associated with slab rollback is produced at the corner between the plates. It also matches the observation that narrow slabs roll back relatively quickly, because the flow around their edges significantly reduces the corner suction[6,7]. The Ionian slab, which is narrower than it is long (Fig. 1), thus appears to be an excellent example of this process.

The truncated shape of this slab at its southern edge is clearly indicated by the abrupt cessation of seismic activity approximately at the north coast of Sicily. About 15 Myr ago this edge was torn away from another slab fragment, now detected by seismic tomography under the coasts of Algeria[23]. The Ionian fragment rolled back to its present position, opening the Tyrrhenian backarc basin, whereas its other fragment sank in the mantle under the coasts of Algeria[24].

The shape of the northern edge of the Ionian slab is less clear. Earthquake distribution and tomographic studies seem to indicate that the Ionian slab is disconnected from the nearby Adriatic slab[25,26] and that the Adriatic slab is decoupled from its mother plate, Adria[26–28]. In Fig. 2a we suggest that here, the sinking Adriatic plate creates a local flow in the asthenosphere that produced the Quaternary magmatism at Mount Vesuvius and nearby. However, the exact subsurface geometry at this triple junction, between the Ionian, Tyrrhenian and Adriatic microplates, is yet to be explored. □

Received 15 April; accepted 16 August 1999.


