27 April 1984, Volume 224, Number 4647

SCIENCE

Interstellar Matter and Chemical Evolution

M. Peimbert, A. Serrano, S. Torres-Peimbert

The chemical composition of the interstellar medium can be determined by studying the emission line spectra of gaseous nebulae. Hydrogen, helium, carbon, nitrogen, oxygen, and neon are the six most abundant elements in our galaxy and in those galaxies for which accurate abundances have been determined. The relative abundance of these elements in the interstellar medium can be derived from H II regions, planetary nebulae, and supernova remnants.

Summary. The determination of chemical abundances of the interstellar medium is related to three different areas of astronomy: pregalactic conditions, stellar evolution, and chemical evolution of galaxies. The observed abundances permit the testing of theories associated with these areas. Knowledge of the abundances of helium, carbon, nitrogen, and neon relative to the abundance of hydrogen in the interstellar medium puts limits on current models of cosmology, stellar evolution, and galactic evolution.

Accurate abundance determinations of H II regions have made it possible to find small differences in the abundance of elements among various galaxies and among different regions of the same galaxy. These small differences occur because material produced in stellar interiors and rich in elements heavier than hydrogen has been injected into the interstellar medium. The enrichment of the interstellar medium is due to loss of mass from stars of intermediate mass that become planetary nebulae before turning into white dwarfs and from massive stars that explode as supernovae. By studying H II regions where almost no star formation has occurred, it is possible to determine pregalactic chemical abundances; and by studying H II regions in galaxies where a substantial fraction of the gas has gone into stars, it is possible to study the chemical evolution of galaxies.

Abundance determinations of supernova remnants are not discussed in this article because their accuracy is considerably smaller than that of abundance determinations of H II regions and planetary nebulae. Similarly, stellar surface abundances are not discussed because, with the exception of a few results for the Magellanic Clouds, there are no individual determinations for stars in other galaxies. Even in our galaxy, the accuracy with which the relative abundances of some key elements can be derived is smaller for stars than for H II regions and planetary nebulae. This is particularly the case for the He/H ratio.

Observational Data on Abundances

H II regions are conglomerations of gas and dust where stars are being formed at present; the most massive H II regions are hot enough to be able to ionize hydrogen atoms (see Fig. 1). The chemical composition of their gaseous component has been affected by stellar enrichment throughout the life of the galaxy. They are called H II regions because most of the gas is in the form of ionized hydrogen. In H II regions of the galactic neighborhood about 90 percent of the atoms are of hydrogen, about 10 percent are of helium, and about 0.1 percent are of all the other elements combined. The fractions of these three constituents, normalized by mass, are called X, Y, and Z, and for the interstellar medium in the solar neighborhood their values amount to 0.70, 0.28, and 0.02, respectively.

Planetary nebulae are shells of gas ejected from and expanding about extremely hot central stars. Stars of intermediate mass go through the planetary nebulae stage after the red giant phase and before they become white dwarfs. The ultraviolet flux of the hot central star ionizes the ejected shells. The physical processes responsible for the emission line spectrum in the outer envelopes of planetary nebulae and in H II regions are very similar.

The emission line intensity of a given ion in a gaseous nebula (planetary nebulae and H II regions are called gaseous nebulae) depends on the distribution along the line of sight of the electron density (N_e), the electron temperature (T_e), and the ionic concentration, $N(X^{+i})$, as well as on atomic parameters that represent the likelihood of photon emission. Values of N_e can be determined from the ratio of two emission lines of the same ion with almost the same excitation energy, and values of T_e can be determined from the ratio of two emission lines of the same ion with very different excitation energies (*1*).

The relative ionic abundances are determined from the ratio of two line intensities of the ions involved. The lines chosen are not affected by radiative transfer effects, and the accuracy of the abundance determinations is given by the accuracy of the line intensity determinations and the precision of atomic constants. There are two types of general processes that produce emission line photons in a gaseous nebula: recombination, which in the case of hydrogen is

The authors are astronomers at Instituto de Astronomía, Universidad Nacional Autónoma de México, México 04510 D.F., México.

Fig. 1. Galactic H II region, called Messier 8 or the Lagoon Nebula. From spectra of its bright regions it is possible to obtain accurate abundances of hydrogen, helium, carbon, nitrogen, oxygen, neon, and heavier elements. This is a reproduction of a photograph taken with the Lick Observatory Crossley 0.9-meter telescope.

responsible for the Balmer lines, and collisional excitation of atomic energy levels followed by radiative de-excitation. The recombination line intensities depend weakly on T_e, whereas the intensities of the collisionally excited lines in the optical and ultraviolet electromagnetic domains depend strongly on T_e. Therefore, high-quality determinations of the He^+/H^+ abundance ratio, which are based on recombination lines, have a typical accuracy of $10^{0.04}$ (≡ 0.04 dex), and high-quality ionic determinations of carbon, nitrogen, oxygen, and neon, relative to hydrogen, based on collisionally excited lines have typical accuracies of about 0.15 dex as a result of their higher dependence on T_e and the lower precision of the atomic constants.

