STAR FORMATION IN DISKS

Richard B. Larson
Yale University Observatory

The basic characteristics of the stellar populations in galaxies are determined by the history of star formation and by the initial mass spectrum with which the stars are formed. Many attempts have been made to model the history of star formation in galaxies using various quasi-theoretical descriptions of star formation, but star formation remains a very poorly understood process and no theoretical understanding with any real predictive power has yet been attained. However, recent work suggests the importance for star formation of some mechanisms that have so far not been very extensively studied, and I shall mention some of these possible processes here, undeveloped though these ideas still are at present.

First, I shall review briefly the observational evidence concerning the history of star formation in galaxies of different Hubble type. Most of the information we have for galaxies other than our own comes from their colors, which can be compared with the predictions of models constructed with various assumed histories of star formation. Ever since the work of Tinsley (1968), it has been clear that most of the observed variation of color with Hubble type can be understood in terms of differing time scales for star formation, the reddest galaxies (E and S0) having formed the bulk of their stars at a relatively early stage, while the much bluer Sc galaxies have formed their stars much more gradually and at a rate that has remained nearly constant up to the present time. Recent studies of normal elliptical galaxies, reviewed by Faber in this volume, confirm that they have no detectable present star formation. At the other extreme, studies of the age distribution of stars in the solar neighborhood of our galaxy, which has type $\sim$ Sbc, confirm that the star formation rate there has remained nearly constant, showing at most a mild decrease with time (Twarog 1980). Information about the history of star formation in the disk of our Galaxy can also be drawn from the

Richard M. West (ed.), Highlights of Astronomy, Vol. 6, 191–198.

https://doi.org/10.1017/S1539299600005128 Published online by Cambridge University Press

distribution of ages of old open clusters: the massive clusters in the outer part of the disk have apparently not suffered severely from disruption, yet they show an age cutoff at about half the age of the globular clusters, suggesting that the outer disk of our Galaxy is significantly younger than the halo (Janes and Adler 1982).

While the colors of most galaxies can thus easily be accounted for, there remain a considerable number of very late-type galaxies, including many irregular and peculiar galaxies, whose colors are too blue to be explained even by a constant rate of star formation. A number of factors that may be relevant for interpreting the colors of such systems were discussed by Larson and Tinsley (1978). Many irregular galaxies are probably metal-poor, like the Magellanic Clouds, and this will make their colors somewhat bluer. Some late-type systems may be significantly younger than the Hubble time; for example, there is evidence that the bulk of the stars in the Large Magellanic Cloud may have formed as recently as $\sim$ 4 Gyr ago (Butcher 1977, Stryker 1981). Recent bursts of star formation in old systems can also lead to very blue colors and to a large scatter of colors in a two-color diagram; this explanation appears to be particularly relevant for interacting galaxies, in which bursts of star formation can apparently be triggered by the interactions. Finally, as an interesting possibility to which not much attention has yet been given, variations in the slope of the initial mass function (IMF) can have a large effect on the colors of blue galaxies; the extreme blueness of very late-type galaxies, and perhaps some of the increase in blueness with Hubble type, could be explained if later-type galaxies tend to have a higher proportion of massive stars. There is some evidence that this may be the case, as discussed in this volume by van den Bergh. Each of the effects that have been mentioned could plausibly make the colors of galaxies bluer by $\sim$ 0.1 in B-V, and some combination of effects could account for the colors of the bluest galaxies.

In any case, it seems clear that the dominant effect governing the colors of galaxies is the increasing overall youth of the stellar content of galaxies of later Hubble type, indicating a longer time scale for the transformation of gas into stars. The relative incompleteness of star formation in later-type galaxies is clear from their higher gas contents, which in some very blue dwarf galaxies appear to dominate the total mass, and from the general metal-poorness of late-type irregular and dwarf galaxies. The primary relation between star formation and galactic morphology may actually be a dependence of star formation time scale on galactic mass, since galaxies of later Hubble type tend to be fainter and less massive than those of earlier Hubble type (de Vaucouleurs 1977). A recent paper by Tully, Mould, and Aaronson (1982) shows that the colors of spiral galaxies are indeed closely correlated with their masses (as measured by rotational velocities) and with their luminosities; in fact, once the mass dependence is accounted for, there is only a marginal residual correlation of color with Hubble type for a given mass.