1. Barberi, F. *et al.* Evolution of a section of the African-Europe plate boundary: paleomagnetic and volcanological evidence from Sicily. *Earth Planet. Sci. Lett.* **21,** 269–276 (1974).
2. Condomines, M., Tanguy, J. C., Kieffer, G. & Allegre, C. J. Magmatic evolution of a volcano studied by $^{230}$Th $^{238}$U disequilibrium and trace elements systematics: The Etna case. *Geochim. Cosmochim. Acta* **46,** 1397–1416 (1982).
3. Tanguy, J. C. Tholeiitic basalt magmatism of Mount Etna and its relations with alkaline series. *Contrib. Mineral. Petrol.* **66,** 51–67 (1978).
4. Allard, P. *et al.* Mantle-derived helium and carbon in groundwater and gases of Mount Etna, Italy. *Earth Planet. Sci. Lett.* **148,** 501–516 (1997).
5. Allard, P. *et al.* Eruptive and diffuse emissions of $CO_2$ from Mount Etna. *Nature* **351,** 387–391 (1991).
6. Nur, A., Dvorkin, J., Mavko, G. & Ben Avraham, Z. Speculations on the origin and fate of backarc basins. *Ann. Geofis.* **36,** 155–163 (1991).
7. Dvorkin, J., Nur, A., Mavko, G. & Ben Avraham, Z. Narrow subducting slabs and the origin of backarc basins. *Tectonophysics* **227,** 63–79 (1993).
8. Wendt, J. I., Regelous, M., Collerson, K. D. & Ewart, A. Evidence for a contribution from two mantle plumes to the island arc lavas from northern Tonga. *Geology* **25,** 611–614 (1997).
9. Turner, S. & Howkesworth, C. Using geochemistry to map mantle flow beneath the Lau basin. *Geology* **26,** 1019–1022 (1998).
10. Gvirtzman, Z. & Nur, A. Plate detachment, asthenosphere upwelling, and topography across subduction zones. *Geology* **27,** 563–566.
11. Monaco, C., Tapponnier, P., Tortorici, L. & Gillot, P. Y. Late quaternary slip rates on the Acireale-Piedimonte normal fault and tectonic origin of Mt. Etna (Sicily). *Earth Planet. Sci. Lett.* **147,** 125–139 (1997).
12. Frepoli, A., Selavagi, G., Chiarabba, C. & Amoto, A. State of stress in southern Tyrrhenian subduction zone from fault-plane solutions. *Geophys. J. Int.* **125,** 879–891 (1996).
13. Lachenbruch, A. H. & Morgan, P. Continental extension, magmatism and elevation: Formal relations and rules of thumb. *Tectonophysics* **174,** 39–62 (1990).
14. Jones, H. J., Unruh, J. R. & Sonder, L. J. The role of gravitational potential energy in active deformation in the southwestern United States. *Nature* **381,** 37–41 (1996).
15. Steinmetz, L., Ferrucci, F., Hirn, C. A., Moreli, C. & Nicholich, R. A 550 km long Moho traverse in the Tyrrhenian Sea from O.B.S. recorded $P_N$ waves. *Geophys. Res. Lett.* **10,** 428–431 (1983).
16. Scarascia, S., Lozej, A. & Cassinis, R. Crustal structures of the Ligurain, Tyrrhenian, and Ionian seas and adjacent onshore areas interpreted from wide-angle seismic profiles. *Boll. Geofis. Teor. Applic.* **36,** 5–19 (1994).
17. Moreli, C. Current knowledge on the crustal properties of Italy. *Ann. Geofis.* **37,** 1113–1130 (1997).
18. Gvirtzman, Z. & Nur, A. Residual topography, lithospheric structure and sunken slabs in the central Mediterranean. *Earth Planet. Sci. Lett.* (submitted).
19. Kastens, K. *et al.* ODP Leg 107 in the Tyrrhenian Sea: Insight into passive and backarc basin evolution. *Geol. Soc. Am. Bull.* **100,** 1140–1156 (1988).
20. Westaway, R. Quaternary uplift of south Italy. *J. Geophys. Res.* **98,** 21741–21772 (1993).
21. Mazzuoli, R., Tortorici, L. & Ventura, G. Oblique rifting in Salina, Lipari, and Vulcano islands (Aeolian islands, southern Italy). *Terra Nova* **7,** 444–452 (1995).
22. Isacks, B. L. & Barazangi, M. in *Island Arcs, Deep Sea Trenches, and Backarc Basins* (eds Talawani, M. and Pitman, W. C.) 99–114 (Maurice Ewing Ser. 1, American Geophysical Union, Washington DC, 1977).
23. Bijwaad, H., Spakman, W. & Engdahl, E. R. Closing the gap between regional and global travel time tomography. *J. Geophys. Res.* **103,** 30055–30078 (1998).
24. Carminati, E., Wortel, M. J. R., Spakman, W. & Sabadina, R. The role of slab detachment processes in the opening of the western-central Mediterranean basin: some geological and geophysical evidence. *Earth Planet. Sci. Lett.* **160,** 651–665 (1998).
25. Mele, G. High frequency wave propagation from mantle earthquakes in the Tyrrhenian Sea: New constraints for the geometry of the south Tyrrhenian subduction zone. *Geophys. Res. Lett.* **25,** 2877–2880 (1998).
26. Amato, A., Alessandrini, B., Cimini, G. B., Frepoli, A. & Selvaggi, G. Active and remnant subducted slabs beneath Italy: evidence from seismic tomography and seismicity. *Ann. Geofis.* **36,** 201–214 (1993).
27. Wortel, M. J. R. & Spakman, W. Structure and dynamics of subducted lithosphere in the Mediterranean region. *Proc. Koninklijke Ned. Acad. Wetensch.* **95,** 325–347 (1992).
28. Spakman, W., van de Lee, S. & van der Hilst, R. Travel time tomography of the European-Mediterranean mantle down to 1400 km. *Phys. Earth Planet. Inter.* **79,** 3–74 (1993).


Correspondence and requests for materials should be addressed to Z.G. (e-mail: zohar@vms.huji.ac.il).

# Decline in Mesozoic reef-building sponges explained by silicon limitation

**Manuel Maldonado, M. Carmen Carmona, María J. Uriz & Antonio Cruzado**

*Centro de Estudios Avanzados de Blanes (CSIC), Camino de Santa Barbara s/n, Blanes 17300, Girona, Spain*

