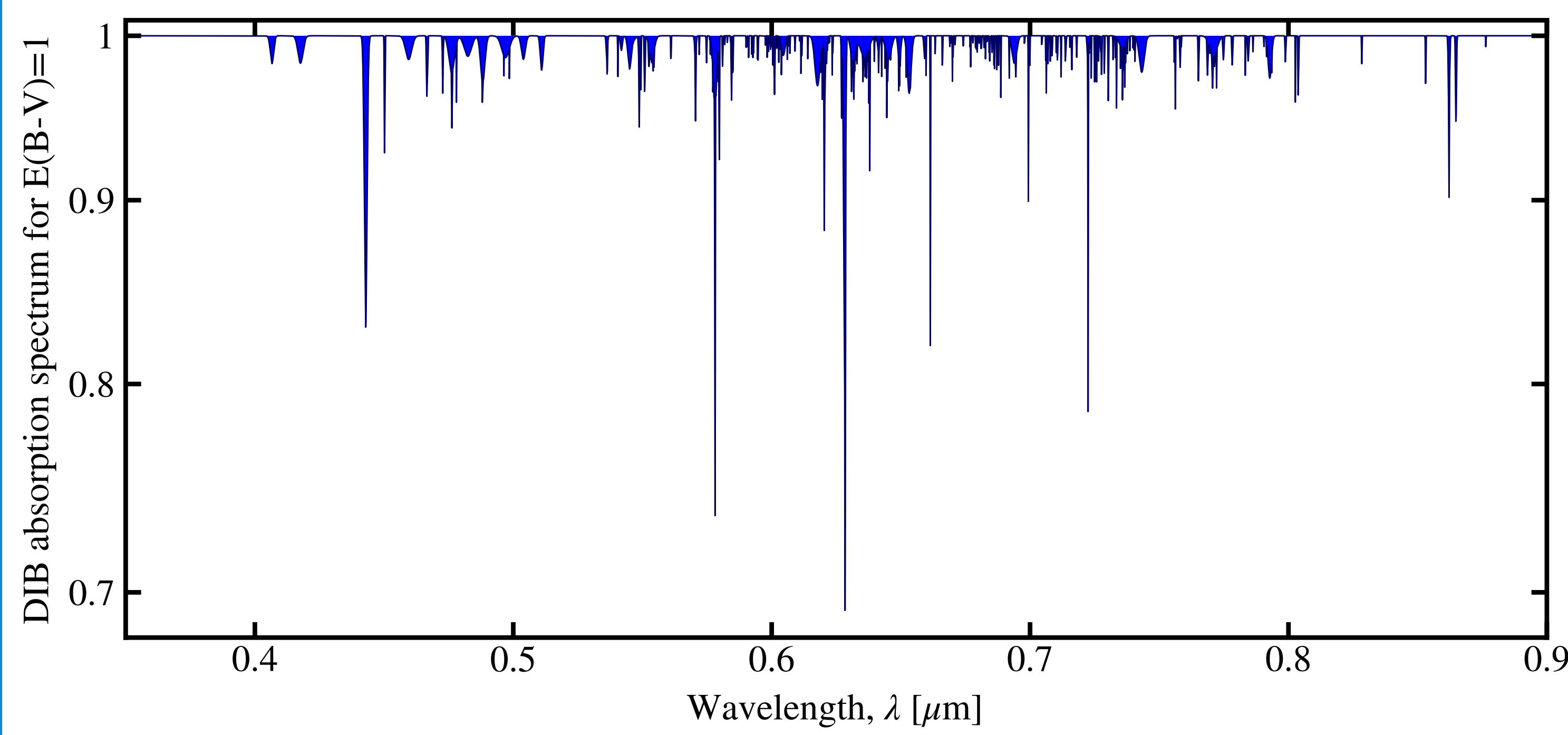


Could Elephant Sharks, Agnathans & Butterflies Be Relevant to ISM Studies?