For some of the atoms we do not observe lines corresponding to all the stages of ionization present in the nebula. Obtaining the total abundances requires evaluation of the fraction of atoms in the unobserved stages of ionization. This estimation is usually based on the degree of ionization and on ionization structure models. Observations at different positions of a given nebula often permit a better determination of this correction.

Some of the more accurate abundance determinations for galactic and extragalactic H II regions are given in Table 1, where the Orion Nebula is representative of the solar neighborhood, and average values for several H II regions in the Large Magellanic Cloud (LMC) and the Small Magellanic Cloud (SMC). Table 1 also shows the abundances for planetary nebulae of different types; the $(He\text{-}N)_{rich}$ and intermediate planetary nebulae are solar neighborhood objects whose parent stars are relatively young (1×10^8 to 3×10^9 years), whereas halo planetary nebulae have very old parent stars, with ages similar to the age of the galaxy (about 10^{10} years). Most of the abundances in Table 1 were obtained from observations made with ground-based optical telescopes; however, abundances of carbon and some ions of nitrogen and neon were obtained from observations made with the *International Ultraviolet Explorer* satellite.

Pregalactic Chemical Abundances

Observational studies of planetary nebulae and supernovae, as well as stellar evolution models have shown that during their lifetimes stars enrich the interstellar medium with elements heavier than hydrogen. Therefore, to find the pregalactic chemical abundances, it is necessary to observe regions where star formation has not been important—that is, to study H II regions located in galaxies where the fraction of mass in the form of interstellar gas (M_{gas}/M_{total}) is large. From abundances determined for extragalactic H II regions, it has been found that generally the higher the values of M_{gas}/M_{total} in the galaxies where the H II regions are located, the smaller the abundances of carbon and heavier elements (*2, 3*). This can also be seen in Table 1, since the values of M_{gas}/M_{total} ratios for the LMC and the SMC are 0.12 and 0.42, respectively (*2*). This result implies that the Z value with which galaxies were formed is less than one-tenth of the value for the Orion Nebula or for the sun. Stars of masses smaller than about one solar mass (1 $M_\odot$) would not be expected to show variations in elements heavier than neon because nuclear reactions in their central regions do not produce or destroy them; therefore their abundances are indicative of the prestellar values. Measuring the abundances of elements heavier than neon relative to the abundance of hydrogen in very old stars has shown that such elements are underabundant compared to their abundances in the Orion Nebula and the sun by about two orders of magnitude. This is the case for sulfur and argon in halo planetary nebulae (*4*), and for iron in globular cluster stars (*5*). From the previous discussion, it follows that the pregalactic Z value is smaller than 0.002 and possibly smaller than 0.0002.

All of the well-observed H II regions in our galaxy and other galaxies show similar He/H abundance ratios, about 0.07 to 0.11, which correspond to Y values between 0.22 and 0.30. One of the main uncertainties in the determination of the He/H abundance ratio is due to the correction for the amount of neutral helium. Fortunately for objects with low O/H values, this correction is small and amounts to 0.01 (or less) of the Y value. From observations of H II regions with O/H values similar to that of the Orion Nebula, Y is found to be between 0.27 and 0.30, whereas from those H II regions with O/H values smaller by a factor of 2 or more than the O/H value of the Orion Nebula, Y is found to be between 0.22 and 0.25 (*2, 6–12*). These values indicate that a very general process occurred before the formation of galaxies that produced $X \approx 0.77$, $Y \approx 0.23$, and almost no heavy elements. This result together with the 3 K background radiation field and the recession of the galaxies (expansion of the universe) are the three strongest pieces of evidence in favor of Big Bang cosmology.

All of the observed H II regions have $Z \neq 0$ and belong to galaxies in which a substantial fraction of the mass has already been converted into stars. The Z value is probably due to enrichment of the interstellar medium by loss of mass from previously formed stars. For a very accurate pregalactic helium abundance (Y_p) to be determined, the observed Y values have to be extrapolated to the case with $Z = 0$. As a first approximation it can be assumed that, as a result of star formation, for any increase in X_{16} there is a corresponding increase in freshly made Y; that is, $\Delta Y/\Delta X_{16} = \beta$, where X_{16} is the oxygen abundance by mass and constitutes about 45 percent of Z (*9*). While one group of observers finds that for the Orion Nebula and for metal-poor H II regions $\Delta Y/\Delta X_{16} \sim 6 \pm 2$ (*2, 7–10*), another group finds that for metal-poor H II regions $\Delta Y/\Delta X_{16} \sim 3 \pm 8$ (*12*). Adopting different determinations of the initial mass function for the solar neighborhood (which gives the relative proportion at birth of stars of different masses) and different stellar evolution models has resulted in the prediction that $\Delta Y/\Delta X_{16}$ is between 2 and 7 (*13*).