**Several unrelated clades of siliceous sponges proliferated on the shelves of the Jurassic Tethys Sea, becoming prominent builders in reefs and near-shore mounds[1–4]. Many of these builders are characterized by massive, rock-like skeletons made of spicules with a characteristic terminal hypersilicification[4,5]. Such hypertrophied spicules are generically known as desmas, irrespective of their phylogenetic origin[5]. Desma-bearing sponges virtually disappeared from reefs and other neritic environments during the Cretaceous and the Early Tertiary[1,2,4,6], but have subsisted in relict populations in deeper, bathyal waters[5,7,8]. The causes of the decline and bathymetric shift of these sponges remain obscure. Here we show experimentally that the concentration of silicic acid in seawater modulates the phenotypic expression of the various spicule types genetically available in a sponge species. We also show that the concentration of this nutrient in Recent surface waters is insufficient for this species to secrete its desmas. These findings indicate that silicon limitation, probably aggravated in shallow waters by the diatom burst around the Cretaceous–Tertiary boundary[9,10], may have forced neritic sponges with desmas to either lighten their skeletons or move to deeper, silicon-rich environments.**

Silicon (Si) is one of the most abundant elements on earth, but only its soluble form, as monomeric, undissociated silicic acid, $Si(OH)_4$, is biologically assimilable. Several groups of marine organisms, such as diatoms, radiolarians, choanoflagellates and sponges, take up $Si(OH)_4$ from seawater to build their opal (amorphous hydrated silica) skeletons. Biological consumption of $Si(OH)_4$ renders present-day oceans undersaturated with this nutrient. Yet deep waters are relatively rich in $Si(OH)_4$, with concentrations varying from 10 to 180 μM, whereas surface waters are notably poorer, with a mean concentration of less than 3 μM[11,12].

Diatoms are responsible for the depletion of $Si(OH)_4$ in the photic zone and largely control the cycling of Si in Recent oceans[13]. Radiolaria are quantitatively important only in the Pacific equatorial bell[14], whereas siliceous sponges and silicoflagellates are considered unimportant on a global scale[10]. However, there is sedimentological evidence that sponges once controlled the Si cycling in neritic environments[10]. Indeed, desma-bearing sponges and hexactinellids, two unrelated groups with substantial siliceous

skeletons, proliferated in neritic assemblages and became prominent builders in reefs and near-shore mounds during the Upper Jurassic[1–4]. We show that the encrusting demosponge *Crambe crambe* (Schmidt), one of the most common macroinvertebrates in the western Mediterranean sublittoral[15], provides some clues to understand why sponges with massive silica skeletons declined in neritic environments.

In most natural populations, skeletons of *C. crambe* contain only a single spicule type, small needles called styles[16]; however, specimens are occasionally found in which the skeleton is supplemented by large styles, scarce C-shaped spicules (called isochelae) or scarce aster-like desmas[16]. This variability has puzzled sponge biologists, as the number of spicule types secreted by a species is a cornerstone character in sponge taxonomy; it is assumed that the spicule set does