Corentin VAN DEN BROEK D'OBRENAN^{1,2} Frédéric GALLIANO¹ Victor BOTEV² Jeremy MINTON² & Ronin WU²

¹Département d'Astrophysique, CEA Saclay, France – ²Iris AI, Bekkestua, Norway

DIFFUSE INTERSTELLAR BANDS (DIBs)

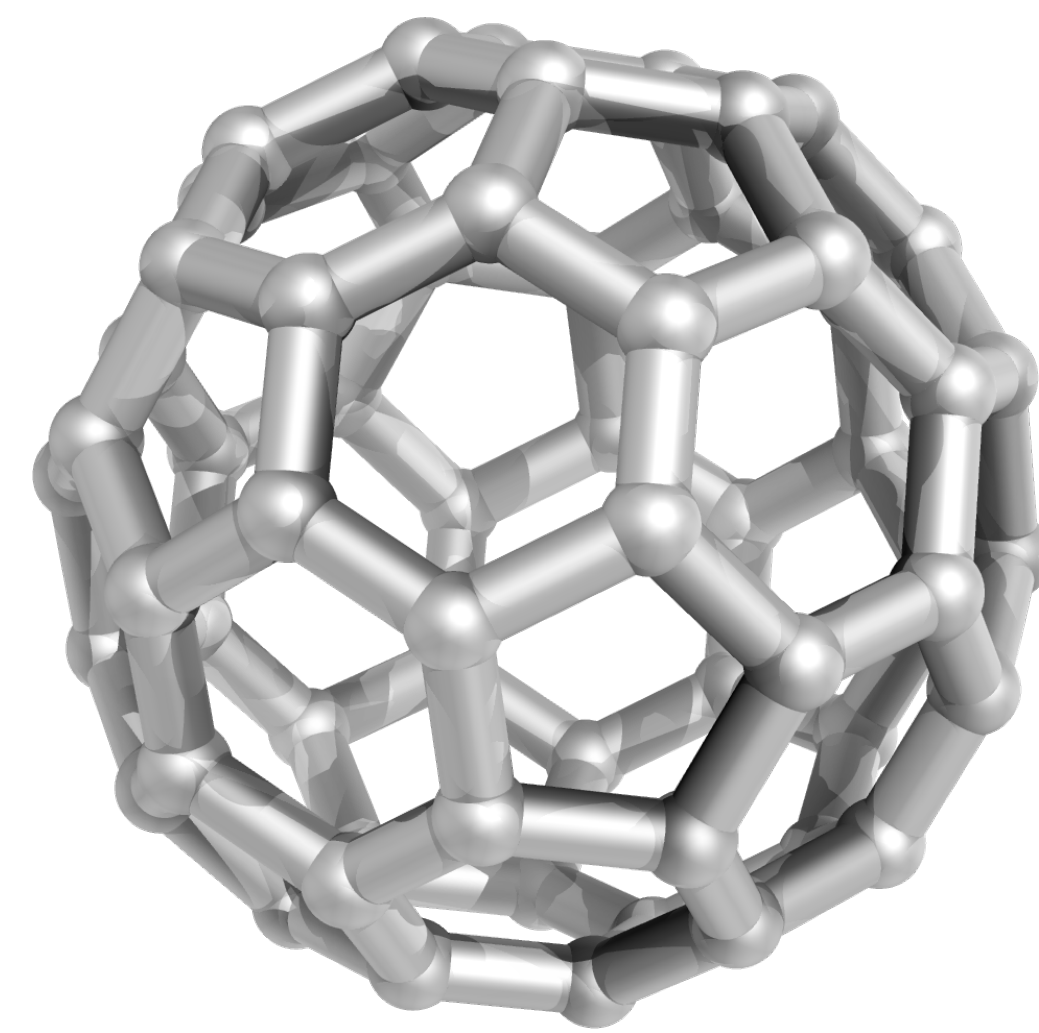


Observational constraints:

- Ubiquitous absorption features (Heger, 1922). More than 500 known bands (visible to NIR range).
- Interstellar origin (Merrill, 1934).
- Strength correlated with extinction.

Physical origin:

- Must come from molecules with ≈ 100 atoms (MacIsaac et al., 2022).
- Must be made of abundant interstellar atoms (H, C, N, O, etc.).
- 2 DIBs attributed to fullerene (C₆₀⁺) (Campbell et al., 2015).



