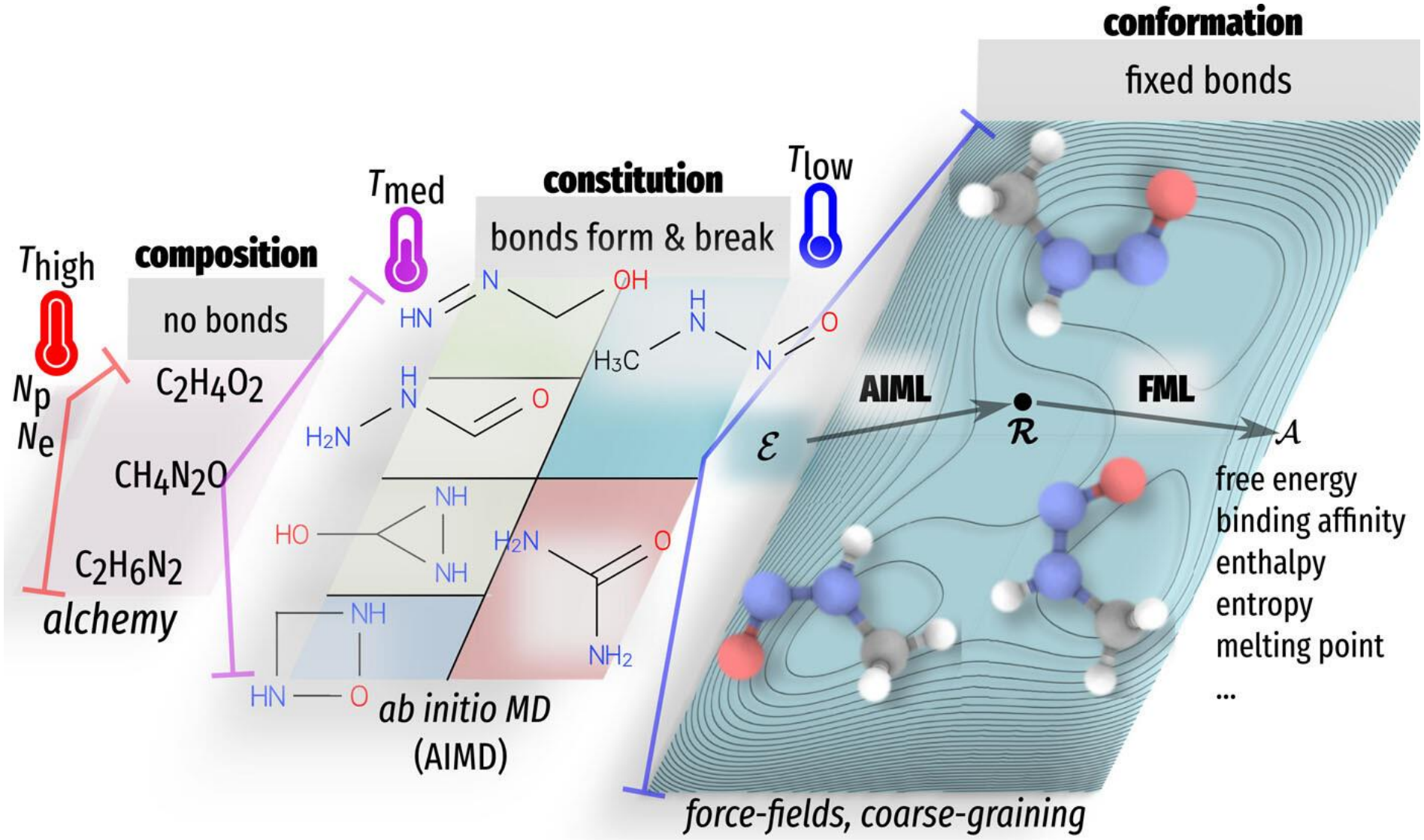
Chemical Space

Hierarchies



https://doi.org/10.1063/5.0095674

Scaling

Scaling with chemical diversity

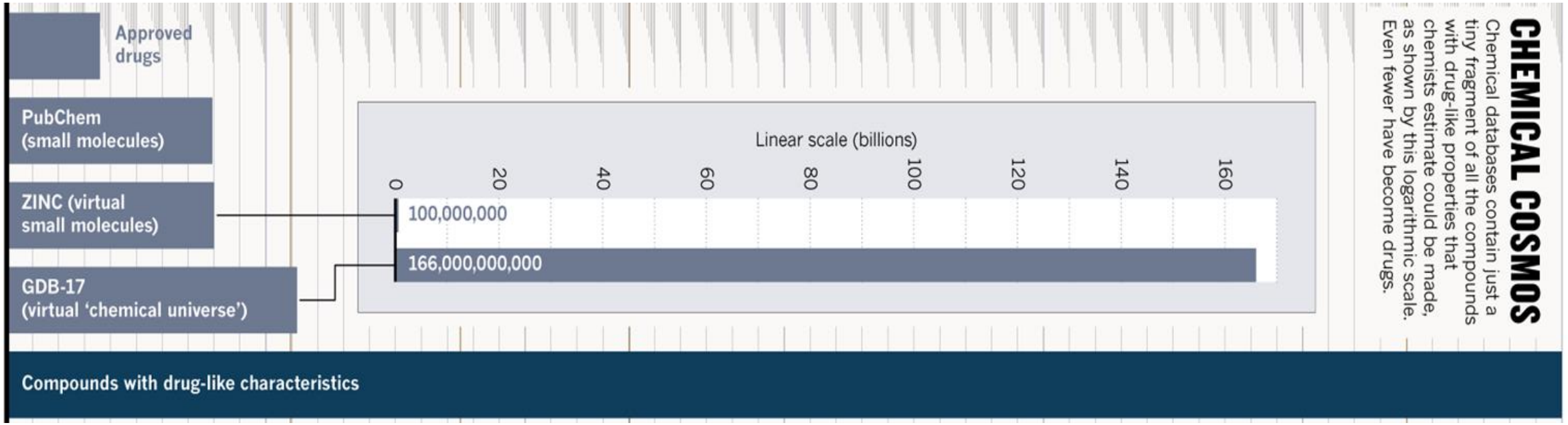
Elements	# atoms	# sum formulas	# graphs [1]	#confomers
CONF	5	169	4,715	16,797[2]
CONFS	5	349	9,917	51,710
CONFSP	5	757	31,550	
CONFSPCl	5	1,142	37,908	
CONFSPClBr	5	1,647	45,132	
CONFSPClBrI	5	2,291	53,285	328,591[2]

Scaling with number of heavy atoms

Elements	# atoms	# sum formulas	# graphs	#confomers [2]
CONF	1	4	4	
CONF	2	19	19	
CONF	3	49	94	
CONF	4	97	621	
CONF	5	169	4,715	
CONF	6	276	42,087	
CONF	7	425	417,923	7,039,390
CONFS	1	5	5	
CONFS	2	28	28	
CONFS	3	82	160	
CONFS	4	180	1,161	
CONFS	5	349	9,917	
CONFS	6	625	97,607	
CONFS	7	1,050	1,064,343	23,016,417

https://gist.github.com/ferchault/9ae958b637a255475b502fc63c652835

Chemistry at scale



Commercial databases

- 164 million molecules
- 15k added daily

Scale

- One person: 1 million compounds/second
- 10 billion people on earth
- 10^{26} universe ages to go through

https://doi.org/10.1038/549445a

Necessity vs Convenience

Necessity

- Only way to cover problem size
- Still open to systematic evaluation
- Often used as prefiltering step
- Complicated chemistry
- Tricky / error-prone reference calculations

Convenience

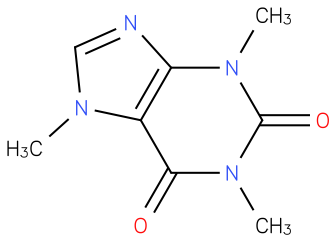
- Can be done more accurately
- Uneconomical/cumbersome reference method
- Often used as direct but optional substitute
- Standard energy calculations of well-behaved systems
- Semi-empirical level sufficient

Summary

1. Almost none of chemical space has been explored.
2. Scaling is a key aspect to think about when comparing methods.
3. Chemical diversity drives molecular diversity.