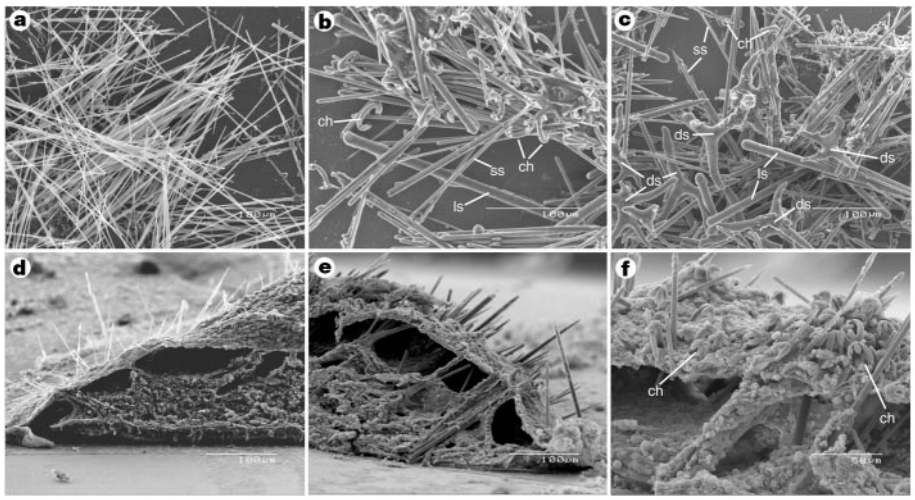


**Figure 1** Skeletons produced under different concentrations of $Si(OH)_4$. **a**, Skeleton of a control sponge built up by a single, needle-like spicule type called style. **b**, Skeleton of a sponge reared in 30 μM $Si(OH)_4$ showing three spicule types: small styles (ss); large styles (ls); and small C-shaped spicules called isochelae (ch). Note differences in size and robustness between styles of control sponges and styles of sponges reared in Si-enriched seawater. **c**, Skeleton of a sponge reared in 100 μM $Si(OH)_4$ showing four spicule types: styles in two size categories (ss, ls); isochelae (ch); and aster-like desmas (ds). **d**, **e**, Skeletal arrangements seen by fracturing sponges reared in controls and in 30 μM $Si(OH)_4$, respectively. **f**, Ectosome of a sponge reared in 30 μM $Si(OH)_4$ showing great abundance of protruding isochelae (ch).

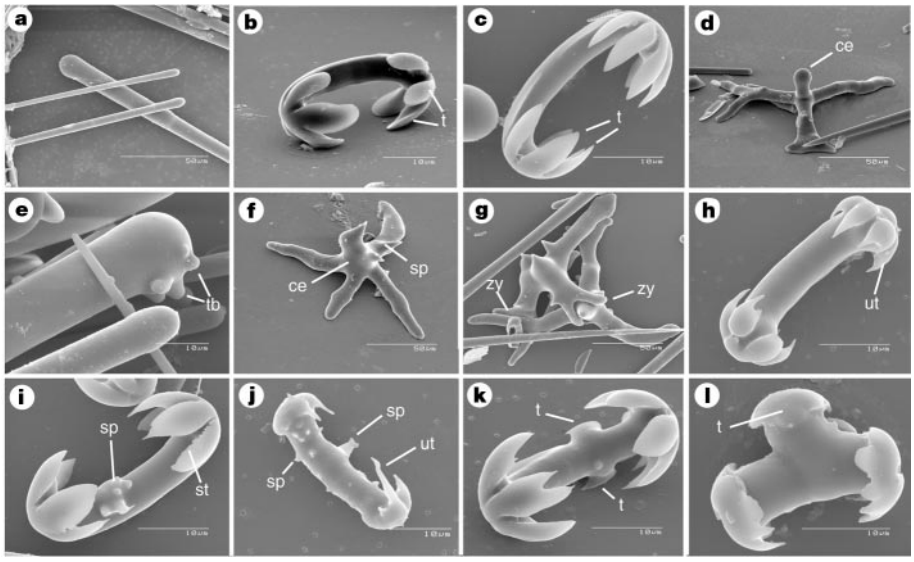


**Figure 2** Spicules produced under two experimental levels of Si supply. **a**–**d**, Spicules produced in 30 μM $Si(OH)_4$. **a**, Styles in two size categories. **b**, **c**, Isochelae showing intraspecific variability in the shape and number of teeth (t), traits assumed to have taxonomic value in some sponge groups. **d**, An aster-like desma with an undeveloped centrum (ce). **e**–**l**, Spicules produced in 100 μM $Si(OH)_4$. **e**, Small and large styles, the latter with a tuberose ornamentation (tb), a trait of presumed taxonomic value. **f**, Aster-like desma showing a well-developed centrum (ce) with three large spines (sp). **g**, Desmas still in contact after treatment with nitric acid, indicating some degree of fusion at the contact points; zy, zygoses. **h**–**l**, Common, malformed isochelae showing teeth reduced to ungues (ut) (**h**), serrated teeth (st) and spines (sp) on the shaft (**i**), reduced teeth (ut) and spines (sp) on the shaft (**j**), and teeth (t) on the shaft (**k**–**l**).

not vary within species.

We examined the production of spicules by newly settled sponges maintained at 3 concentrations of $Si(OH)_4$ for 14 weeks (August to November 1998). One set of 36 sponges was cultured in seawater from their natural habitat. Concentration of $Si(OH)_4$ in these controls ranged from 0.03 to 4.5 μM (mean ± s.e. = 0.741 ± 0.133 μM). Two other sets were reared at elevated concentrations of $Si(OH)_4$, referred to hereafter as Si-medium (30.235 ± 0.287 μM) and Si-high (100.041 ± 0.760 μM) treatments.