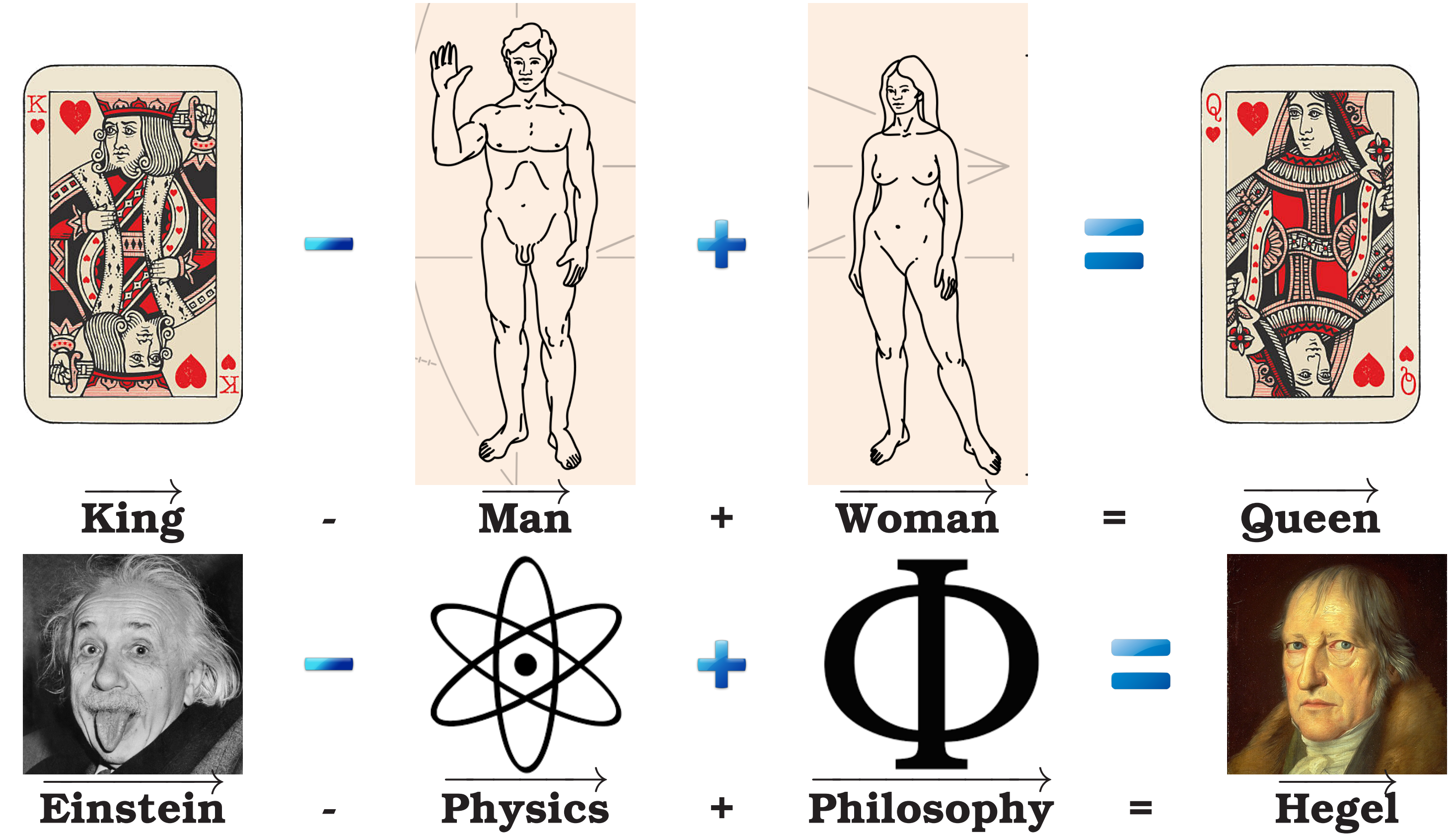
Buckminsterfullerene

NATURAL LANGUAGE PROCESSING (NLP)

NLP: machine-learning method applied to written texts \Rightarrow explore large corpuses, make parallel between disciplines, etc.

Word2vec (Mikolov et al., 2013): the meaning of a word is characterized by its context, i.e. the other words around it.

Word-embedding: each word is a vector in a high dimension space. Vector operations can then be performed to explore concepts, such as the following equations:



\Rightarrow We used NLP to explore the interdisciplinary literature looking for what would be $\overrightarrow{\text{DIBs}} - \overrightarrow{\text{astrophysics}} + \overrightarrow{\text{other discipline}}$.

