Equivalence relations

between molecules, including proteins

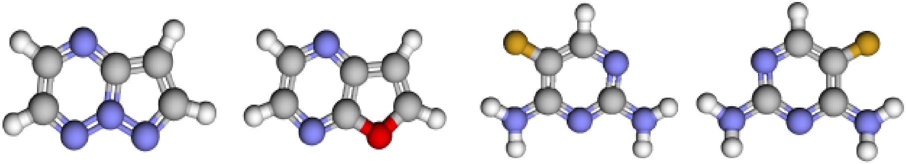
open-table discussion with new students
and colleagues in applied areas

THE ROYAL SOCIETY

Key question: same or different?

Data: clouds of (un)ordered points representing atoms in molecules, including proteins.

different representations may refer to the *same* object but what do we really mean by **"same"?**



Sacchi et al. **Same or different - that is the question**: identification of crystal forms from crystal structure data. CrystEngComm, 2020.

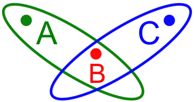
Three axioms of an equivalence

A relation $A \sim B$ between any data objects is called an *equivalence* if the three axioms hold:

(1) *reflexivity*: any object $A \sim A$;

(2) *symmetry*: if $A \sim B$ then $B \sim A$;

(3) *transitivity*: if $A \sim B$ and $B \sim C$, then $A \sim C$.



The transitivity axiom guarantees that all objects are in disjoint classes. Any justified classification needs an equivalence.

Equality is an equivalence: $0.5 = 50\% = \frac{1}{2} = 2 \div 4$

Examples or non-examples?

Question. Are the following binary relations between real numbers $x, y \in \mathbb{R}$ equivalences?

(1) $x < y$ (strict); (2) $x \leq y$ (non-strict);

(3) distance $|x - y| \leq \varepsilon$ for any fixed $\varepsilon > 0$.

Answer. (1) fails the reflexivity: $x < x$ is false.

(2) fails the symmetry axiom: if $x \leq y$ then $y \leq x$ holds only for $x = y$, not for all $x, y \in \mathbb{R}$.

(3) fails the transitivity axiom: the Euclidean distance $|\pm\varepsilon - 0| = \varepsilon$, but $|-\varepsilon - \varepsilon| = 2\varepsilon > \varepsilon$.

Equivalence classes

For any fixed equivalence, all objects can be classified (split) into disjoint *classes* consisting of all objects that are equivalent to each other.

Any object A defines the equivalence class $[A] = \{\text{all objects } B \text{ equivalent to } A\}$.

Take any $C \notin [A]$ and form the class $[C] = \{\text{all objects } B \text{ equivalent to } C\}$ and so on.

If classes overlap: $B \in [A] \cap [C]$, they should coincide by the transitivity axiom: $A \sim B \sim C$.

Sorites paradox (of a heap of sand)

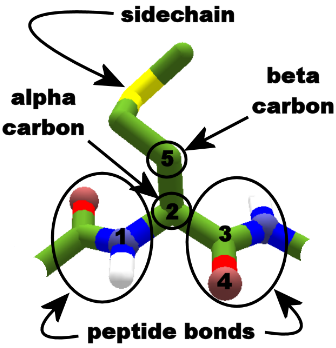
If a heap is reduced by a single grain at a time, when does it cease to be considered the [same] heap?

If $x \in \mathbb{R}$ (or any object given by real numbers) is considered *equivalent to* (the same as) $x \pm \varepsilon$ for any fixed $\varepsilon > 0$, all objects become equivalent by the transitivity axiom due to a long enough chain of equivalences $x \sim x_1 \sim \cdots \sim x_n \sim y$ in a connected space where objects are compared.

Paradox solution: any grain changes the heap.

The backbone of a protein chain

The *primary structure* of a *protein chain* is a sequence of amino acid residues whose side chains R_i are joined to α-carbon atoms A_i.



A *protein backbone* is a sequence of ordered triplets of the atoms (1) nitrogen N_i, (2) alpha-carbon A_i, (3) another carbon C_i embedded in space $\mathbb{R}^3$, where $i = 1, \ldots, m$ (# residues).

Weaker vs stronger equivalences

By length: backbones can be called the same if their lengths (number m of residues) are equal.

The equivalence by length is *weaker* than **by sequence** of amino acids because many different sequences have the same length.

If backbones $S, Q \subset \mathbb{R}^3$ coincide as ordered sets of atoms, this **coincidence equivalence** is **too strong** because we can rigidly move a protein and hence change atomic coordinates without changing its functional properties.

Different equivalence relations

Chemical: crystals $A \sim B$ if A, B have the same composition. Ok, but diamond and graphite with vastly different properties are in the same class.

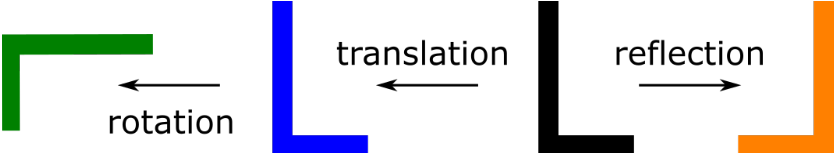
By property: compounds $A \sim B$ if A, B have the same property. Ok, but crystals that share one property can differ by many other properties.

By symmetry: $A \sim B$ if A, B have isomorphic space groups. Fedorov and Schoenflies (1891): 230 classes. Then NaCl, MgO, TiC, LaN, NaI, RbF, SrS, ... have the same group (225, Fm$\bar{3}$m).

What is the strongest relation?

Many real-life objects are rigid and should be considered equivalent under **rigid motion** =

a composition of translations and rotations ($\cong$);

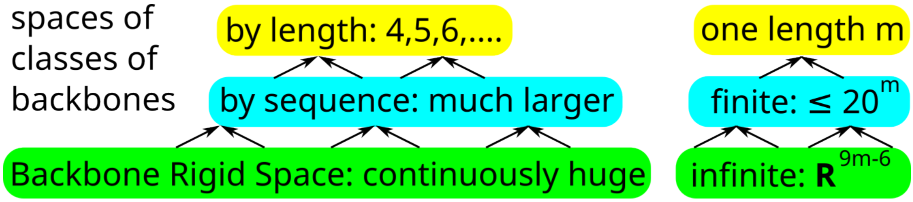


or *isometry* = rigid motion + reflections in $\mathbb{R}^n$.

In a general metric space, an **isometry** is any map that preserves all inter-point distances ($\simeq$).

Spaces of equivalence classes

All protein backbones form a finite space (of equivalence classes) by *length*, a much larger finite space by *primary structures* (sequences), and a huge infinite *Backbone Rigid Space* of all rigid classes: any noise changes a rigid class.



How can we distinguish between rigid classes?

Descriptors vs invariants

Real objects are often described by ambiguous *descriptors*, e.g. lists of x, y, z coordinates, that easily change under important equivalences. An **invariant** I is a function (property) whose values are preserved under a given equivalence.

If molecules $S \cong Q$ are exactly matched under rigid motion, then $I(S) = I(Q)$. Equivalently, if $I(S) \neq I(Q)$, then $S \not\cong Q$ are rigidly different.

The number m of atoms is invariant under rigid motion. A photo is a descriptor, not an invariant.