right\} \cup \\ &\left\{K^s_r \mid r \in R' \wedge (s,r) \in E_{\text{modifier}}\right\} \cup \\ &\left\{n^s_r \mid r \in R' \wedge (s,r) \in E_{\text{modifier}}\right\} \end{aligned} \tag{16}$$

where

- k_r is the kinetic constant of reaction r
- K^s_r is the apparent dissociation constant of modifier s in reaction r
- n^r_s the hill coefficient of modifier s in reaction r

1.3 Convenience Kinetics and Modular Rate Laws

$$v = E_{\text{total}} \cdot f_{\text{reg}} \cdot \frac{k_{\text{cat}}^{\text{for}} \prod_i \left(\frac{S_i}{K_{m,S_i}}\right)^{n_{-i}} - k_{\text{cat}}^{\text{back}} \prod_j \left(\frac{P_j}{K_{m,P_j}}\right)^{n_j}}{\prod_i \left(1 + \sum_{a=1}^{i} \left(\frac{S_i}{K_{m,S_i}}\right)^a\right) + \prod_j \left(1 + \sum_{a=1}^{j} \left(\frac{P_j}{K_{m,P_j}}\right)^a\right) - 1}$$

$\boldsymbol{v}$ amount of substance that is converted in the reaction

$\boldsymbol{E_{\text{total}}}$? oh, wait, wtf????? Is it computable? E_{total} is the enzyme concentration

$\boldsymbol{f_{\text{reg}}}$?

$\boldsymbol{k_{\text{cat}}^{\text{for}}}$ constant of reaction moving "forward" when the reaction is reversible (why cat?)

$\boldsymbol{k_{\text{cat}}^{\text{back}}}$ turnover rate!!! (same for forward)

$\boldsymbol{K_{m,S_i}, K_{m,P_j}}$ constant which somehow reduces the probability of reaction of that species, what does that m stand for?

$$f_{\text{reg}} = \begin{cases} 1 & \text{if no regulation is present} \\ \prod\left(\frac{M}{K_A+M} \cdot \frac{K_I}{K_I+M}\right) & \text{otherwise (resp. positive and negative regulation)} \end{cases}$$

where

- M is the concentration of the modifier
- K_A, K_I measured in concentration units (values denote concentrations, at which the inhibitor or activator has its half-maximal effect)

now:

- the denominator should somehow "slow down" the reaction… right?
  - well, in the worst case the denominator is exactly 1
- is it $>$ or $<$ of 1? It must be at least 1
- what is the domain of $K_{m,S}$? Is it $>= 0$? Yeah, it must be.
- Where are the modifiers? Are the modifiers in E_{total}?

$$K_V = \left(K_{\text{cat}}^{\text{for}} \cdot K_{\text{cat}}^{\text{back}}\right)^{\frac{1}{2}}$$

$$E_{\text{total}}$$

pare pericolosa, perché è la somma della concentrazione degli enzimi + la somma del prodotto

Definition 3 (Biological model satisfiability problem)**.** Given a biological model $B = (G, \mathcal{K})$ and a let $\mathcal{S}$ be the set of concentrations of species of the species in the network, $\mathcal{S} = \{s \mid s \in S\}$ and $S_{\text{avg}} = \{s_{\text{avg}} \mid s \in S\}$ the set of average concentrations of the species, $T \in \mathbb{R}^+$ the time horizon, $\varphi \in [0, 1]$ the following constraints must hold:

$$\forall s \in \mathcal{S} \quad 0 \leq [s] \leq 1 \tag{17}$$

$$\begin{aligned} &\forall k_{r_1}, k_{r_2} \in \mathcal{K}, s \in S \\ &\quad (s, r_1, i) \in E \wedge (s, r_2) \in (E_{m^+} \cup E_{m^-}) \wedge r_1 \neq r_2 \rightarrow k_{r_1} < k_{r_2} \end{aligned} \tag{18}$$

$$\forall s \in \mathcal{S}_{\text{avg}} \quad s(\varphi \cdot T) - s(T) = 0 \tag{19}$$

