GALACTIC EVOLUTION

RICHARD B. LARSON Yale University

ABSTRACT Observational evidence and theoretical notions concerning the evolution of the halo and disk components of our Galaxy are reviewed, and the implications of some recent results are discussed. Data on the ages and other properties of globular clusters suggest that the formation of the galactic halo was a prolonged and chaotic process, possibly involving the merging of galaxylike subsystems. There appears to be approximate chronological and chemical continuity between the halo and the disk, which suggests that disk formation took up where halo formation left off some 10 to 12 Gyr ago. Theoretical considerations and direct evidence concerning the history of star formation in the local disk and in typical Sc galaxies are both consistent with a simple model in which the star formation rate is approximately proportional to the gas content. With current data, the variation of the star formation rate (SFR) in such a model is moderate, and the ratio of past average to present SFR is only about 2. Models with a rapidly declining SFR and a strongly bimodal IMF are probably excluded by current constraints on the amount of unseen mass in the local disk. Gas replenishment by infall appears to be a minor effect at present, although it was almost certainly much more important at early times.

1. INTRODUCTION

A major motivation for studies of stellar evolution is to gain a better understanding of how galaxies like ours have evolved and how they may have formed. We cannot yet speak of a truly predictive "theory" of galactic evolution, although there are ideas about possibly important processes and schematic models to illustrate how they might operate; instead, the only real way to learn about how our Galaxy evolved is to apply our knowledge of stellar evolution to try to reconstruct the Galaxy's past history from the fossil record. The primary astronomical fossils are, of course, stars and star clusters, and the main focus of this review will be the history of star formation in our Galaxy. Chemical evolution will be mentioned only briefly because it has been well reviewed in this volume by Matteucci. An excellent and extensive review of many aspects of galactic evolution has been given by Tinsley (1980).

Our Galaxy contains distinct spheroidal and flat components that will be discussed separately here and called simply the "halo" and the "disk," without regard to whether additional subcomponents can be distinguished. However, the radial structure

of the galactic halo and the properties of the central bulge region are of special interest and will also be discussed briefly.

2. EVOLUTION OF THE GALACTIC HALO

The galactic halo is of great interest as the oldest part of our Galaxy; information about its evolution therefore bears directly on the question of how the Galaxy formed. The traditional view is that our Galaxy and others formed by the collapse of extended and nearly spherical protogalactic gas clouds, and evidence supporting this view was presented in a classic paper by Eggen, Lynden-Bell, and Sandage (1962). These authors envisioned a homogeneous protogalactic cloud in which the only motions present were rotation and overall collapse, and they argued that the observed highly non-circular motions of halo stars can be understood only if the collapse was rapid and lasted no longer than 2×10^8 years. However, this simple model does not account for the fact that almost half of the stars in the galactic halo are actually on retrograde orbits. This fact led Larson (1969) to consider models of clumpy and turbulent protogalaxies; these models had longer collapse times that sometimes exceeded 109 years. An even more inhomogeneous picture of galaxy formation was proposed by Toomre (1977), who suggested that large galaxies form by the merging of smaller ones; this view was further elaborated by Tinsley and Larson (1979). Searle (1977) and Searle and Zinn (1978) also proposed that the properties of the outer halo globular clusters are best understood if they formed over a period of more than 1 Gyr in a number of separately evolving "protogalactic fragments."

2.1 Globular Cluster Chronology

Clearly, a key discriminant among these possibilities is the age spread of the globular clusters, since a small age spread (< 1 Gyr) would be consistent with the picture of rapid and ordered collapse proposed by Eggen, Lynden-Bell, and Sandage (1962), whereas a large age spread (> 1 Gyr) would suggest a more prolonged and chaotic process of galaxy formation, possibly involving mergers. A summary of some recent results suggesting a substantial age spread of the order of 5 Gyr among globular clusters has been given by Larson (1989). The strongest result is that three groups of investigators have all concluded that the small outer halo cluster Pal 12 is only about 11 Gyr old, significantly younger than the typical globular cluster age of about 15 Gyr (e.g., Stetson et al. 1989). Another result of several investigators, although one not universally accepted (see Hesser, this volume), is that the clusters NGC 288 and NGC 362, which have similar metallicities, differ in age by an amount that is at least 3 Gyr (Bolte 1989) and perhaps as large as 5.Gyr (Demarque et al. 1989). Supporting evidence for this result is provided by the fact that a similar age difference can account for the otherwise unexplained gross difference in horizontal-branch structure between these clusters, i.e., for the "second parameter effect" (Lee, Demarque, and Zinn 1988, 1990). This result supports the hypothesis that the second parameter for horizontal-branch structure is in fact age, as advocated by Searle and Zinn (1978) and Zinn (1986).

A recent study by Sarajedini and King (1989) finds evidence not only for an age spread but also for a correlation between age and metallicity in a sample of 31 globular clusters. This work did not take into account the strong variation of oxygen-to-iron ratio with metallicity found by Abia and Rebolo (1989), which will reduce the slope of

the correlation between age and metallicity; such a revision will not, however, affect the evidence for age differences between clusters of nearly the same metallicity, such as NGC 288 and NGC 362. In fact, even NGC 288 and Pal 12 have only a modest difference in metallicity, so the result that their ages differ by about a factor of 1.5 is not likely to be greatly altered. Thus there appears to be respectable, if not yet uncontested, evidence for a substantial age range among the galactic globular clusters.