Values of Y versus O/H presented in

Fig. 2 are based on observations of H II regions in spiral, irregular and blue compact galaxies. In this figure, we show the three best Y determinations that have been reported (*7–9*), with estimated standard deviation errors (1σ), together with additional very good data reported by others (*2, 11, 12*). We consider our Y values (*7–9*) to be better than those of the other three groups of observers (*2, 11, 12*) because the procedure used to estimate the amount of neutral helium inside the H II regions was based on an empirical ionization correction formula that included O^+ and O^{2+} and, in addition, was based on the condition that the dispersion of the He/H ratio would be minimum for a set of H II regions of a given galaxy. The Mexican group (*2, 7–9, 11*), which finds $\Delta Y/\Delta X_{16} \sim 6$, used a subset of the data in Fig. 2 to obtain $Y_p = 0.225$ by extrapolating to $Z = 0$. Kunth and Sargent (*12*) used another subset of the data in Fig. 2, and letting $\Delta Y/\Delta X_{16} = 0$, obtained $Y_p = 0.245$. An exhaustive discussion of the internal and external errors of these determinations has not been reported, but from unpublished work the error in Y_p is estimated to be about 0.01 at the 1σ level (*14*). Furthermore, it is not known if ΔY should be linear with ΔZ for values smaller than the minimum observed O/H ratio. The point is very important because $Y_p = 0.225$ is consistent with a standard (that is, the simplest) hot Big Bang model with only two families of two component light neutrinos ($m_\nu \ll 1$ million electron volts, where m_ν is the rest mass of the neutrinos), whereas $Y_p = 0.245$ is consistent with the standard model with three families of two component light neutrinos (*15*). If $Y_p < 0.23$, either the τ neutrino might be heavy ($m_\nu > 1$ MeV) or a nonstandard Big Bang model would have to be adopted.

To increase the accuracy of the Y_p determinations from H II region observations, we need (i) more accurate measurements of several He/H emission line ratios in a given H II region, (ii) measurements in several areas of a given H II region, (iii) better H II region ionization structure models, and (iv) observations of several H II regions in a galaxy with a well mixed interstellar medium.

Table 1. Chemical abundances expressed as log N(X) and scaled to log N(H) = 12.00.

Type	Element						References
	H	He	C	N	O	Ne	
			H II regions				
Orion	12.00	11.01	8.57	7.68	8.65	7.80	(*9*)
LMC	12.00	10.92	7.86	7.03	8.34	7.44	(*2, 7, 10*)
SMC	12.00	10.89	7.00	6.41	7.89	7.03	(*2, 8, 10*)
			Planetary nebulae				
$(He\text{-}N)_{rich}$	12.00	11.18	8.3–9.1	8.6	8.6	7.9	(*19*)
Intermediate	12.00	11.04	8.4–9.2	8.1	8.7	8.1	(*17*)
Halo	12.00	10.99	8.9	6.7–8.3	7.9	6.2–8.0	(*18, 40*)

Interstellar Medium Enrichment

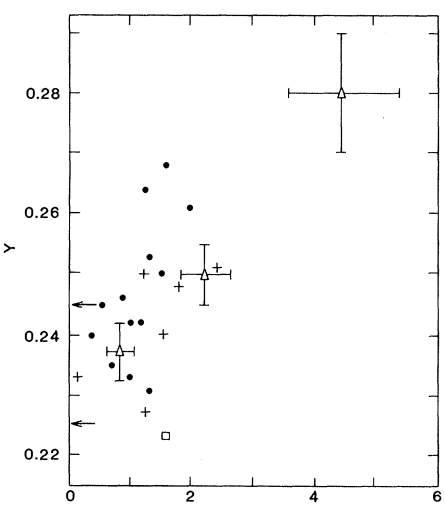
The previous discussion suggests that galaxies are formed with $X \approx 0.77$, $Y \approx 0.23$, and almost no heavy elements.

The theory of stellar evolution predicts that in the lifetime of the galaxy, only stars more massive than $\sim 1\ M_\odot$ have evolved and ejected material that has modified the chemical composition of the interstellar medium. The stars more massive than $\sim 8\ M_\odot$ produce supernovae that enrich the interstellar medium with oxygen and heavier elements such as neon, sulfur, argon, iron, and nickel. The intermediate mass stars produce planetary nebulae that enrich the interstellar medium with helium, nitrogen, and carbon; this can be seen by comparing the abundances of these elements in $(He\text{-}N)_{rich}$ and intermediate planetary nebulae of the solar neighborhood with the Orion Nebula values (Table 1).

The stellar evolution predictions for the chemical composition of planetary nebulae shells by Iben and his collaborators (*16*) have been compared with observations (*17–20*). For example, a comparison of theory with observations is shown in the plot of log C/O versus He/H in Fig. 3. In general, the agreement is good and indicates that intermediate mass stars produce large amounts of carbon and helium. Moreover, these stars are responsible for most of the carbon and nitrogen and a substantial amount of the helium enrichment of the interstellar medium (*13, 21*).