DUG-OUT INTERDISCIPLINARY ARTICLES

	Article	Transitions	Closest DIB
Chromophores within proteins	Davies et al. (2009a) Elephant shark eye pigment	441.9 \pm 1.0 nm	442.82 \pm 0.17 nm
		493.7 \pm 2.6 nm	494.74 \pm 0.01 nm
		496.3 \pm 0.1 nm	496.39 \pm 0.01 nm
		498.3 \pm 0.3 nm	498.21 \pm 0.01 nm
		498.7 \pm 0.3 nm	498.74 \pm 0.01 nm
		504.1 \pm 1.0 nm	505.48 \pm 0.01 nm
		509.5 \pm 0.5 nm	509.21 \pm 0.01 nm
		510.1 \pm 0.2 nm	510.10 \pm 0.01 nm
		520.9 \pm 2.0 nm	521.79 \pm 0.01 nm
	534.2 \pm 1.0 nm	534.25 \pm 0.01 nm	
	547.8 \pm 2.2 nm	548.08 \pm 0.01 nm	
	Davies et al. (2009b) Agnathan eye pigments	501.0 \pm 0.1 nm	500.36 \pm 0.01 nm
		535.5 \pm 3.3 nm	535.88 \pm 0.01 nm
		544.1 \pm 5.0 nm	543.35 \pm 0.01 nm
	Wakakuwa et al. (2010) Butterfly eye pigment	425 nm	425.90 \pm 0.01 nm
453 nm		450.18 \pm 0.02 nm	
563 nm		563.50 \pm 0.01 nm	
620 nm		619.90 \pm 0.01 nm	
Dove et al. (1995) Coral pigment	640 nm	640.05 \pm 0.01 nm	
	560 nm	560.09 \pm 0.01 nm	
	580 nm	580.66 \pm 0.01 nm	
Spady et al. (2006) Cichlid eye pigment	590 nm	590.06 \pm 0.01 nm	
	423 nm	425.90 \pm 0.01 nm	
	456 nm	450.18 \pm 0.15 nm	
	472 nm	472.68 \pm 0.02 nm	
	518 nm	517.81 \pm 0.01 nm	
Fasick et al. (1998) Dolphin eye pigment	528 nm	529.8 \pm 0.01 nm	
	561 nm	560.98 \pm 0.01 nm	
	488 nm	488.00 \pm 0.12 nm	
	545 nm	545.06 \pm 0.83 nm	
Heme & other	Rémigy et al. (2003) Bacterium cytochrome	420 nm	425.90 \pm 0.01 nm
		525.2 nm	525.18 \pm 0.01 nm
	Schoot Uiterkamp et al. (1976)	545.4 nm	545.06 \pm 0.83 nm
		653 nm	653.65 \pm 0.01 nm
	755 nm	755.94 \pm 0.01 nm	
	Wolfbeis et al. (2001)	488 nm	488.00 \pm 0.01 nm
	Filosa (2001)	695 nm	694.46 \pm 0.01 nm
Maréchal et al. (2007)	445 nm	442.82 \pm 1.69 nm	

DISCUSSION ABOUT THE FOUND MOLECULES

Training corpus:

Generic, cross-domain corpus: 1.5×10^6 open access, English articles.

DIB specification: 939 DIB articles found on ADS.