2.2 Globular Cluster Formation

To understand more about how the halo of our Galaxy formed, we need to know something about the origin of the globular clusters themselves: did they form in isolation, or as parts of larger systems? And how can we understand their continuing formation over a period of several Gyr throughout the galactic halo? If globular clusters form with anything like a normal stellar mass function, the formation of a globular cluster must be a spectacular event producing thousands of O stars in a very small volume of space; such a phenomenon would probably be called a starburst. Searle (1977), Searle and Zinn (1978), and Larson (1988a) have argued that the homogeneity of the iron abundances within most globular clusters is best understood if the globular clusters formed in larger protogalactic subsystems that remained internally well mixed while becoming chemically enriched. A similar picture is suggested by the fact that there appears to be no sharp distinction between globular and open clusters and therefore no reason to suppose that globular clusters are formed in a fundamentally different way from open clusters, which are presently observed to form with a low efficiency of only $\sim (1-2) \times 10^{-3}$ in large star-forming molecular complexes (Larson 1990). This suggests that the globular clusters may have formed in much larger systems with masses exceeding $10^8 M_{\odot}$, i.e., systems with the masses of small galaxies. An example of such a system may be the metal-poor blue compact galaxy I Zw 18 = Mkn 116, whose central starburst region may presently be forming a very massive cluster. A few remnants of this early population of cluster-forming systems may have survived in the outskirts of our Galaxy as the local dwarf spheroidal galaxies, the largest of which, Fornax, contains several apparently normal globular clusters.

We are thus led to a picture in which the halo of our Galaxy formed in a chaotic fashion and over an extended period of time from subsystems that may have resembled compact galaxies. Eventually these subsystems were disrupted, perhaps by collisions with each other, and their stars and clusters were dispersed to form what is now the halo of our Galaxy. If the second parameter determining the structure of the horizontal branch is in fact age, then as argued by Searle and Zinn (1978) and Zinn (1986), the inferred age spread in the inner part of the galactic halo is small and the clusters there are all relatively old, while in the outer halo the age spread is larger and the average cluster age is younger. It is consistent with this inference that the youngest known globular cluster, Pal 12, is located in the outer part of the Galactic halo. It may therefore be the case that globular cluster formation began at about the same time everywhere but ceased first in the inner halo, continuing for progressively longer times at greater distances from the galactic center. This would be expected if the destruction of the cluster-forming subsystems was caused by collisions. The Magellanic Clouds, which are apparently now being disrupted and may soon merge with our Galaxy, may continue the process of building up the galactic halo from subsystems. One might then view the halo as being built up from the inside out in a process of gradual emergence of order from chaos.

2.3 The Central Bulge

How does the central bulge of our Galaxy fit into such a picture for the formation of the halo? While less information is available about the bulge, what is known about it suggests that it contains almost exclusively old stars and is very metal-rich, with an average iron abundance about twice that of the Sun (Frogel 1988). The globular clusters in the inner part of our Galaxy are also relatively metal-rich and may be among the oldest ones in the Galaxy, as noted above, although ages have not yet been directly measured for any clusters in the bulge region. The high metallicity of the bulge is a strong clue to its formation, and almost certainly excludes the possibility that most of its stars originated in smaller galaxy-like systems that merged, since smaller systems always have much lower metallicities; metallicities as high as those in the bulge are observed only in the inner parts of large galaxies. Thus, the bulge stars probably formed in situ from highly enriched gas that settled dissipatively into the central part of our Galaxy during its formation. Apparently this process occurred relatively early, and even before the formation of some of the younger globular clusters like Pal 12 in the outer halo. If so, the bulge may actually have been the first part of our Galaxy to form or to become organized into its present structure, and it may then have served as a nucleus around which the rest of the Galaxy was built up from the inside out. Note that such an "inside-out" picture of galaxy formation is diametrically opposite to the outside-in process of development that occurs in collapse models such as those of Larson (1969); in these models, the nucleus of a galaxy is the last part to form. Clearly, more information about the chronology of the inner part of our Galaxy will be of great interest in allowing us to address such fundamental questions about galaxy formation.

3. THE HALO-DISK TRANSITION

The age range discussed above (see also Sarajedini and King 1989) suggests that the bulk of the galactic halo formed between about 17 and 12 Gyr ago, and that the formation of at least the well-studied and relatively nearby part of the halo ended around 11 Gyr ago with the formation of Pal 12. In this case, the ages of the youngest halo objects are comparable to the age of the local galactic disk indicated by several types of evidence. Twarog (1980) found that the ages of local field stars determined from isochrone fitting range up to about 12 Gyr, although this number is probably fairly uncertain. According to Malaney and Fowler (1989), nucleochronology also sets an upper limit of about 12 Gyr on the age of the local galactic disk. Iben and Laughlin (1989) and Van Horn (this volume) have argued that the abrupt downturn of the luminosity function of white dwarfs at low luminosities is evidence that the local galactic disk cannot be much older than about 9 Gyr. The oldest known open cluster, NGC 6791, may be as old as 10 to 12 Gyr (Janes 1988), but a younger age of ~ 8 Gyr seems more likely (Demarque, private communication). The oldest well-studied disk object may actually be the globular cluster 47 Tuc, which has been assigned by Zinn (1985) and Armandroff (1989) to the disk population of globular clusters and which has an age of about 13 Gyr, according to the summary of results presented by Larson (1989).