All sponges reared in control cultures formed only a single spicule type: small styles (Fig. 1a, d). In contrast, all sponges reared in Si-medium conditions secreted at least three spicule types in great abundance: small styles, large styles and 28–33-μm long isochelae (Figs 1b, e–f, 2a–c). Fifty per cent of these sponges also formed a small number (1–5) of a fourth spicule type: aster-like desmas (Fig. 2d). In Si-high conditions, all sponges produced all four spicule types in great abundance (Figs 1c, 2e–l). Two of these sponges also secreted a fifth category: 8–10-μm long isochelae (spicule not shown). Notably, all sponges used in the experiments were from a large population of *C. crambe* studied for more than 15 years, in which we never found large styles, isochelae or desmas.

The concentration of $Si(OH)_4$ had a striking effect on both the size and shape of spicules (Figs 1a–c, 2). In Si-high conditions, a few isochelae formed spines and tooth-like structures at the middle of the shaft (Fig. 2k, l), which resembled spicules from the fossil record[4,17]. Differences in $Si(OH)_4$ concentration also affected its uptake, as determined by analyses during the first five weeks of the experiment. Consumption curves of $Si(OH)_4$ were parallel in all three treatments, with a peak immediately after metamorphosis (Fig. 3). Sponges took up $Si(OH)_4$ more efficiently in Si-medium than in Si-high conditions, indicating that silica transport into skeletal secretory cells (sclerocytes) may be inhibited at unusually high concentrations. This may explain the production of numerous malformed isochelae in Si-high conditions (Fig. 2h–l). More significantly, $Si(OH)_4$ uptake was higher in the two Si-enriched treatments than in controls (Fig. 3), which indicates that *C. crambe* is limited by Si availability in coastal waters.

Our results show that *C. crambe* is genetically capable of producing spicule types that are not normally found in natural populations. Limited silicic acid in coastal waters appears to be the reason for this phenotypic skeletal inhibition, which affects entire populations of thousands of individuals. Our results also indicate that, in contrast to what is believed, each spicule type may be secreted by a specific sclerocyte type, which is activated by a particular $Si(OH)_4$ concentration threshold. Significantly, aster-like desmas are consistently formed only at concentrations of $Si(OH)_4$ (100 μM) that are not found in most surface waters today. Sponges with aster-like desmas were fairly common in Jurassic–Cretaceous assemblages, but all these are now extinct, except the sublittoral genus *Crambe* and, perhaps, the monotypic, deep-water genus *Vetulina*[4,16,18].

The possibility that $Si(OH)_4$ concentration influenced the skeletal evolution of sponges is discussed elsewhere[4,19]. Two peaks of volcanic activity in the Triassic (Fassanian and Sevatian), which probably increased the amount of dissolved Si in seawater, may have favoured the radiation and proliferation of sponges with massive siliceous skeletons during the Jurassic[4,19]. The hypothesis that seawater was saturated with dissolved Si at that time is also supported by the robust skeletons of fused spicules and state of preservation of fossils[9,20]. The study of chert beds indicates that domination of the Si cycle by diatoms was completed only in the Eocene, after a long period of expansion and radiation beginning during the Cretaceous[21]. The contribution of sponge spicules to neritic cherts began to decline concurrently with the radiation and proliferation of diatoms[9,10]. The palaeontological record indicates that both desma-bearing and hexactinellid sponges had disappeared from neritic environments and penetrated the bathyal and abyssal zones by the Lower Tertiary[1,4,6]. During the Cenozoic, siliceous spicules of neritic sponges became progressively less robust and fused spicules less frequent[6,9]. The few Recent hexactinellids still found at shallow depths occur either in microhabitats with enhanced Si supply[22] or at high latitudes[23,24], where the concentration of $Si(OH)_4$ in surface water is higher. There is also correlative evidence that the expansion of diatoms reduced both diversity and abundance of other Si-dependent organisms, such as radiolarians[25].