The location of NGC 6302 and N97 in Fig. 3 indicates that, in the stellar models, the ratio of the convective mixing length to the pressure scale height, α, has to be about 2. There are two other arguments in favor of this value of α; a smaller value of α would produce a C/H ratio larger than that observed in the solar neighborhood (*21*); a smaller value of α would be inconsistent with the decrease of the O/H ratio with increasing N/H found in $(He\text{-}N)_{rich}$ planetary nebulae (*18, 21, 22*).

Another result that can be derived from Fig. 3 is that the predicted mass for the observed chemical composition of NGC 6302 and N97 is around 8 $M_\odot$. This result is important because it indicates



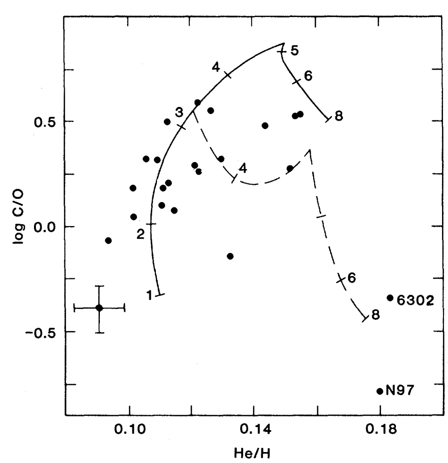


Fig. 2 (left). Graph of Y versus O/H for galactic and extragalactic H II regions. Crosses are the data from (*2*); the square corresponds to NGC 5471 in M101 (*11*), and the filled circles are the data from (*12*). For these points, the errors are not labeled to avoid crowding. Two possible values of the pregalactic helium abundance, Y_p, are marked by arrows. Fig. 3 (right). C/O versus He/H abundance ratios. From stellar evolution models, the predicted values for the planetary nebula shell are presented for two values of α, the ratio of the convective mixing length to the pressure scale height. The solid line is for $\alpha = 0$, and the broken line is for $\alpha = 2$. The numbers in the figure denote the mass of the progenitor stars. The filled circles are the observed values for planetary nebulae. Standard deviation errors (1σ) shown for one object are typical for the sample.

that at least some stars with initial masses around 8 $M_\odot$ do not become supernovae. However, before this result is accepted, other tests of the stellar evolution theory of intermediate mass stars should be made. An independent observation that supports this result is the detection, in galactic clusters, of white dwarfs that had progenitors with masses around 7 $M_\odot$ (*23*); white dwarfs and nuclei of planetary nebulae have masses smaller than 1.4 $M_\odot$, and therefore the excess mass was lost by stellar winds and planetary nebulae shell ejection.

In spiral galaxies, like ours, the M_{gas}/M_{total} ratio increases from the center outward (*24*), and therefore the amount of heavy elements relative to hydrogen should decrease from the center outward. This is indeed what is seen in our galaxy and other galaxies (*11, 25*). The quantitative study of these gradients is related to the problem of the chemical evolution of galaxies.

Chemical Evolution of Galaxies

To explain the observed abundances in the interstellar medium of our galaxy and other galaxies, it is necessary to construct models for the chemical evolution of galaxies (*2, 26, 27*); in these models the abundance at a given time in a given region of a galaxy depends on (i) the initial mass function, (ii) the birth rate of stars and the chemical composition of mass lost by the stars during their evolution, and (iii) the existence of large-scale mass flows, like infall from the halo or radial flows across the disk of a galaxy. Since these physical parameters are not generally known in galaxies, the observed abundances can be used as constraints of the galactic chemical evolution models.

In closed models, in which there are no mass flows, the total mass (gas plus stars) is conserved and abundances depend only on the fraction of the initial mass of the galaxy that has been converted into stars independently of the details of the birth rate function (*28*). This result holds as long as the time scale for gas consumption in the galaxy is much longer than the lifetime of the stars that produce predominantly a given element; this is the instantaneous recycling approximation (IRA). Such a simple model (closed plus IRA) results in abundances that increase logarithmically with the gas fraction M_{gas}/M_{total}. All information about the contribution of individual stars is condensed in the "yield," which is the mass of the newly formed elements that stars eject to the interstellar medium in units of the mass that is never returned to it (mass permanently trapped in "remnants"). The yield is defined for a stellar generation and weighted by the initial mass function, and it depends only on the assumed stellar mass loss; that is, the amount and chemical composition. In a diagram of abundance versus the logarithm of the gas fraction, such as Fig. 4, the yield of a simple model corresponds to the slope of the galactic evolution line. Furthermore, the simple model predicts that if the yield is the same for all galaxies, they should lie on a straight line in Fig. 4.