Thus, within the uncertainties, it appears that the local galactic disk began to form at about the same time that halo formation ended, some 10 to 12 Gyr ago. Chronological continuity between halo and disk would be expected if the disk was formed out of gas left over from the halo subsystems, which were probably still quite gas-rich at

the time of their destruction. In any case, whatever the origin of the disk gas, it would probably have been impossible for this gas to settle permanently into a disk during the chaotic period of halo formation discussed above.

The chemical abundances of halo and old disk objects also suggest chemical continuity between halo and disk, the halo-disk transition occurring at an iron abundance of [Fe/H] = -0.8, according to Zinn (1985) and Armandroff and Zinn (1988). Apparently, after the iron abundance in the halo reached a value of this order, the remaining gas settled into a disk and then continued to be enriched by star formation in the disk. Twarog (1980) found for the oldest disk stars an iron abundance of $[Fe/H] \sim -0.55$, not very different from the value of -0.7 for 47 Tuc; the average [Fe/H] of disk stars then increases steadily with decreasing age, until reaching almost current values at an age of about 4 Gyr. In addition to the continuing increase of metallicity with time, another trend that appears to continue through the halo-disk transition is a decrease with time in the typical size of the clusters formed; prior to 10 Gyr ago our Galaxy seems to have formed only globular clusters, whereas after this time only open clusters have been formed. Similar trends are seen in some other nearby galaxies, such as the Magellanic Clouds.

It should be noted that these evolutionary trends apply in the well-studied and relatively nearby parts of our Galaxy, but not necessarily in the central regions or the outer parts of the halo. In fact, there is some evidence not only from the "second parameter" effect noted above but also from directly measured cluster ages (Sarajedini and King 1989) that the inner part of the Galactic halo evolved more rapidly than the outer part and reached high metallicities at earlier times. The Magellanic Clouds, at the other extreme, have evolved much more slowly, reaching metallicities as high as [Fe/H] = -0.8 only within the past few Gyr (Mould and Da Costa 1988); moreover, there is no clear halo/disk division in the Magellanic Clouds, and star formation appears to have proceeded in a relatively leisurely pace in them until recently.

4. EVOLUTION OF THE LOCAL GALACTIC DISK

We have direct information only about the evolution of a local part of the galactic disk consisting of an annular zone at the Sun's distance from the galactic center that includes the full vertical thickness of the disk; all of the subsequent discussion will refer to such a region. The most basic feature to be understood about the evolution of the local galactic disk, or of galactic disks generally, is the time dependence of the star formation rate (SFR): was the SFR much higher at early times, or did it decrease only moderately with time, or did it perhaps even remain essentially constant? It has proven surprisingly difficult to answer this question definitively, but it may now be possible at least to exclude some of the more extreme possibilities.

4.1 Simple Models

For purposes of discussion, it will be useful to consider three simple and classical models for the time dependence of the gas content and the SFR in a galactic disk. In these "closed box" models, it is assumed that no matter enters or leaves the system during its evolution; also, the gas depletion rate is assumed to be proportional to the SFR, as would be approximately true if the stellar Initial Mass Function (IMF) were independent of time. Each model is then specified by an assumption relating the SFR to the gas

content $M_{\rm g}$. Following Schmidt (1959, 1963) and many subsequent authors, and also the theoretical considerations of Section 4.2 below, the SFR will be assumed to vary as a power of $M_{\rm g}$. For illustration, we apply these models to the local galactic disk, for which we assume an age of 12 Gyr, a present gas surface density of 13 $M_{\odot}/{\rm pc}^2$, and a total surface density in stars, gas, and stellar remnants of 55 $M_{\odot}/{\rm pc}^2$ (see Section 4.4).

The model property of most interest for comparison with various observational constraints is the ratio $\langle SFR \rangle / SFR_0$ of the past average star formation rate $\langle SFR \rangle$ to the present rate SFR_0 ; in addition, we shall be interested in the present timescale for gas depletion, $\tau_g = -M_g/(dM_g/dt)$. The three standard models and their basic properties, for the above parameter values, are as follows:

- 1. Constant SFR model with SFR = constant \neq f(M_g). For this model, $\langle SFR \rangle / SFR_0 = 1.0$, and the (constant) gas consumption rate is 3.5 $M_{\odot} / \text{pc}^2 / \text{Gyr}$, yielding a present gas depletion timescale of $\tau_g = 3.7$ Gyr.
- 2. Linear model with SFR $\propto M_{\rm g}$. In this case, both the gas content and the SFR decline exponentially with time. For the above parameters, $\langle {\rm SFR} \rangle / {\rm SFR_0} = 2.2$ and the present gas consumption rate is 1.6 $M_{\odot}/{\rm pc^2}/{\rm Gyr}$, yielding a timescale of $\tau_{\rm g} = 8$ Gyr.
- 3. Quadratic model with SFR $\propto M_{\rm g}^2$. In this case, $M_{\rm g}$ and the SFR decline more rapidly at first and more slowly at later times than in the linear model. For the above parameters, the ratio $\langle {\rm SFR} \rangle / {\rm SFR_0}$ is 3.2 and the gas timescale is $\tau_{\rm g} = 12$ Gyr.