The hypothesis that low $Si(OH)_4$ concentration in shallow waters might limit sponges has been disregarded[7,26,27], as the Si-uptake system of sponges is expected to have adapted to the decline in the availability of this nutrient. In contrast, our results indicate that Si concentration is a limiting factor on shallow-water sponges and that this pressure may be strong enough to shape their skeletons and restrict their bathymetric distributions. Chronic Si limitation is also supported by the only study on Si uptake in a sublittoral sponge, *Halichondria panicea*. This showed Michaelis–Menten-uptake kinetics with saturation at $Si(OH)_4$ concentrations far higher than the maximum available in the natural habitat of the species[28]. We propose that chronic limitation may be responsible not only for much of the unexplained skeletal variability that puzzles sponge taxonomists, but also for much of the skeletal change observed in shallow-water sponges during the Cenozoic. In which sponge groups the 'skeletal change' represents evolution, and in which it merely reflects phenotypic inhibition of certain spicule types remain unclear. Therefore, we cannot rule out the possibility that by rearing more sponges in seawater saturated with silicic acid some apparently extinct species or spicules may 'come back to life' for taxonomists. □

## Methods

The sponges used in the experiment were grown in the laboratory from free-swimming

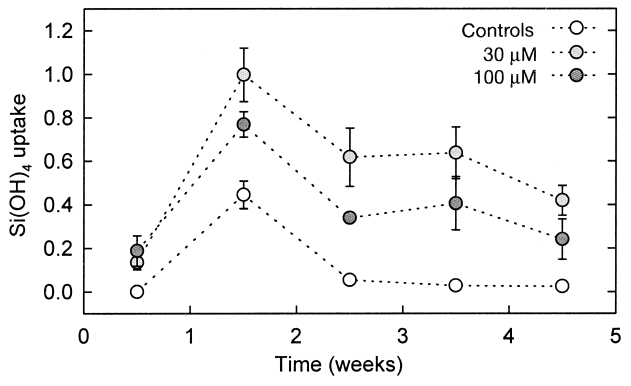


**Figure 3** Consumption (mean ± s.e.) of $Si(OH)_4$ by sponges (μmol per individual per week) as a function of substrate concentration and time during the first five weeks of the experiment.

larvae collected randomly in the field (2° 48.12′ N, 41° 40.33′ E) by SCUBA. Between 5 and 10 juveniles were recruited successfully in each of 15, 1 l polystyrene containers ($n = 15$), the bottom of which was covered with an acetate sheet that served as substratum for sponge attachment. Containers were then randomly distributed in 3 groups, and sponges in each group were reared for 14 weeks in 3 different concentrations of $Si(OH)_4$: 0.741 ± 0.133, 30.235 ± 0.287 and 100.041 ± 0.760 μM (mean ± s.e.). All cultures were prepared using 0.22 μm polycarbonate-filtered seawater, which was collected from the sponge habitat, handled according to standard methods to prevent Si contamination[29] and enriched in dissolved silica, when treatments required, by using $Na_2SiF_6$. During the experiment, all sponges were fed by weekly addition of 2 ml of a bacterial culture ($40–60 \times 10^6$ bacteria $ml^{-1}$) to each container[30]. The seawater was replaced weekly, with regeneration of initial food and $Si(OH)_4$ levels. The concentration of $Si(OH)_4$ in cultures was determined on 3 replicates of 1 ml seawater samples per container by using a Bran-Luebbe TRAACS 2000 nutrient autoanalyser. After week 5, the accidental contamination of some culture containers by diatoms rendered subsequent estimates of Si uptake by sponges unreliable, so we discarded them for the study.

For the study of the skeleton, sponges were treated according to standard methods[30] and examined in a Hitachi S-2300 scanning electron microscope (SEM).

Received 21 April; accepted 16 August 1999.