Definition 4 (Optimization problem)**.**

2 Queries

2.1 Multiple compartments per species

Because the functions of biologic molecules critically depend on their subcellular locations, chemically identical entities located in different compartments are represented as distinct physical entities. (TODO: this is a citation)

```
MATCH (physicalEntity:PhysicalEntity)
WHERE EXISTS {
    MATCH
        (physicalEntity)--(compartment1:Compartment),
        (physicalEntity)--(compartment2:Compartment)
    WHERE
        compartment1 <> compartment2
}
RETURN COUNT(DISTINCT physicalEntity) // 6347
```

Listing 1: Entities connected to multiple compartments

```
MATCH (physicalEntity:PhysicalEntity)
WHERE EXISTS {
    MATCH
        (physicalEntity)-[:compartment]-(compartment1:Compartment),
        (physicalEntity)-[:compartment]-(compartment2:Compartment)
    WHERE
        compartment1 <> compartment2
}
RETURN COUNT(DISTINCT physicalEntity) // 656
```

Listing 2: Entities with multiple compartments only using the `[:compartment]` relation

```
MATCH (physicalEntity:PhysicalEntity)
WHERE EXISTS {
    MATCH
        (physicalEntity)-[:compartment]-(compartment1:Compartment),
        (physicalEntity)-[:compartment]-(compartment2:Compartment)
    WHERE
        compartment1 <> compartment2
        AND NOT EXISTS {
            MATCH (compartment1)-[:surroundedBy]-(compartment2)
        }
}
RETURN COUNT(DISTINCT physicalEntity) // 491
```

Listing 3: Entities with multiple compartments where one compartment is not `[:surroundedBy]` the other

```
MATCH (physicalEntity:PhysicalEntity)
WHERE
    EXISTS {
        MATCH
            (physicalEntity)-[:compartment]-(compartment1:Compartment),
            (physicalEntity)-[:compartment]-(compartment2:Compartment)
        WHERE
            compartment1 <> compartment2
            AND NOT EXISTS {
                MATCH (compartment1)-[:surroundedBy]-(compartment2)
            }
    }
    AND "EntitySet" IN labels(physicalEntity)
RETURN COUNT(DISTINCT physicalEntity) // 491
```

Listing 4: Entities with multiple unrelated compartments which are entity-sets (how do I treat these?)

2.2 Reactions with multiple compartments

```
MATCH (reaction:ReactionLikeEvent)
WHERE EXISTS {
    MATCH
        (reaction)--(compartment1:Compartment),
        (reaction)--(compartment2:Compartment)
    WHERE
        compartment1 <> compartment2
}
RETURN COUNT (DISTINCT reaction) // 36322 (out of 93672)
```

Listing 5:

```
MATCH (reaction:ReactionLikeEvent)
WHERE EXISTS {
    MATCH
        (reaction)--(compartment1:Compartment),
        (reaction)--(compartment2:Compartment)
    WHERE
        compartment1 <> compartment2
        AND NOT EXISTS {
            MATCH (compartment1)-[:surroundedBy]-(compartment2)
        }
}
RETURN COUNT (DISTINCT reaction) // 20974
```

Listing 6:

```
MATCH (reaction:ReactionLikeEvent)
WHERE EXISTS {
    MATCH
        (reaction)--(compartment1:Compartment),
        (reaction)--(compartment2:Compartment)
    WHERE
        compartment1 <> compartment2
        AND NOT EXISTS {
            MATCH (compartment1)--(compartment2)
        }
}
RETURN COUNT (DISTINCT reaction) 20833 (if you remove :surroundedBy)
```

Listing 7: https://reactome.org/PathwayBrowser/#/R-HSA-9646399&FLG=R-HSA-9640114&FLGINT

https://sbml.org/documents/faq/

https://raw.githubusercontent.com/combine-org/combine-specifications/main/specifications/files/sbml.level-3.version-2.core.release-2.pdf

```
MATCH (reaction:ReactionLikeEvent)
WHERE EXISTS {
    MATCH
        (reaction)--(compartment1:Compartment),
        (reaction)--(compartment2:Compartment)
    WHERE
        compartment1 <> compartment2
        AND NOT EXISTS {
            MATCH (compartment1)-[:surroundedBy]-(compartment2)
        }
        AND EXISTS {
            MATCH (compartment1)--(compartment2)
        }
}
RETURN COUNT (DISTINCT reaction) // 189 "componentOf"
```

```
MATCH (:Compartment)-[relation]-(:Compartment)
RETURN DISTINCT type(relation)
```