Theoretical estimates of the yield of "heavy elements" (carbon and heavier) give values between 0.003 and 0.01 (*2, 26, 27*). The same range is obtained when simple models with a yield independent of metallicity are used to explain observed abundances and gas fractions in galactic and extragalactic H II regions; this is the "great success of the simple model." But although the order of magnitude of the heavy elements predicted by this model is correct, there are galaxies with the same mass fraction and with different values for abundances of heavy elements (represented in Fig. 4 by the abundance of oxygen with respect to hydrogen). Therefore, unless we accept an arbitrary yield that varies from galaxy to galaxy, the simple model has to be dropped. Models with a yield independent of metallicity and infall of gas with primordial composition were proposed in order to solve this problem (*29*). Models with infall can be characterized by the parameter γ, the ratio of accretion to the rate of star formation. For example, if the model is closed, $\gamma = 0$, and if infall (or accretion) exactly balances star formation so that the mass in the form of gas is constant, then $\gamma = 1$. These IRA models with a yield independent of metallicity are shown in Fig. 4 as dashed lines. In the case of the extreme infall model, $\gamma = 1$, the oxygen abundance

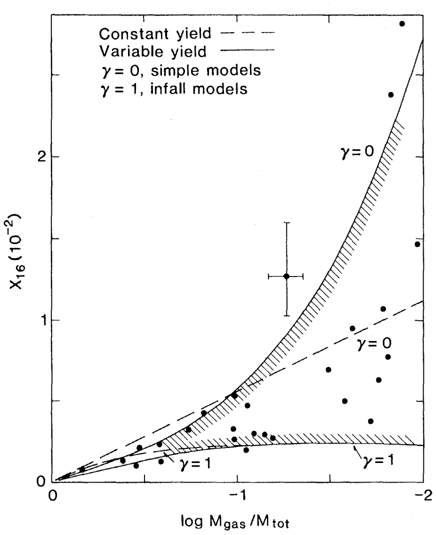


Fig. 4. Effects of accretion of primordial gas and of yields increasing with metallicity on the relation between the metallicity, represented by X_{16}, and the gas fraction M_{gas}/M_{total}. Filled circles represent observations of galactic and extragalactic H II regions (*3*). Also presented are models with a constant yield (p_{16} = 0.0023) and models with a yield increasing with metallicity (p_{16} = 0.0009 + 0.6 X_{16}). The infall models coincide for log M_{gas}/M_{total} smaller than −1. Standard deviation errors (1σ) presented for one object are typical for the sample; note that the vertical error is proportional to the oxygen abundance.

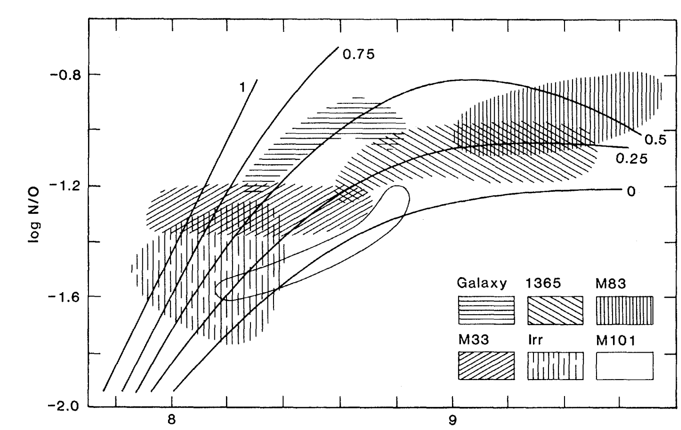


Fig. 5. Graph of N/O versus O/H, assuming secondary production of nitrogen and a yield increasing with metallicity. Observations of galactic and extragalactic H II regions are schematically represented by shaded areas. As in Fig. 4, curves are labeled by γ. Accretion increases with γ, defining the region between the curves $\gamma = 0$ and $\gamma = 1$ as the permitted region. Lines of constant age are nearly horizontal in this diagram, with age increasing with N/O. Lines of constant gas fraction are nearly vertical, with M_{gas}/M_{total} decreasing with O/H. Galaxies identified in the boxes are our own galaxy; spiral galaxies NGC 1365, M33, M83, and M101; and a group of irregular and blue compact galaxies designated Irr.

quickly saturates to the value given by the yield (a horizontal line in Fig. 4) because an equilibrium is reached in which dilution due to accretion of primordial material does not allow abundances to increase. If galaxies diminish their gas fraction as the result of star formation, they should lie between the simple model ($\gamma = 0$) and the extreme infall model ($\gamma = 1$) in Fig. 4. Thus, by assuming different values of the ratio of infall to star formation rates, it would be possible to explain a range of O/H values at a given value of M_{gas}/M_{total}.