Models (2) and (3), the linear and quadratic models, were first proposed and studied by Schmidt (1959), and have remained popular ever since because they are based on a simple and plausible dependence of the SFR on the gas content. The linear model is the simplest analytically and has been most widely used. Model (1), the constant-SFR model, is less plausible physically but has been advocated recently as providing a better match to various observed properties of the stellar populations in the solar neighborhood and in typical late-type galaxies (e.g., Twarog 1980; Kennicutt 1983, 1986; Gallagher, Hunter, and Tutukov 1984). If the constant-SFR model is correct, then either the SFR is independent of the gas content, which is not consistent with theoretical expectations or with other evidence (see Section 4.2), or the gas supply is replenished at a rate that just balances the SFR, which also does not seem consistent with the evidence (Section 4.5). This dilemma and possible resolutions to it have been reviewed by Kennicutt (1986). It is argued below that there are good theoretical and observational reasons for believing that the SFR does depend at least linearly on the gas content, so that lack of dependence of the SFR on gas content is not a plausible resolution.

4.2 Theoretical Considerations

What controls the SFR in galactic disks? While it is clear that effects like density waves and tidal interactions can organize or enhance star formation in galaxies, it is also clear that such effects are not required to drive it, since many galaxies form stars vigorously without such help. Instead, the essential requirement for large-scale star formation to occur seems to be simply that the gas surface density be high enough for self-gravity to be important in the gas layer (e.g., Kennicutt 1989). Most stars form in large gravitationally-bound molecular clouds, which often belong to still larger self-gravitating complexes or spiral arm segments with sizes of order 1 kpc and masses of

order $10^7 M_{\odot}$ (Elmegreen and Elmegreen 1983, 1987). Thus star formation seems to require the formation of large self-gravitating aggregates of gas in galactic disks; the star formation rate will then depend on the rate at which such aggregates are formed.

Larson (1987, 1988b) has discussed several mechanisms that are likely to play a role in the creation of massive star-forming clouds; these include gravitational instability of the gas layer, swing amplification (a kind of limited instability of shearing disks), gravitational accretion by a seed cloud in a shearing gas layer, and cloud growth by random collisions. The timescales required for these processes to build massive clouds depend mainly on two parameters, the surface density μ and the velocity dispersion c of the gas layer. For example, the timescale for gravitational instability or swing amplification is approximately $c/\pi G\mu$. If the velocity dispersion c is the same everywhere, as seems observationally to be true to a first approximation, then this timescale depends only on the surface density μ and is proportional to μ^{-1} . If the efficiency of star formation in massive clouds is approximately constant (Larson 1987, 1988b), then the SFR per unit area in a galactic disk is predicted to be proportional to μ^2 , similar to the law first proposed by Schmidt (1959) but depending physically on surface density rather than volume density.

This simple quadratic law does not, however, readily account for the fact that galaxies of later Hubble types are presently consuming their gas more slowly than earlytype galaxies, despite the fact that their gas surface densities are not very different; Donas et al. (1987) find, in fact, a systematic increase of the estimated gas depletion time τ_g with Hubble type among late-type galaxies. This result could be understood if the velocity dispersion c is not strictly constant but is influenced by the large-scale dynamics of the disk, as happens for example in numerical simulations where gravitational acceleration or "heating" effects tend to maintain the stability parameter $Q = c\kappa/\pi G\mu$ nearly constant at a value of about 2; here κ is the local epicyclic frequency in the disk. If Q is constant, the above timescale becomes proportional to κ^{-1} , and the SFR per unit area then becomes proportional to $\kappa\mu$; in this case the SFR depends linearly on μ , but there is an additional dependence on epicyclic frequency that allows the Hubble-type dependence noted above to be understood (Larson 1988b). Probably reality lies somewhere between the two possibilities that have been discussed, in which case the SFR per unit area would vary with a power between 1 and 2 of the gas surface density. These considerations add some theoretical plausibility to simple star formation laws like those considered by Schmidt (1959), and they suggest that something intermediate between the linear and quadratic models of galactic evolution may be most realistic.

How well do the observed star formation rates in galactic disks conform to such predicted dependences on gas surface density? Viallefond (1988), in a re-analysis of the data of Donas et al. (1987), finds that in a large sample of spiral galaxies the average SFR per unit area is well correlated with the average gas surface density, being proportional to about the 1.6 power of the gas surface density. For a smaller sample of galaxies with complete data on both atomic and molecular gas, Buat, Deharveng, and Donas (1989) find a very similar relation with an exponent of about 1.65, although a smaller exponent is also possible if different assumptions are made about extinction. The SFR per unit area is found by these authors to correlate well with the total (atomic plus molecular) gas surface density, but poorly with the molecular surface density alone. Very similar results have been found by Kennicutt (1989). These results suggest that it is the total surface density of gas that controls the SFR, as expected theoretically, and not just the molecular surface density.

Kennicutt (1989) has also found, in agreement with earlier results of Skillman

(1987) and with theoretical expectations for star formation driven by gravitational instability (Quirk 1972), that there is a threshold gas surface density that is required for significant star formation to occur. The empirical threshold surface density corresponds to a value for Q of about 2, which is approximately the value below which swing amplification effects become important (Toomre 1981); this result is therefore in good agreement with the idea that swing amplification plays a major role in driving star formation in galactic disks (Larson 1983). For surface densities well above the threshold, Kennicutt (1989) finds that the SFR per unit area varies approximately as the 1.3 power of the gas surface density. Some of the confusion that has previously existed in this subject (see Freedman 1986 for a review) may have resulted from failure to recognize the existence of such a threshold and the consequent departure from simple power-law relationships.