1. Hartman, W. D., Wendt, J. W. & Wiedenmayer, F. Living and fossil sponges. Notes for a short course. *Sedimentia* **8,** 1–274 (1980).
2. Ghiold, J. The sponges that spanned Europe. *New Scient.* **129,** 58–62 (1991).
3. Leinfelder, R. R. Upper Jurassic reef types and controlling factors. *Profil* **5,** 1–45 (1993).
4. Wiedenmayer, F. Contributions to the knowledge of post-Paleozoic neritic and archibental sponges (Porifera). *Schweiz. Paläont. Abh.* **116,** 1–147 (1994).
5. Lévi, C. in *Fossil and Recent Sponges* (eds Reitner, J. & Keupp, H.) 72–82 (Springer, New York, 1991).
6. Moret, L. Contribution à l'étude des spongiaires siliceux du Miocene de l'Algerie. *Mém. Soc. Géol. Fr.* **1,** 1-27 (1924).
7. Vacelet, J. Indications de profondeur donnés par les Spongiaires dans les milieux benthiques actuels. *Géol. Méditerr.* **15,** 13–26 (1988).
8. Maldonado, M. & Young, C. M. Bathymetric patterns of sponge distribution on the Bahamian slope. *Deep-Sea Res. I* **43,** 897–915 (1996).
9. Lowenstam, H. A. & Weiner, S. *On Biomineralization* (Oxford Univ., Oxford, 1989).
10. Maliva, R. G., Knoll, A. H. & Siever, R. Secular change in chert distribution: a reflection of evolving biological participation in the silica cycle. *Palaios* **4,** 519–532 (1989).
11. Nelson, D. M., Tréguer, P., Brzezinski, M. A., Leynaert, A. & Quéguiner, B. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Glob. Biochem. Cycles* **9,** 359–372 (1995).
12. Tréguer, P. *et al.* The silica balance in the world ocean: a reestimate. *Science* **268**, 375–379 (1995).
13. Calvert, S. E. in *Silicon Geochemistry and Biogeochemistry* (ed. Aston, S. R.) 143–186 (Academic, London, 1983).
14. Lisitzyn, A. P. Sedimentation in the world ocean. *Soc. Econ. Palaeon. Mineral. Spec. Pub.* **17,** 1–218 (1972).
15. Weinberg, S. *Découvrir la Méditerranée* (Éditions Nathan, Paris, 1993).
16. Maldonado, M. & Uriz, M. J. Skeletal morphology of two controversial poecilosclerid genera (Porifera, Demospongiae): *Discorhabdella* and *Crambe*. *Helgoländer Meeresunters.* **50**, 369–390 (1996).
17. Hinde, G. J. & Holmes, W. M. On the sponge remains in the Lower Tertiary strata near Oamaru, New Zealand. *J. Linn. Soc. Zool.* **24,** 177–262 (1892).
18. Kelly-Borges, M. & Pomponi, S. A. Phylogeny and classification of lithisthid sponges (Porifera: Demospongiae): a preliminary assessment using ribosomal DNA sequence comparisons. *Mol. Mar. Biol. Biotech.* **3,** 87–103 (1994).
19. Mostler, H. Poriferenspiculae der alpinen Trias. *Geol. Paläont. Mitt. Innsbruck.* **6,** 1–42 (1976).
20. Palmer, T. J. & Fürsich, F. T. Ecology of sponge reefs from the Upper Bathonian of Normandy. *Palaeontology* **24,** 1–23 (1981).
21. Burckle, L. H. in *Introduction to Marine Micropaleontology* (eds Haq, B. U. & Boersma, A.) 245–266 (Elsevier, Amsterdam, 1978).
22. Austin, B. Underwater birdwatching. *Canadian Tech. Rep. Hydro. Ocean. Sci.* **38,** 83–90 (1984).
23. Koltun, V. M. in *The Biology of the Porifera* (ed. Fry, W. G.) 285–297 (Academic, London, 1970).
24. Tabachnick, K. R. in *Sponges in Time and Space* (eds van Soest, R. M. W., van Kempen, T. M. G. & Braekman, J. C.) 225–232 (A. A. Balkema, Rotterdam, 1994).
25. Harper, H. E. & Knoll, A. H. Silica, diatoms and Cenozoic radiolarian evolution. *Geology* **3,** 175–177 (1975).
26. Hartman, W. D. in *Silicon and Siliceous Structures in Biological Systems* (eds Simpson, T. L. & Volcani, B. E.) 453–493 (Springer Verlag, New York, 1981).
27. Pisera, A. Upper Jurassic siliceous sponges from Swabian Alb: taxonomy and paleoecology. *Palaeont. Pol.* **57,** 1–216 (1997).
28. Reincke, T. & Barthel, D. Silica uptake kinetics of *Halichondria panicea* in Kiel Bight. *Mar. Biol.* **129,** 591–593 (1997).
29. Grasshoff, K., Ehrardt, M. & Kremling, K. *Methods of Seawater Analysis* (Verlag Chemie, Weinheim, 1983).
30. Maldonado, M. & Uriz, M. J. An experimental approach to the ecological significance of microhabitat-scale movement in an encrusting sponge. *Mar. Ecol. Prog. Ser.* **185,** 239–255 (1999).