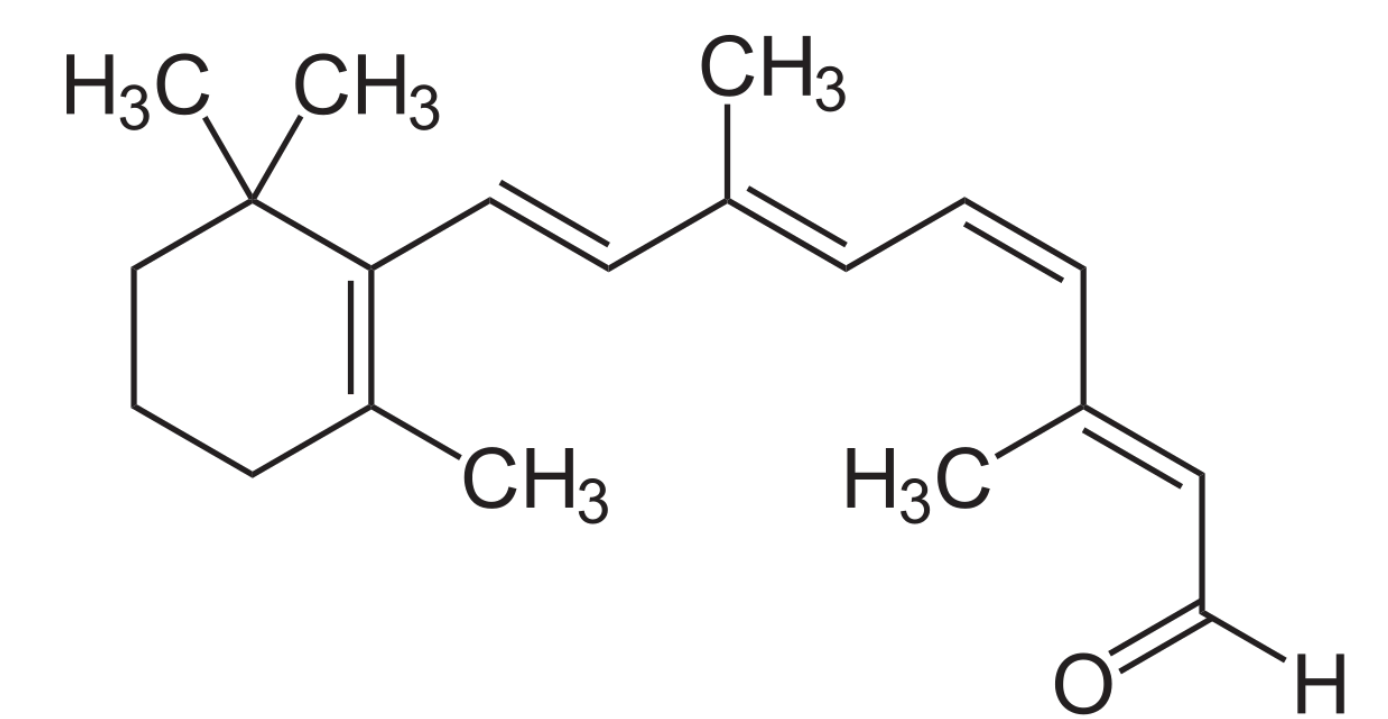
Unit disambiguation: select only papers with transitions matching DIBs.

Chromophores:

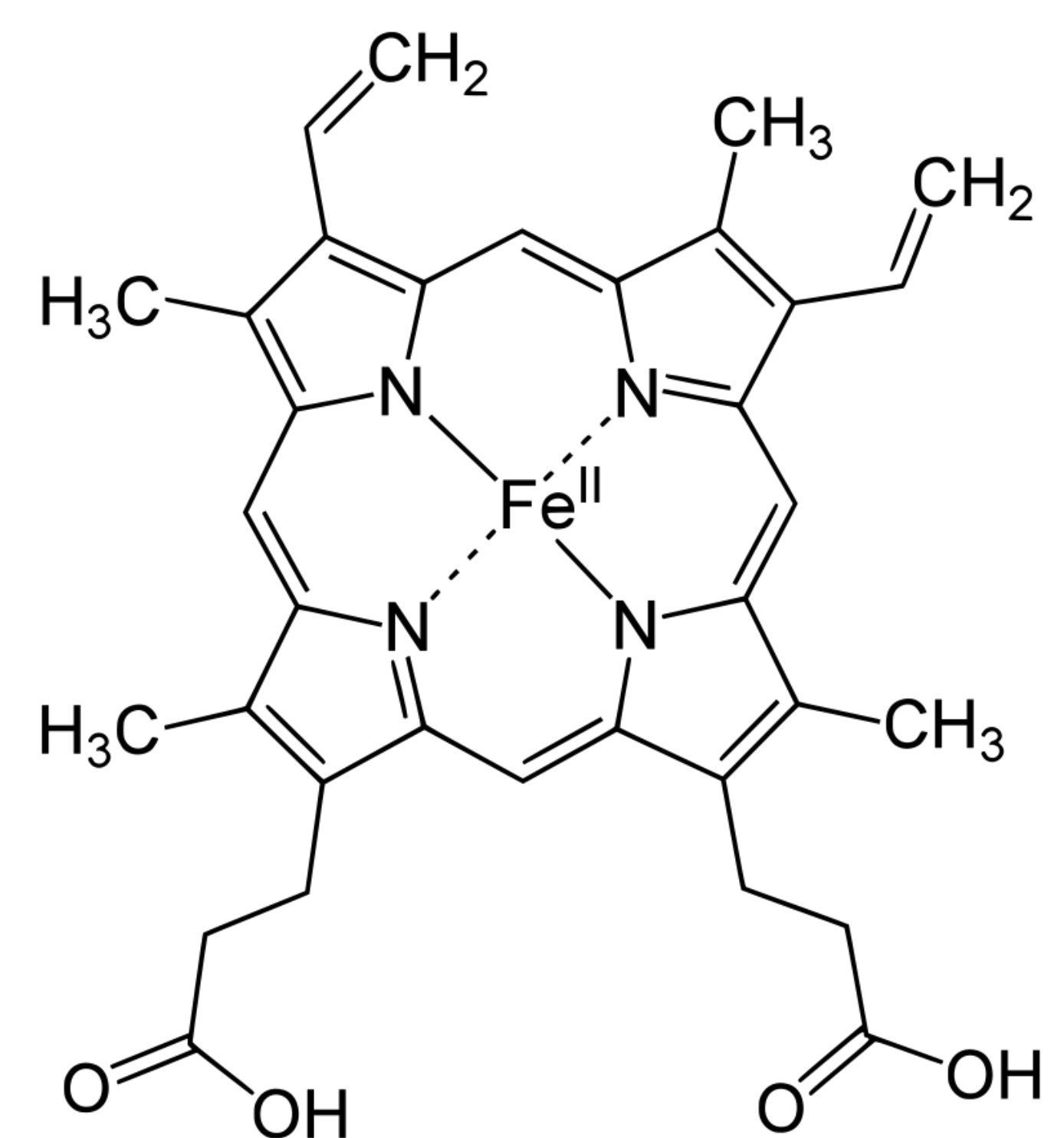
- Numerous papers with several matching transitions deal with chromophores.
- Chromophores could form moieties on nano-grains in the ISM (Jones, 2016a).