In summary, there appear to be persuasive theoretical and observational reasons to believe that the SFR in galactic disks depends importantly on the gas content. Empirically, the SFR per unit area seems to vary as about the 1.3 power of the gas surface density, and much more strongly near the threshold. As a result, since our Galaxy is apparently not accreting gas as rapidly as it is being consumed, we would expect that the SFR was significantly higher in the past.

4.3 Constraints on the History of Star Formation

Constraints on the time dependence of the SFR in the solar neighborhood, and in galactic disks generally, are provided by the observed properties of the stellar populations in these systems. In the solar neighborhood it is possible, at least in principle, to determine directly the age distribution of the local disk stars by isochrone fitting. This was attempted by Twarog (1980) for a large sample of nearby F stars; he found, under what he considered to be the most likely assumptions, that the average formation rate of these stars over the past 12 Gyr was about 1.5 times the present rate. Twarog argued further that a constant SFR is not excluded, and showed that models with a constant SFR and with gas infall yield a good fit to the observed age-metallicity relation. Tinsley (1981) also adopted a model with a constant SFR for the local galactic disk.

Twarog's result depends, however, on the assumed age dependence of the scale height of local field stars, which is not well constrained observationally. Twarog assumed that it is constant for ages between 0 and 4 Gyr, and then increases linearly with age to a value 6 times as large for the oldest stars. However, the scale height almost certainly has a much stronger age dependence for ages between 0 and 4 Gyr, because the velocity dispersion of local stars shows a strong increase with age in this range (Wielen 1977; Carlberg et al. 1985). If the formation rate of Twarog's stars is re-calculated assuming a scale height that increases with age in proportion to the velocity dispersion derived by Wielen (1977), the SFR then shows a broad peak at a time about 6 Gyr ago, with a smooth decline toward both smaller and larger ages; the maximum SFR is about 3.5 times the present rate, and the ratio of past average to present SFR is about 2.4. Clearly there remains considerable uncertainty in any such estimate, but a value for (SFR)/SFR₀ of approximately 2 at least does not seem excluded by this evidence.

Kennicutt (1983, 1986) has argued for a nearly constant SFR in the solar neighborhood and in similar late-type galaxies on the grounds that a nearly constant SFR is required in order for stellar population models to match the observed UBV colors and $H\alpha$ equivalent widths of these systems. Kennicutt modeled the $H\alpha$ equivalent widths of a large sample of Sc galaxies, and obtained a median value for $\langle SFR \rangle / SFR_0$ of only

about 1.3; this value is too small, although not by a large factor, to agree with simple linear or quadratic models fitted to the observed gas contents of these galaxies, for which the median values of $\langle SFR \rangle / SFR_0$ are 2.3 and 3.3 respectively, almost the same as for the solar neighborhood.

We can re-evaluate the constraint placed by UBV colors on the star-forming histories of galaxies by using the recent stellar population models of Yamanaka (1987); these models yield slightly bluer colors and larger $H\alpha$ equivalent widths for galaxies with a constant SFR than most earlier models, although this difference is probably no larger than the uncertainties that still remain. For a set of such models with an age of 15 Gyr, an exponentially declining SFR, and a Salpeter IMF, the B-V color is uniquely related to the ratio $\langle SFR \rangle / SFR_0$; hence the models can be used to estimate this ratio for galaxies with known colors. Adopting average galaxy colors from de Vaucouleurs (1977) as a function of Hubble type, we then obtain the following results:

Hubble type	B-V	⟨SFR⟩/SFR₀
S0	0.85	20.0
\mathbf{Sa}	0.75	7.0
$\mathbf{S}\mathbf{b}$	0.64	3.4
\mathbf{Sc}	0.52	1.8
Sd	0.43	1.1
Irr	0.37	0.7

Our Galaxy is usually considered to have a Hubble type intermediate between Sb and Sc (e.g., de Vaucouleurs 1983); we would then infer from the above table that the average value of $\langle SFR \rangle / SFR_0$ for our Galaxy is between 1.8 and 3.4. The integrated B-V color of the local galactic disk has been estimated by de Vaucouleurs (1983) to be about 0.53, which suggests that at least the local part of our Galaxy is more similar to an Sc than to an Sb galaxy, and that the local value of $\langle SFR \rangle / SFR_0$ is about equal to 2. Note that, according to the above table, even typical Sc galaxies have experienced a significant decline of the SFR with time; only the very latest Hubble types, *i.e.*, types Sd and later, appear to have had nearly constant star formation rates.

We conclude that a value for $\langle SFR \rangle / SFR_0$ of about 2 is consistent with the observed properties of the stellar populations in both the solar neighborhood and typical Sc galaxies; the simple linear model with $\langle SFR \rangle / SFR_0 = 2.2$ is therefore not excluded as a description of the evolution of these systems. However, the constant-SFR model with $\langle SFR \rangle / SFR_0 = 1$ and the quadratic model with $\langle SFR \rangle / SFR_0 = 3.2$ appear less likely because they would predict colors that are too blue and too red, respectively, assuming that a Salpeter IMF is always applicable. We note in passing that the white dwarf luminosity function can in principle provide additional information about the past SFR in the solar neighborhood (Meusinger 1989; Noh and Scalo 1990), but present constraints from this method are very broad and do not exclude any of the possibilities that have been discussed here.