However, as can be seen in Fig. 4, this permitted region does not agree with observations, since there is a much larger observed range of X_{16} than constant-yield models predict for small gas fractions. To explain this fact it is necessary for the yield to increase with the abundance of heavy elements. In Fig. 4, the variable yield presented (p) is an expression of the form $p = p_0 + aZ$ (*3*). This in turn implies that galaxies with low O/H are more efficient in the formation of low mass stars (less than 1 $M_\odot$). These models are shown as continuous lines in Fig. 4. In this case the simple model ($\gamma = 0$) curves upward in contrast with closed models with a yield independent of metallicity, which are straight lines. On the other hand, extreme infall models ($\gamma = 1$) are very similar in both cases. The high metallicity objects are better explained with a yield increasing with metallicity.

In the simple model, there is a clear distinction between primary and secondary elements, where primary elements are those that are directly synthesized from hydrogen and helium, and secondary elements are those synthesized from heavy elements that were already present in the star when it was formed. For example, in an N/O versus O/H diagram (Fig. 5), primary nitrogen production corresponds to a horizontal line (N/O is constant), whereas secondary nitrogen production corresponds to a 45° line (N/O $\propto$ O/H) (*30*). It is clear from Fig. 5 that there is no unique correlation of either type between these two abundance ratios. The spread in Fig. 5 makes it more difficult to find out whether nitrogen is a primary or secondary element. The assumption of IRA breaks down for nitrogen because the average star that produces nitrogen has a smaller initial mass (around 3 $M_\odot$) than the average star that produces oxygen (around 25 $M_\odot$). Since the lifetime of a star is a strongly decreasing function of its mass, this means that there is a delay in the injection of nitrogen into the interstellar medium with respect to oxygen. Moreover, there are usually no strong gradients of N/O across disks of spiral galaxies, whereas O/H usually shows well-defined gradients. This led to the suggestion that nitrogen is primary, so that each galaxy is represented by a horizontal line in Fig. 5, and that there is a vertical spread in this figure caused by the effect of delays in the ejection of nitrogen (*31*).

Since the oxygen content in a simple model (closed plus IRA) is determined by the gas fraction, the spread in Fig. 5 requires a second physical parameter whose variation from galaxy to galaxy explains the dispersion in N/O at a given O/H. Such a second parameter is also needed to explain the variation of O/H, at N/O nearly constant, from place to place across the disks of spiral galaxies. It has been proposed that the spread in Fig. 5 is due to variations in (i) the initial mass function, (ii) the star formation rate, (iii) the history of the star formation rate, and (iv) the accretion rate. It has been found that Fig. 5 is best explained by models with accretion of uncontaminated gas, secondary nitrogen, and a yield increasing with O/H (*32*). Each of these three assumptions is necessary: primary nitrogen would fill up the forbidden region at high O/H and low N/O; closed models result in a too narrow diagram; and, finally, a constant yield would not predict galaxies with high O/H at all. In these models a given galaxy has the same age across its disk; O/H gradients can exist under small differences in the local gas fractions; and much smaller N/O gradients are explained by gradients in the ratio of infall to the star formation rate. Under these assumptions, the stars that produce nitrogen have masses of about 1.3 $M_\odot$ and larger.

If the explanation of Fig. 5 (*32*) is correct, most of the nitrogen of the interstellar medium is of secondary origin and has been produced by intermediate mass stars, progenitors of planetary nebulae. This is contrary to the suggestion that novae inject most of the nitrogen present in the interstellar medium (*33*), because the nitrogen produced by novae is of primary origin.

The models of chemical evolution of galaxies require infall of gas to be important in order to explain the observations presented in Figs. 4 and 5. Moreover, the chemical history of the solar neighborhood requires also that the rate of infall of gaseous material from outside the disk has been approximately equal to half the star formation rate as indicated by the age-metallicity relationship in stars (*34*) and by isotopic abundance ratios (*35*). The existence of a substantial infall rate in galaxies is a subject of great controversy, although recent observations have shown that it exists in both irregular and spiral galaxies (*36*).

A study of a sample of irregular galaxies shows that there is no relation between the mass gas fraction and the O/H abundance ratio. This also indicates that even if the Z value of a given galaxy is consistent with the closed model, the galaxies as a group are not; in addition, the observations are best explained by an infall model (*37*). Moreover, the constancy of the star formation rate determined for isolated irregular galaxies cannot be explained by the simple model; an alternative explanation could be steady accretion from a bound gaseous halo or other types of accretion. Other explanations are also possible (*38*).

The suggestion that formation of low mass stars is more efficient in low-metallicity environments is also controversial, since a simple gravitational instability argument would suggest exactly the opposite. Other theories, however, predict that the relative number of low mass stars decreases with metallicity (*39*). A yield increasing with metallicity, as proposed, would produce a positive gradient in the observed mass-to-luminosity ratio in spiral galaxies and would be consistent with a substantial amount of hidden mass in the form of low mass stars in galactic halos.