2.3 Species which are both inputs and outputs of a reaction

```
MATCH (reaction:ReactionLikeEvent)-[:input]->(p:PhysicalEntity)
WHERE EXISTS {
    MATCH (n)-[:output]->(p)
}
RETURN COUNT(n) // 1062
```

```
MATCH (n:ReactionLikeEvent)-[:input]->(p:PhysicalEntity)
WHERE EXISTS {
    MATCH (n)-[:output]->(p), (m)-[:output]->(p)
    WHERE n <> m
}
RETURN COUNT (n) // 895
```

```
MATCH (n:ReactionLikeEvent)-[:input]->(p:PhysicalEntity)
WHERE EXISTS {
    MATCH (n)-[:output]->(p), (n)-[:output]->(q)
    WHERE p <> q
}
RETURN COUNT (n) // 741
```

```
MATCH (n:ReactionLikeEvent)-[r1:input]->(p:PhysicalEntity)
WHERE EXISTS {
    MATCH (n)-[r2:output]->(p), (n)-[:output]->(q)
    WHERE
        p <> q
        AND r1.stoichiometry <> r2.stoichiometry
}
RETURN COUNT (n) // 36
```

Di queste 36, 24 sono “electron transfer”

```
MATCH (n:ReactionLikeEvent)-[r1:input]->(p:PhysicalEntity)
WHERE EXISTS {
    MATCH (n)-[r2:output]->(p), (n)-[:output]->(q)
    WHERE
        p <> q
        AND r1.stoichiometry <> r2.stoichiometry
}
AND NOT n.displayName STARTS WITH 'Electron transfer'
RETURN COUNT (n) // 12
```

2.4 Enzymes which are both positive and negative regulators

Reactions that are driven by an enzyme are described as requiring a catalyst activity, modeled in Reactome by linking the macromolecule that provides the activity to the GO molecular function term [10,11] that describes the activity. In addition, the Reactome data model allows reactions to be modulated by positive and negative regulatory factors. When a precise regulatory mechanism (‘positive allosteric regulation’, ‘noncompetitive inhibition’) is known, this information is captured.

```
MATCH (reaction:ReactionLikeEvent)
WHERE EXISTS {
    MATCH
        (reaction)-[:catalystActivity]->
            (:CatalystActivity)-[:physicalEntity]->(enzyme),
        (reaction)-[:regulatedBy]->
            (:NegativeRegulation)-[:regulator]->(inhibitor)
    WHERE enzyme = inhibitor
}
RETURN COUNT(DISTINCT reaction) // 4
```

```
MATCH (reaction:ReactionLikeEvent)
WHERE EXISTS {
    MATCH
        (reaction)-[:regulatedBy]->
            (:PositiveRegulation)-[:regulator]->(enzyme),
        (reaction)-[:regulatedBy]->
            (:NegativeRegulation)-[:regulator]->(inhibitor)
    WHERE enzyme = inhibitor
}
RETURN COUNT(DISTINCT reaction) // 4
```

```
MATCH
    (reaction:ReactionLikeEvent),
    path1 = (reaction)-[:catalystActivity]->
            (:CatalystActivity)-[:physicalEntity]->(),
    path2 = (reaction)-[:regulatedBy]->
        (:NegativeRegulation)-[:regulator]->()
WHERE
    reaction.dbId IN [164341, 164087, 164136, 164055]
RETURN path1, path2
UNION
MATCH
    (reaction:ReactionLikeEvent),
    path1 = (reaction)-[:regulatedBy]->
        (:PositiveRegulation)-[:regulator]->(),
    path2 = (reaction)-[:regulatedBy]->
        (:NegativeRegulation)-[:regulator]->()
WHERE
    reaction.dbId IN [8950347, 10181912, 10290972, 10761215]
RETURN path1, path2
```

- top ones are all in either *"Mus musculus"* or in *"Rattus norvegicus"*
- bottom ones are basically the same reaction, but inferred to multiple species

```
MATCH
    path = (r1 {dbId: 1482894})-[:reverseReaction]-(r2),
    path2 = (r1)--(:PhysicalEntity),
    path3 = (r2)--(:PhysicalEntity)
RETURN path, path2, path3
```