4.4 Constraints on the Initial Mass Function

The properties of stellar populations depend not only on the star formation rate as a function of time, but also on the IMF and its possible variations; in fact, the SFR and the IMF are so closely linked in their observational implications that we usually cannot derive one without making some assumption about the other. Simple models like those considered above assume that the IMF is constant and not greatly different from that originally derived by Salpeter (1955); an extensive review of this subject has been given by Scalo (1986). The IMF derived from the luminosity function of local field stars depends on the assumed history of star formation, and most derivations assume not only that the IMF has been constant but also that the SFR has been constant in time. If the SFR has not been constant, then the derived IMF will be modified.

If the SFR has decreased with time, then the upper IMF must be corrected upward relative to the lower IMF to account for the larger number of massive stars that formed and died during past galactic history, relative to the number of low-mass stars still observed; for example, if $\langle SFR \rangle / SFR_0 = 2$, the upper IMF must be corrected upward by a factor of 2 relative to the lower IMF. Such a change would have the effect of flattening the derived IMF in the vicinity of 1 M_{\odot} ; in fact, a second peak is found in the derived IMF if the SFR is assumed to have decreased very significantly with time (Scalo 1986). The occurrence of a jump in the IMF near 1 M_{\odot} has usually been regarded as implausible, and this in turn has been taken as an argument for a nearly constant SFR (e.g., Tinsley 1977; Miller and Scalo 1979). However, if the possibility of a non-monotonic IMF is allowed, then models are in principle possible in which the SFR has decreased strongly with time and in which, as noted by Quirk and Tinsley (1973) and Green (1980), a large amount of mass is predicted to be in stellar remnants, which might then account for the unseen mass believed to be present in the solar neighborhood.

Motivated by the possibility of accounting for the unseen mass, by previous suggestions that star formation is bimodal (Güsten and Mezger 1983), and by possible physical reasons for variability of the IMF (Larson 1985), Larson (1986a) proposed a model of galactic evolution in which the IMF is bimodal and consists of two components that peak below and above 1 M_{\odot} . If only the high-mass component declines strongly with time, it is then possible to generate a large amount of mass in remnants without violating the constraint that the SFR for stars of near solar mass has decreased only moderately with time. A variable-IMF model similar in its basic properties to this bimodal model had earlier been proposed by Schmidt (1963); Schmidt showed that if the formation of massive stars was favored at early times, as first suggested by Schwarzschild and Spitzer (1953), then a solution could be provided to the "G-dwarf problem," i.e., the relative paucity of metal-poor stars. The bimodal model also offers a solution to this problem, while, as noted by Larson (1986b), the model of Schmidt also predicts a large amount of mass in remnants.

Scalo (1987a) has emphasized that direct evidence for bimodality or strong variability of the IMF is weak, and Scalo and other authors (e.g., Meusinger 1989) have pointed out a number of difficulties with such models. However, the most direct constraint, which may now even eliminate this type of model, is that provided by improved determinations of the amount of unseen mass in the solar vicinity. The bimodal model of Larson (1986a) was constructed to account for the surface density of 67 M_{\odot}/pc^2 favored by Bahcall (1984a, b) for the local galactic disk from dynamical studies. With a surface density of $22 M_{\odot}/\text{pc}^2$ in visible stars and $8 M_{\odot}/\text{pc}^2$ in gas, the model accounted for the remaining $37 M_{\odot}/\text{pc}^2$ as stellar remnants, mostly white dwarfs. More recent estimates of the total column density of matter in the solar neighborhood have yielded

smaller values than those favored by Bahcall and previous investigators: Kuijken and Gilmore (1989) estimate from new data a column density of only 46 M_{\odot}/pc^2 , while Gould (1990) finds, from a re-analysis of the same data, about 54 M_{\odot}/pc^2 . It is clear that much more work will be needed to establish an accurate value for this number, but for illustration we shall suppose that a total column density of 55 M_{\odot}/pc^2 is more likely to be correct than the 67 M_{\odot}/pc^2 assumed earlier.

Another revision suggested by recent work is a higher column density of visible stars, mainly due to larger adopted scale heights for old stellar populations. For example, Rana (1987) adopts a scale height of 580 pc for old disk stars, much larger than the 325 pc assumed by Miller and Scalo (1979) and Scalo (1986), and as a result he finds a column density of 35 M_{\odot}/pc^2 for visible main-sequence stars. Kuijken and Gilmore (1989) adopt a column density of 35 $M_{\odot}/\mathrm{pc^2}$ for all known stars, including about 4 M_{\odot}/pc^2 for white dwarfs. In addition, a larger value is probably required for the column density of gas, which may be as large as 13 M_{\odot}/pc^2 (Kulkarni and Heiles 1987; Kuijken and Gilmore 1989). With these revisions, the column density of matter not yet accounted for becomes only $55-35-13=7~M_{\odot}/{\rm pc^2}$, and the maximum possible column density in remnants becomes about 11 $M_{\odot}/\mathrm{pc^2}$, or 30 percent of the amount in the model of Larson (1986a). In order to agree with this revised constraint, the highmass mode of star formation in the model would have to be scaled down by at least a factor of 4 in its average SFR; a similar revision would be required to the model of Schmidt (1963). Bimodality or variability of the IMF would then no longer be of major importance, and models with a less strongly variable or even constant IMF might work equally well or better (Olive 1986; Rana 1987; Meusinger 1989). Clearly, in any case, measurements of the amount of unseen mass in the solar neighborhood can place strong constraints on models of galactic evolution by limiting the past variability of both the SFR and the IMF.