References and Notes

1. D. E. Osterbrock, *Astrophysics of Gaseous Nebulae* (Freeman, San Francisco, 1974).
2. J. Lequeux, M. Peimbert, J. F. Rayo, A. Serrano, S. Torres-Peimbert, *Astron. Astrophys.* **80**, 155 (1979).
3. M. Peimbert and A. Serrano, *Mon. Not. R. Astron. Soc.* **198**, 563 (1982).
4. T. Barker, *Astrophys. J.* **237**, 482 (1980); *ibid.* **270**, 641 (1983).
5. H. L. Helfer, G. Wallerstein, J. L. Greenstein, *ibid.* **129**, 700 (1959); G. Wallerstein and H. L. Helfer, *Astron. J.* **71**, 350 (1966).
6. M. Peimbert, *Annu. Rev. Astron. Astrophys.* **13**, 113 (1975); B. E. J. Pagel, *Philos. Trans. R. Soc. London Ser. A* **307**, 19 (1982); T. D. Kinman and K. Davidson, *ibid.*, p. 37. A thorough discussion on the pregalactic helium abundance is given in *Primordial Helium*, P. A. Shaver, D. Kunth, K. Kjar, Eds. (European Southern Observatory, Garching, 1983).
7. M. Peimbert and S. Torres-Peimbert, *Astrophys. J.* **193**, 327 (1974).
8. ———, *ibid.* **203**, 581 (1976).
9. ———, *Mon. Not. R. Astron. Soc.* **179**, 217 (1977); S. Torres-Peimbert, M. Peimbert, E. Daltabuit, *Astrophys. J.* **238**, 133 (1980).
10. R. J. Dufour, G. A. Shields, R. J. Talbot, Jr., *Astrophys. J.* **252**, 461 (1982); R. J. Dufour, *Int. Astron. Union Symp.*, in press.
11. J. F. Rayo, M. Peimbert, S. Torres-Peimbert, *Astrophys. J.* **255**, 1 (1982).
12. D. Kunth and W. L. W. Sargent, *ibid.* **273**, 81 (1983).
13. A. Serrano and M. Peimbert, *Rev. Mex. Astron. Astrofis.* **5**, 109 (1981); A. Maeder, *Astron. Astrophys.* **101**, 385 (1981); C. Chiosi and F. M. Matteucci, *ibid.* **105**, 140 (1982).
14. K. Davidson and T. D. Kinman, *Astrophys. J.*, in press; M. Peimbert and S. Torres-Peimbert, in preparation.
15. J. Yang, M. S. Turner, G. Steigman, D. N. Schramm, K. A. Olive, *Astrophys. J.*, in press.

16. A. Renzini and M. Voli, *Astron. Astrophys.* **94**, 175 (1981); I. Iben, Jr., and A. Renzini, *Annu. Rev. Astron. Astrophys.* **21**, 271 (1983).
17. L. H. Aller, *Int. Astron. Union Symp.* **103**, 1 (1983); J. B. Kaler, *ibid.*, p. 245; M. Peimbert, in *Physical Processes in Red Giants* (Reidel, Dordrecht, Netherlands, 1981), p. 409.
18. S. Torres-Peimbert, in *Stellar Nucleosynthesis* (Reidel, Dordrecht, Netherlands, in press).
19. M. Peimbert, *Int. Astron. Union Symp.* **76**, 215 (1978); _____ and S. Torres-Peimbert, *ibid.* **103**, 233 (1983).
20. M. J. Barlow, S. Adams, M. J. Seaton, A. J. Willis, A. R. Walker, *ibid.*, p. 538.
21. M. Peimbert, *Int. Astron. Union Symp.*, in press.
22. A. Renzini, in *Stellar Nucleosynthesis* (Reidel, Dordrecht, Netherlands, in press).
23. D. Reimers and D. Koester, *Astron. Astrophys.* **116**, 341 (1982).
24. M. Peimbert, *Int. Astron. Union Coll.* **45**, 149 (1977).
25. L. Searle, *Astrophys. J.* **168**, 327 (1971); M. Peimbert, *Int. Astron. Union Symp.* **84**, 307 (1979); P. A. Shaver, R. X. McGee, L. M. Newton, A. C. Danks, S. R. Pottasch, *Mon. Not. R. Astron. Soc.* **204**, 53 (1983); H. E. Bond, in preparation.
26. R. J. Talbot, Jr., and W. D. Arnett, *Astrophys. J.* **186**, 51 (1973).
27. B. M. Tinsley, *Fundam. Cosmic Phys.* **5**, 287 (1980); B. E. J. Pagel and M. G. Edmunds, *Annu. Rev. Astron. Astrophys.* **19**, 77 (1981); B. A. Twarog, *Int. Astron. Union Symp.*, in press.
28. L. Searle and W. L. W. Sargent, *Astrophys. J.* **173**, 25 (1972).
29. R. B. Larson, *Nature (London) Phys. Sci.* **236**, 21 (1972).
30. R. J. Talbot, Jr., and W. D. Arnett, *Astrophys. J.* **190**, 605 (1974).
31. M. G. Edmunds and B. E. J. Pagel, *Mon. Not. R. Astron. Soc.* **185**, 77 (1978).
32. A. Serrano and M. Peimbert, *Rev. Mex. Astron. Astrofis.* **8**, 117 (1983).
33. R. E. Williams, *Astrophys. J.* **261**, L77 (1982).
34. B. A. Twarog, *ibid.* **242**, 242 (1980).
35. M. Tosi, *ibid.* **254**, 699 (1982).
36. I. F. Mirabel, *ibid.* **256**, 112 (1982); J. M. Van der Hulst *et al.*, *ibid.* **264**, L37 (1982); I. F. Mirabel and R. Morras, *ibid.*, in press.
37. D. A. Hunter, J. S. Gallagher, D. Rautenkranz, *Astrophys. J. Suppl. Ser.* **49**, 53 (1982).
38. J. S. Gallagher, D. A. Hunter, A. V. Tutukov, *Astrophys. J.*, in press.
39. J. I. Silk, *ibid.* **214**, 718 (1977); V. C. Reddish, *Stellar Formation* (Pergamon, Oxford, 1978), p. 208.
40. S. Adams, M. J. Seaton, I. D. Howarth, M. Aurriere, J. R. Walsch, *Mon. Not. R. Astron. Soc.*, in press.