Heme:

- Several papers deal with heme or heme-related molecules.
- Jones (2016b) notes that heme could result from the reaction of N with a-C(:H) grains and exist in the ISM.



11-cis retinal (chromophore)



Heme b

REFERENCES

- Campbell, E. K., Holz, M., Gerlich, D., & Maier, J. P. 2015, Nature, 523, 322
- Davies, W. L., Carvalho, L. S., Tay, B. H., et al. 2009a, Genome research, 19(3), 415-426
- Davies, W. L., Collin, S. P., & Hunt, D. M. 2009b, Molecular biology and evolution, 26(8), 1803-1809
- Dove, S. G., Takabayashi, M., & Hoegh-Guldberg, O. 1995, The Biological bulletin, 189(3), 288-
- Fasick, J. I., Cronin, T. W., Hunt, D. M., & Robinson, P. R. 1998, Visual neuroscience, 15(4), 643-651
- Filosa, A. 2001, PhD thesis, Concordia University, unpublished
- Heger, M. L. 1922, Lick Observatory Bulletin, 10, 141
- Jones, A. P. 2016a, Royal Society Open Science, 3, 160223
- Jones, A. P. 2016b, Royal Society Open Science, 3, 160224
- MacIsaac, H., Cami, J., Cox, N. L. J., et al. 2022, arXiv e-prints, arXiv:2203.01803
- Maréchal, A., Mattioli, T. A., Stuehr, D. J., & Santolini, J. 2007, The Journal of biological chemistry, 282(19), 14101-14112
- Merrill, P. W. 1934, PASP, 46, 206
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. 2013, CoRR, abs/1301.3781
- Rémigy, H. W., Aivaliotis, M., Ioannidis, N., et al. 2003, Photosynthesis research, 78(2), 153-160
- Schoot Uiterkamp, A. J., Evans, L. H., Jolley, R. L., & Mason, H. S. 1976, Biochimica et biophysica acta, 453(1), 200-204
- Spady, T. C., Parry, J. W., Robinson, P. R., et al. 2006, Molecular biology and evolution, 23(8), 1538-1547
- van den Broek d'Obrenan, C., Galliano, F., Botev, V., Minton, J., & Wu, R. 2022, JIMIS, in press
- Wakakuwa, M., Terakita, A., Koyanagi, M., et al. 2010, PLoS One, 5(11):e15015
- Wolfbeis, O. S., Opitz, D., Werner, T., & Quart, A. 2001, Journal of molecular recognition, 14(1), 13-17