If we therefore return for simplicity to the assumption that the IMF has not varied with time, but also assume that $\langle SFR \rangle / SFR_0 = 2$ as discussed in Section 4.3, then the IMF derived using the data of Scalo (1986) shows a modest jump by about a factor of 1.5 between 0.7 and 1.2 M_{\odot} and has a second peak at 1.2 M_{\odot} . Considering the many large uncertainties involved in deriving the IMF, including its strong sensitivity to the exact slope of the mass-luminosity relation and also to assumptions about scale heights, as noted above, it is not clear that this second peak is a real feature, and it is probably reasonable to adopt a smoothed and broadly peaked IMF, perhaps like that suggested by Rana (1987), which approaches a power law at high masses but drops progressively farther below this power law at masses below 1 M_{\odot} . Because the average slope of this function over the relevant range of masses is still not very different from the Salpeter law, it would probably not predict very different galaxy colors when used in stellar population models; the conclusions of Section 4.3 may therefore not require much revision. Another check on such a model with $\langle SFR \rangle / SFR_0 = 2$ and a constant IMF is that the predicted column density of stellar remnants is in this case about 7 M_{\odot}/pc^2 , which is quite consistent with the revised dynamical constraints discussed above.

4.5 Gas Depletion and the Role of Infall

The rate of evolution of a star-forming system is controlled by the rate $dM_{\rm g}/dt$ at which its gas content $M_{\rm g}$ is depleted, which is conveniently expressed in terms of the timescale $\tau_{\rm g} = -M_{\rm g}/(dM_{\rm g}/dt)$. Many attempts have been made to estimate this quantity for the solar neighborhood and for other galaxies, and it has often been concluded that

the resulting timescale is too short to be readily understood unless the gas is somehow replenished. For example, in early work van den Bergh (1957) estimated a local gas depletion time of only 0.7 Gyr (for an assumed disk age of 5 Gyr) and suggested that the gas may be replenished by a radial outflow in the galactic disk, while Larson (1972) estimated a timescale of 3 to 5 Gyr which he still regarded as uncomfortably short, and suggested that gas might be re-supplied by infall from outside the Galaxy.

In a later study, Larson, Tinsley, and Caldwell (1980) derived a local gas depletion time of only 2 Gyr, and again suggested that infall is important. However, this number is almost certainly too small because it assumes that half of the gas is turned into unseen brown dwarfs; since it now appears unlikely that brown dwarfs are a major constituent of the local galactic disk (indeed, the revised dynamical constraints discussed above leave little room for brown dwarfs), this number should probably be increased to 4 Gyr. For the same reason, the median value of τ_g estimated by Larson, Tinsley, and Caldwell for a sample of 36 spiral galaxies should probably be increased from 4 to 8 Gyr. Kennicutt (1983) estimated a median τ_g of about 5 Gyr for a larger sample of 85 galaxies, and he considered this result to support the conclusions of Larson, Tinsley, and Caldwell; Kennicutt assumed no brown dwarfs, however, so his result is not subject to the same upward revision. A recent study by Donas et al. (1987) of star formation rates in 149 galaxies yields a median τ_g that is about 7 Gyr when calculated using the same assumptions as Kennicutt (1983). Very similar results were obtained by Sandage (1986) in a discussion of gas depletion rates in the solar neighborhood and several Local Group dwarf galaxies. Thus, recent estimates of the gas depletion time in galaxies have generally yielded somewhat larger values than earlier estimates, and it is no longer clear that these times are too short, especially when compared with the predictions of simple models such as those of Section 4.1.

Using the parameters for the solar neighborhood adopted in Section 4.4, we can make a new estimate of the gas depletion timescale $\tau_{\rm g}$ in the solar neighborhood. If the total surface density is $55~M_{\odot}/{\rm pc^2}$ and the amount remaining in gas is $13~M_{\odot}/{\rm pc^2}$, then the amount of gas that has been removed by star formation is $42~M_{\odot}/{\rm pc^2}$, and the average rate of gas consumption over the past 12 Gyr is $3.5~M_{\odot}/{\rm pc^2}/{\rm Gyr}$. If we assume that $\langle {\rm SFR} \rangle/{\rm SFR_0} = 2$ and that the IMF has not varied, so that the ratio of the past average gas depletion rate to the present rate is also equal to 2, then the present gas depletion rate is $1.8~M_{\odot}/{\rm pc^2}/{\rm Gyr}$. Note that, because of the fairly high gas return fraction of ~ 0.48 in the model considered here, this corresponds to a present SFR of about $3.4~M_{\odot}/{\rm pc^2}/{\rm Gyr}$. The present gas depletion timescale is then $\tau_{\rm g}=13/1.8$, or about 7 Gyr. This is larger than previous estimates chiefly because of the higher surface density of gas adopted here.

Comparing this result for $\tau_{\rm g}$ with the properties of the simple models of Section 4.1, we see that it is in good agreement with the linear model, for which $\tau_{\rm g}$ is about 8 Gyr, but not with the constant-SFR or quadratic models, for which $\tau_{\rm g}$ is about 4 and 12 Gyr, respectively. This is of course expected because the value for $\langle {\rm SFR} \rangle / {\rm SFR_0}$ adopted here is nearly the same as the value of 2.2 appropriate for the linear model. For typical Sc galaxies, there is also reasonable agreement with the linear model, since a similar argument assuming $\langle {\rm SFR} \rangle / {\rm SFR_0} = 1.8$ and a median gas fraction of 22% (Kennicutt 1986) yields a gas depletion time of about 6 Gyr, compared with 8 Gyr for the linear model. Since the simple linear model is not in conflict with any of the observational or theoretical considerations that have been discussed in this review, there is no convincing case from these timescale estimates for gas replenishment by infall.