Potential Energy Surfaces

Concept



$$\hat{\mathrm{H}}(\mathbf{R}_I, Z_I, n, \sigma) \rightarrow \Psi \rightarrow \text{Energy} \quad E(\mathbf{R}_I | Z_I, n, \sigma)$$

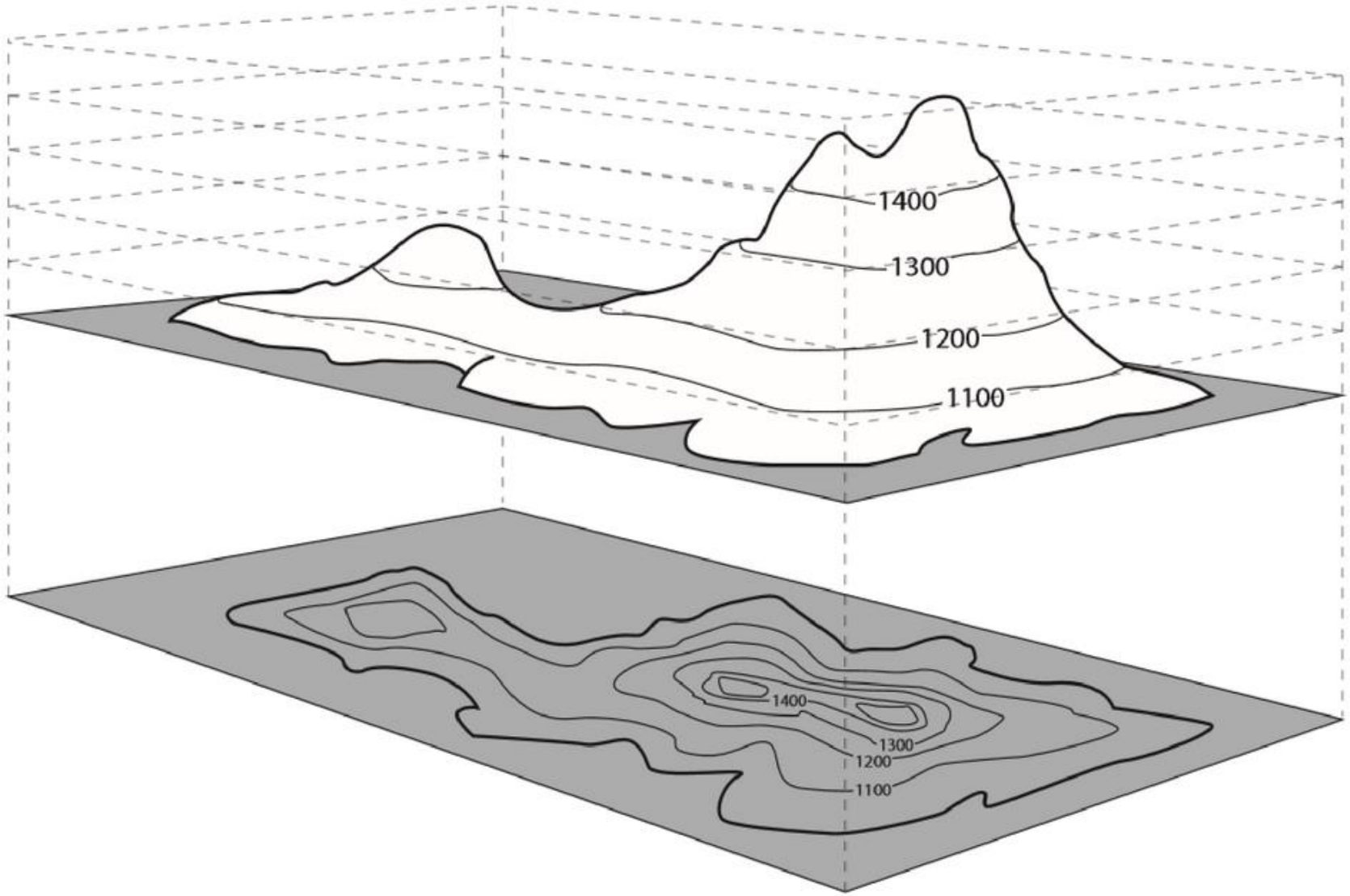
Numbers of electrons

Electronic spin

Nuclear charges

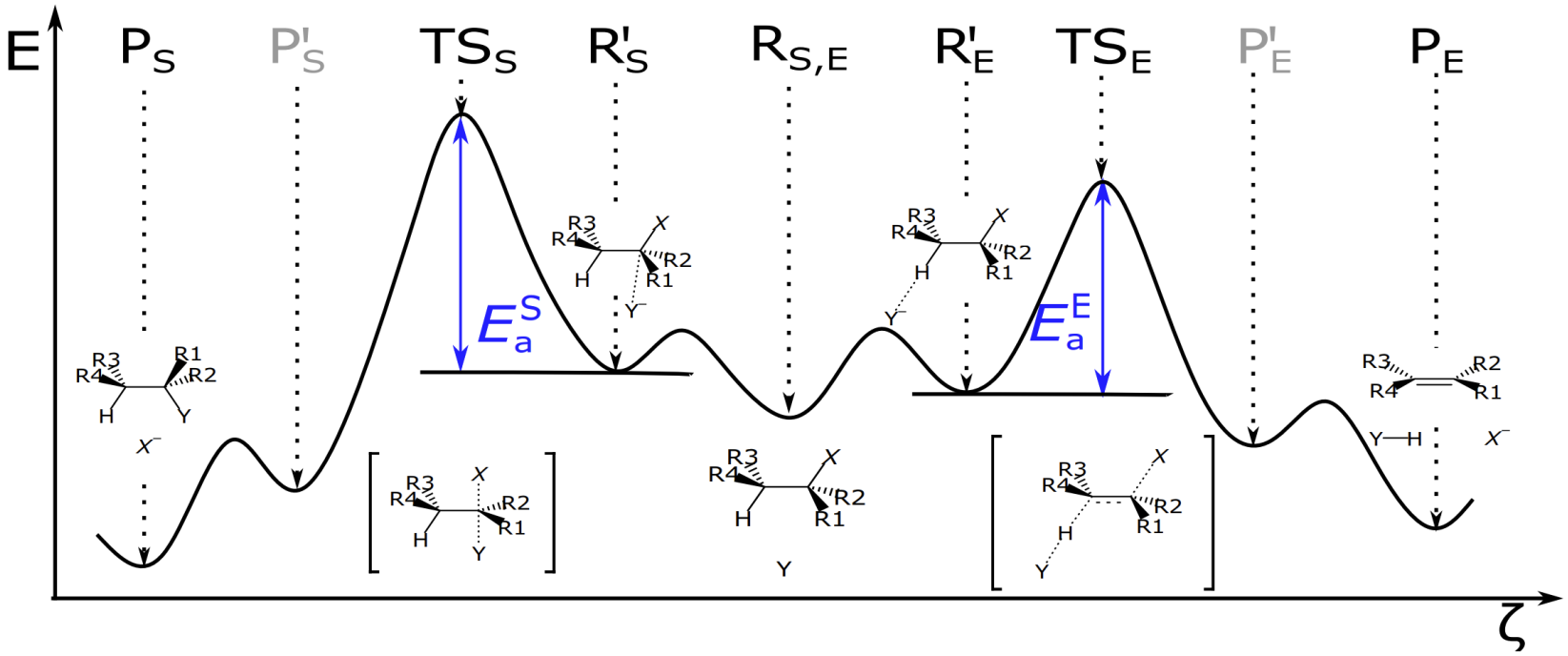
Nuclear coordinates

Topographical map



https://mapswithnoroads.com

Critical points



- Local minima — Meta-stable, ensembles
- Global minima — Most stable
- Saddle points — Barriers / access

- Attractive basins

https://doi.org/10.1088/2632-2153/aba822

Properties

Curvatures:
- Vibrational frequencies
- Normal modes
- Vibrational energy levels

Barriers:
- Tunneling
- Thermalisation

Thermally accessible regions:
- Ensemble of configurations