Thus the primary fact to be explained seems to be that the more massive galaxies turn their gas into stars faster than the less massive galaxies. As a secondary effect, star formation may also go faster in galaxies that are more centrally condensed. Another possible trend, for which more observational evidence is needed, is that galaxies of later Hubble type or smaller mass may form a higher proportion of very massive stars. These trends are not readily accounted for by the traditional approaches attempting to relate star formation to purely local quantities such as the gas density; rather, the data seem to point toward a relation between star formation and the large-scale dynamics of galaxies, and in particular to a dependence on galactic rotation as an important parameter. Also, as is shown by the evidence for bursts in interacting galaxies, the star formation rate is very sensitive to large-scale disturbances in the dynamics of galaxies, such as those produced by interactions between galaxies.

Suggestions that star formation depends on dynamical processes are, of course, not new; theories in which star formation is induced by shocks, etc., have been widely discussed in recent years. In particular, this idea is an essential part of the theories of self-propagating star formation that have been applied to explain many properties of spiral and irregular galaxies (Gerola, Seiden, and Schulman 1980). Also, it is not new to suggest that star formation depends on galactic rotation, since this is implied by density-wave theories of spiral structure and star formation. However, these theories all have ad hoc features or undetermined parameters, and they have so far not yielded any very clear a priori predictions about rates of star formation. In the theory of self-propagating star formation, for example, the initiation of star formation is ad hoc, while density-wave theory has undetermined parameters such as the wave amplitude and pattern speed, for which empirical determinations have so far yielded at best ambiguous results. In any case, neither of these theories is likely to apply to all galaxies, and it would seem that the explanation of such a universal characteristic as the color-mass relation among galaxies should be sought in a more general mechanism of star formation. It is worth noting that even elliptical galaxies show a color-mass relation, and although it is generally attributed to metallicity rather than age variations, a plausible explanation is that the efficiency of star formation is an increasing function of the mass of the subsystems from which elliptical galaxies are formed (Tinsley and Larson 1979). These subsystems were probably more irregular than present-day spirals, and probably experienced interactions that triggered bursts of star formation. Both normal star formation and bursts in interacting systems may be qualitatively understandable in terms of a recently revived theory of spiral structure whereby spiral arms represent transient gravitational instabilities of disks, sometimes triggered by disturbances such as interactions between galaxies.

There is little doubt that star formation involves self-gravitational phenomena on large scales, much larger than those of individual protostars or protoclusters. Star formation occurs in dense molecular clouds which are generally parts of larger, less dense complexes; these "giant molecular cloud" complexes have typical dimensions of $\sim 10^2$pc and masses of 10^5 - 10^6 $M_\odot$. These may be grouped in turn into still larger complexes of molecular and atomic gas having dimensions of a kiloparsec or more and masses up to 10^7 $M_\odot$ (Elmegreen and Elmegreen 1982), which are probably the basic constituents of spiral arms. Except perhaps for the very largest scales, molecular clouds and complexes of all sizes are gravitationally bound objects, since in them there is approximate virial balance between the gravitational potential energy and the kinetic energy of internal motions (Larson 1981). This implies that gravity must play an important role in the dynamics and evolution of molecular clouds and cloud complexes, and almost certainly in their formation as well.

There are at least two ways in which gravity can play a role in the formation of molecular clouds out of more diffuse gas in the disk of a spiral galaxy: (1) gravitational instabilities in the gas layer can lead to clumping, in the form of transient spiral density enhancements; and (2) gravitational accretion of matter by a small cloud or density enhancement in a shearing gas layer can cause it to grow into a large molecular cloud before star formation begins to disrupt the cloud. Probably these two effects are not completely separable, and both occur as part of the same phenomenon; large-scale aspects may be describable in terms of a gravitational instability, whereas smaller-scale aspects and the properties of individual clouds may be determined by gravitational accretion processes.