Evolution of Proteolytic Enzymes

Hans Neurath

Summary. Proteolytic enzymes have many physiological functions, ranging from generalized protein digestion to more specific regulated processes such as the activation of zymogens, blood coagulation and the lysis of fibrin clots, the release of hormones and pharmacologically active peptides from precursor proteins, and the transport of secretory proteins across membranes. They are present in all forms of living organisms. Comparisons of amino acid sequences, three-dimensional structures, and enzymatic reaction mechanisms of proteases indicate that there are distinct families of these proteins. Changes in molecular structure and function have accompanied the evolution of proteolytic enzymes and their inhibitors, each having relatively simple roles in primitive organisms and more diverse and more complex functions in higher organisms.

Complex molecules, such as the proteins, are the products of evolutionary processes in which, according to Jacob (*1*), "constraints, at every level, specify the rules of the game and define what is possible with these systems; and historical circumstances determine the actual course of events." Proteolytic enzymes, or proteases, are enzymes that catalyze the cleavage of peptide bonds in other proteins. They are presumed to have arisen in the earliest phases of biological evolution since even the most primitive organisms must have required them for digestion and for the metabolism of their own proteins. In fact, present digestive proteases can be shown to have a common ancestry with those of microbial origin, and they are considered to have originated some billion years ago (*2*). If one accepts the generally held view that biological evolution proceeded from the relatively simple to the more complex levels of organization, one would expect that in the course of evolutionary development, the very proteases that originally served purely digestive functions have adapted to the more specific and complex tasks of physiological regulation (*3*). In this process, they acquired a higher degree of specialization by restricting their action to a select number of peptide bonds located at specific sites in specific protein substrates. Such limited proteolysis does not destroy the protein substrate altogether but modifies its properties and physiological role. Many biological processes are regulated by this type of protease action, for example, blood coagulation and fibrinolysis, the release of protein hormones from precursor molecules, the transport of secretory proteins across membranes, the assembly of macromolecular structures such as collagen fibers or certain viruses, fertilization, and the control of proteolytic digestion itself (*3*). In this article, I examine the changes in molecular structure that have accompanied the evolution of proteolytic enzymes from primitive organisms, serving relatively primitive functions, to the diverse and more complex functions that they fulfill in higher organisms including man.

Proteolytic enzymes are not only a physiological necessity but also a potential hazard, since, if uncontrolled, they can destroy the protein components of cells and tissues. Hence a discussion of protease evolution necessitates a consideration of the mechanisms whereby the action of proteolytic enzymes themselves is regulated. Two principal regulatory mechanisms, serving the same end, have been devised by nature: (i) activation of inactive protease precursors (zymogens) by limited proteolysis (*4*) and (ii) inactivation of proteases by forming complexes with protein inhibitors (*5*). The physiological importance of both mechanisms has been demonstrated in certain disease states that are related to deficiencies of functional zymogens or protease inhibitors, respectively. For instance, several types of familial hemophilias are due to deficiencies of one or the other of plasma proteases that normally activate zymogens in the blood coagulation cascade (*6*). In another case, familial emphysema, a genetic mutation in α_1-inhibitor impairs its secretion from the liver, resulting in abnormally low plasma concentrations and consequently in destruction of connective tissue of the pulmonary alveoli by leukocyte elastase (*7*).

Hans Neurath is Professor Emeritus in the Department of Biochemistry, University of Washington, Seattle 98195. This article is based on the text of the Margaret Hastings Distinguished Lecture, which was presented at the Scripps Clinic and Research Foundation, La Jolla, California, 22 April 1983.