### Acknowledgements

We thank S. Pla for help with nutrient analyses, technicians of the Servicio de Microscopia for help with SEM, and E. Ballesteros, C. M. Young, A. Pisera and R. Rycroft for comments on the manuscript.


Correspondence and requests for materials should be addressed to M.M. (e-mail: maldonado@ceab.csic.es).

# Learning the parts of objects by non-negative matrix factorization

**Daniel D. Lee*** **& H. Sebastian Seung*†**

** Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA*
*† Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

**Is perception of the whole based on perception of its parts? There is psychological[1] and physiological[2,3] evidence for parts-based representations in the brain, and certain computational theories of object recognition rely on such representations[4,5]. But little is known about how brains or computers might learn the parts of objects. Here we demonstrate an algorithm for non-negative matrix factorization that is able to learn parts of faces and semantic features of text. This is in contrast to other methods, such as principal components analysis and vector quantization, that learn holistic, not parts-based, representations. Non-negative matrix factorization is distinguished from the other methods by its use of non-negativity constraints. These constraints lead to a parts-based representation because they allow only additive, not subtractive, combinations. When non-negative matrix factorization is implemented as a neural network, parts-based representations emerge by virtue of two properties: the firing rates of neurons are never negative and synaptic strengths do not change sign.**

We have applied non-negative matrix factorization (NMF), together with principal components analysis (PCA) and vector quantization (VQ), to a database of facial images. As shown in Fig. 1, all three methods learn to represent a face as a linear combination of basis images, but with qualitatively different results. VQ discovers a basis consisting of prototypes, each of which is a whole face. The basis images for PCA are 'eigenfaces', some of which resemble distorted versions of whole faces[6]. The NMF basis is radically different: its images are localized features that correspond better with intuitive notions of the parts of faces.

How does NMF learn such a representation, so different from the holistic representations of PCA and VQ? To answer this question, it is helpful to describe the three methods in a matrix factorization framework. The image database is regarded as an $n \times m$ matrix $V$, each column of which contains $n$ non-negative pixel values of one of the $m$ facial images. Then all three methods construct approximate factorizations of the form $V \approx WH$, or

$$V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia} H_{a\mu} \qquad (1)$$

The $r$ columns of $W$ are called basis images. Each column of $H$ is called an encoding and is in one-to-one correspondence with a face in $V$. An encoding consists of the coefficients by which a face is represented with a linear combination of basis images. The dimensions of the matrix factors $W$ and $H$ are $n \times r$ and $r \times m$, respectively. The rank $r$ of the factorization is generally chosen so that $(n + m)r < nm$, and the product $WH$ can be regarded as a compressed form of the data in $V$.

The differences between PCA, VQ and NMF arise from different constraints imposed on the matrix factors $W$ and $H$. In VQ, each column of $H$ is constrained to be a unary vector, with one element equal to unity and the other elements equal to zero. In other words, every face (column of $V$) is approximated by a single basis image (column of $W$) in the factorization $V \approx WH$. Such a unary encoding for a particular face is shown next to the VQ basis in Fig. 1. This unary representation forces VQ to learn basis images that are prototypical faces.