Gravitational instabilities in shearing gas layers were originally suggested as a theory of spiral structure by Goldreich and Lynden-Bell (1965). This idea has been revived and extended in recent work on the stability of disk systems; most of this work still awaits publication, but a summary has been given by Toomre (1981) in a discussion of the "swing amplifier" mechanism. As was originally found by Goldreich and Lynden-Bell, a perturbation in a marginally stable, differentially rotating gas disk characteristically results in a rapidly growing, shearing (i.e. winding up) spiral density enhancement. A point emphasized by

Toomre is that the shear itself contributes strongly to the growth of the perturbation, and numerical simulations illustrate that under suitable conditions a shearing density enhancement can be amplified by a very large factor. The condition required for the swing amplifier to work is similar in form to the usual criterion for the growth of axisymmetric instabilities in a disk, and for a pure gas disk is approximately

$$\mu > 0.30\ c\kappa/G$$

where μ is the surface mass density, c is the velocity dispersion of the gas, and κ is the epicyclic frequency. Because of the amplifier effect, shearing modes can grow even in a disk that is stable to axisymmetric modes. The initial density perturbation can be of external origin, for example a tidal interaction with a companion galaxy, or it can result from random density fluctuations in a marginally stable disk. The amplifier effect works particularly well for an initially leading spiral density enhancement that swings around into a trailing one. This situation might in fact occur naturally if the initial density enhancement is created when differential rotation in a clumpy medium brings one clump past another one with a slightly larger orbit; then when the inner clump is approaching the outer one, the resulting density enhancement initially has a leading orientation.

The criterion for swing amplification involves a critical surface density of gas that is proportional to the velocity dispersion of the gas and to the epicyclic frequency. If the velocity dispersion is approximately constant at $\sim$ 10 km/s, as seems to be the case rather generally in spiral galaxies, the critical surface density is proportional to the epicyclic frequency κ, which for a flat rotation curve varies inversely with radius. Thus there may be a radially decreasing threshold surface density μ_c such that for $\mu > \mu_c$, strong instabilities, large density enhancements, and rapid star formation occur, whereas for $\mu < \mu_c$, little happens unless the disk is perturbed, for example by a tidal encounter; in this case there can result transient, although possibly spectacular, spiral structure and star formation. Toomre (1981) suggests that M51, with its nearby companion and spectacular spiral pattern, may be an example of such an occurrence.

If most galaxies have depleted their gas to the point where $\mu \sim \mu_c$ and star formation is proceeding at only a modest pace, the star formation rate may be particularly sensitive to disturbances such as tidal encounters, and this could account for the occurrence of bursts of star formation in many interacting galaxies. The sensitivity of the star formation rate to interactions might also explain why star formation has generally gone farther to completion in galaxies that are in groups and clusters; cluster galaxies generally have lower gas contents than field galaxies, and clusters contain a higher proportion of S0 galaxies than the field, the *fraction of S0's increasing steadily with increasing ambient* density of galaxies (Dressler 1980), as might be expected if star formation is enhanced by interactions.

While the occurrence of star formation may thus depend on large-scale gravitational instabilities, the properties of individual molecular clouds may be determined by gravitational accretion processes. It can be shown, with plausible assumptions, that a small clump or seed mass in a shearing gas layer will accrete matter and grow to attain the mass of a "giant molecular cloud" in an interestingly short time; after this point, the time scale for internal evolution becomes shorter than the growth time, and star formation and its effects may destroy the cloud. The accretion rate and the maximum cloud mass depend on the surface density μ and the shear rate A of the gas layer, and for typical solar neighborhood values the mass attained by a cloud before it is disrupted is predicted to be around 10^5 $M_\odot$, typical of "giant molecular clouds." The growth time is of the order of 10^7 years, consistent with the inferred lifetimes of such objects, and the typical velocities with which matter is accreted are several km/s, comparable with the observed internal velocities in molecular clouds. Thus, gravitational accretion may account for some of the basic characteristics of molecular clouds, such as their masses and internal turbulent velocites. It may even be possible to account in this way for the correlation between turbulent velocity and cloud size found by Larson (1981).

The overall rate of star formation will presumably depend not only on the growth rate of molecular clouds but also on factors such as the rate of production of suitable seed clouds. Whatever the details, it seems likely that galactic shear plays a role and that an important time scale is the shear time scale A^{-1}; in this case the two main parameters governing the formation and evolution of molecular clouds are the surface density of gas μ and the shear rate A. The rates of the relevant processes, and hence the overall rate of star formation, should increase with both μ and A; thus galaxies or parts of galaxies that have higher surface density and/or higher shear should form stars on a shorter time scale than regions with lower μ or A. A dependence of the star formation time scale on μ would probably be predicted qualitatively by almost any theory of star formation, but the predicted dependence on shear rate A is a more unique and perhaps more fundamental prediction of gravitational instability and accretion theories. Instead of inhibiting instabilities, as was once thought, differential rotation or shear can actually strongly enhance the growth of gravitational instabilities via the swing amplifier mechanism. As a result, it may be possible to understand the dependence of star formation time scale on galactic properties in terms of a dependence primarily on the shear rate A. Galaxies of larger mass have larger rotational velocities and hence larger values of A at each radius, so we might expect the time scale for star formation to decrease with increasing galactic mass. For a given mass, the shear rate increases with increasing central concentration of the mass distribution, so we would expect that galaxies with greater central concentration (e.g., a larger bulge/disk ratio) should form stars faster than less centrally condensed galaxies. Finally, for galaxies whose rotation curves are approximately flat, the shear rate A increases inwards, so that star formation should have proceeded farther to completion in the inner parts of such galaxies than in the outer parts. All of these predicted trends agree qualitatively with observed trends.

As a final very speculative comment, the dependence of molecular cloud properties on galactic parameters predicted by the gravitational accretion mechanism of cloud formation may also imply a dependence of the stellar IMF on galactic properties. There is evidence that the IMF depends on the masses of the molecular clouds in which stars form, in the sense that more massive and condensed clouds form more massive stars, and the most massive stars form only in the most massive clouds (Larson 1982). The gravitational accretion theory predicts, with plausible assumptions, that the characteristic cloud mass varies as $\mu^3 A^{-4}$, so that if μ does not vary greatly between galaxies, those with less shear should form larger clouds. This is basically because in galaxies with less shear, the clouds or cloud complexes have more time to build up very large masses. Possible radial variations in cloud mass within galaxies are harder to predict because the effects of radial variations in μ and A tend to cancel, but it seems likely that the dependence on A will usually dominate, in which case cloud masses should be larger in the outer parts of spiral galaxies. Thus late-type irregular galaxies and the outermost parts of giant spirals may share the common characteristics that the shear is small and the time scale for star formation is long, but the clouds attain very large masses and form a relatively high proportion of very massive stars. All of these qualitative predictions seem consistent with present knowledge; in particular, it appears that small irregular galaxies such as the LMC and the outermost parts of large spirals like M101 do contain very massive gas complexes and extremely luminous HII regions, while earlier-type galaxies generally contain much less prominent HII regions and therefore presumably fewer very massive stars, as expected from the above very speculative argument. Whether such effects can be confirmed and their generality extended by further work will be a particularly interesting topic for further study.

REFERENCES

Butcher, H. R.: 1977, Astrophys. J. 216, 372.
de Vaucouleurs, G.: 1977, in The Evolution of Galaxies and Stellar Populations, ed. B. M. Tinsley and R. B. Larson, Yale University Observatory, New Haven, p. 43.
Dressler, A.: 1980, Astrophys, J. 236, 351.
Elmegreen, B. G., and Elmegreen D. M.: 1982, Mon. Not. Roy. Astron. Soc., in press.
Gerola, H., Seiden, P. E., and Schulman, L. S.: 1980, Astrophys. J. 242, 517.
Goldreich, P., and Lynden-Bell, D.: 1965, Mon. Not. Roy. Astron. Soc. 130, 125.
Janes, K., and Adler, D.: 1982, Astrophys. J. Suppl. 49, 425.
Larson, R. B.: 1981, Mon. Not. Roy. Astron. Soc. 194, 809.
Larson, R. B.: 1982, Mon. Not. Roy. Astron. Soc. 200, 159.
Larson, R. B., and Tinsley, B. M.: 1978, Astrophys. J. 219, 46.
Stryker, L. L.: 1981, Ph.D. thesis, Yale University.
Tinsley, B. M.: 1968, Astrophys. J. 151, 547.
Tinsley, B. M., and Larson, R. B.: 1979, Mon. Not. Roy. Astron. Soc. 186, 503.
Toomre, A.: 1981, in The Structure and Evolution of Normal Galaxies, ed. S. M. Fall and D. Lynden-Bell, Cambridge University Press, p. 111.
Tully, R. B., Mould, J. R., and Aaronson, M.: 1982, Astrophys. J. 257, 527.
Twarog, B. A.: 1980, Astrophys. J. 242, 242.