Nevertheless, there is good evidence for a net inflow of gas into our Galaxy, albeit

at a considerably lower rate than was originally estimated (e.g., Oort 1970; Larson 1972). According to Mirabel (1989), the total rate of infall of gas into our Galaxy is likely to be between 0.2 and 0.5 M_{\odot} per year; if we assume (although this is highly uncertain) that this gas falls in uniformly over a disk of radius 15 kpc, the infall rate per unit area is then between 0.3 and 0.7 $M_{\odot}/\text{pc}^2/\text{Gyr}$. This is between 0.16 and 0.4 times the local gas depletion rate of 1.8 $M_{\odot}/\mathrm{pc^2/Gyr}$ estimated above; thus gas infall at this rate would lengthen the timescale for gas depletion by between 16 and 40 percent. If these numbers are correct, it would appear that gas infall in the solar neighborhood is a minor but not completely negligible effect at the present time. Infall may nevertheless have played a significant role in the past chemical evolution of the galactic disk, because infall was almost certainly much more important at early times. As we have seen, there was apparently a prolonged period of halo formation preceding the formation of the local galactic disk, and there may well have been an equally prolonged period of several Gyr during which the leftover gas continued to settle into a disk. Both the gas content and the SFR in the disk would then have increased with time for several Gyr following the formation of the disk, before they began to decline again. This possibility would be consistent with the reconstruction of the past SFR in the solar neighborhood described in Section 4.3, and with the galaxy formation models of Larson (1976); an analytic model of this type was also suggested by Lynden-Bell (1975) to account for galactic chemical evolution. Models basically similar to the type of model suggested here have been studied in more detail by a number of authors, including Chiosi (1980), Vader and de Jong (1981), and most recently Matteucci and François (1989), and have been found to produce an acceptable account of the chemical evolution of our Galaxy.

5. CONCLUSIONS

If the globular clusters have an age spread like that suggested by recent results, it would appear that the formation of the halo of our Galaxy was a prolonged and chaotic process involving the formation, evolution, and eventual disruption of subsystems that may have resembled compact galaxies. Such a complex process of galaxy formation cannot readily be modeled in any simple fashion; therefore the only way to obtain a better understanding of this stage of galactic evolution is through continuing study of the fossil record. Better information about globular cluster ages will obviously be of key importance. A central goal of such research would be to learn more about the properties of the hypothetical halo subsystems.

After halo formation ceased and the residual gas settled into a disk, the evolution of our Galaxy apparently began to proceed in a more orderly fashion. Many uncertainties are involved in establishing a reliable account even of the evolution of the disk of our Galaxy, but if the data that have been adopted in this review are correct, relatively simple models may suffice and some long-standing problems may be less serious than has been thought. For example, the amount of unseen mass in the disk may in fact be small, and consistent with the amount expected to be in dark remnants in a simple model having a modest decline of the SFR with time. The discrepancy that has been thought to exist between the nearly constant SFR inferred empirically and the declining SFR expected theoretically may not be real, since the observations seem consistent with a ratio of past average to present SFR of about 2 that is essentially the same as the value predicted for a model in which the SFR is proportional to the gas content and both decline exponentially with time. The estimated gas depletion timescale also appears to

be consistent with such a model, in which case there is no need for gas replenishment by infall to be important at present. Finally, a constant IMF not radically different from that first proposed by Salpeter may be an adequate approximation for the local galactic disk and typical spiral galaxies. In other words, simple and physically plausible models of galactic evolution may work after all.

However, many of the modifications to simple models that have been proposed to solve the various problems mentioned above may yet find their place as necessary refinements in a more complete description of galactic evolution. For example, gas infall, which seems to be of minor importance at present, was almost certainly more important in the past, and infall has in fact become a standard feature of models of galactic chemical evolution, as reviewed in this volume by Matteucci. Radial gas flows may also play a role, as suggested by Tinsley and Larson (1978), Lacey and Fall (1985), and Clarke (1989). Variability of the IMF almost certainly occurs at some level, and bimodality or other kinds of "fine structure" cannot be excluded, although such effects may not be dominant; in any case, in models like that suggested above, the local IMF must be significantly flatter than a Salpeter function at masses of $\sim 1~M_{\odot}$ and less.

Finally, lest we become too complacent with our ability to model the evolution of galactic disks, we must recall that chaos can sometimes intervene even in the evolution of disks, as in the evolution of many types of complex systems (Scalo 1987b). For example, perturbations caused by the near passage or capture of a companion galaxy can alter the evolution of a galaxy, causing large fluctuations in the star formation rate and even significantly changing the structure of the galaxy. Although our Galaxy has apparently not experienced major disturbances in its recent history, considerable chaos may be in store if and when the Magellanic Clouds are accreted by our Galaxy; models like those that have been discussed here may then need major revision.

REFERENCES

Abia, C. and Rebolo, R. 1989, Ap. J., 347, 186.

Armandroff, T. E. 1989, A.J., 97, 375.

Armandroff, T. E. and Zinn, R. 1988, A.J., 96, 92.

Bahcall, J. N. 1984a, Ap. J., 276, 169.

1984b, Ap. J., **287**, 926.

Bolte, M. 1989, A.J., 97, 1688.

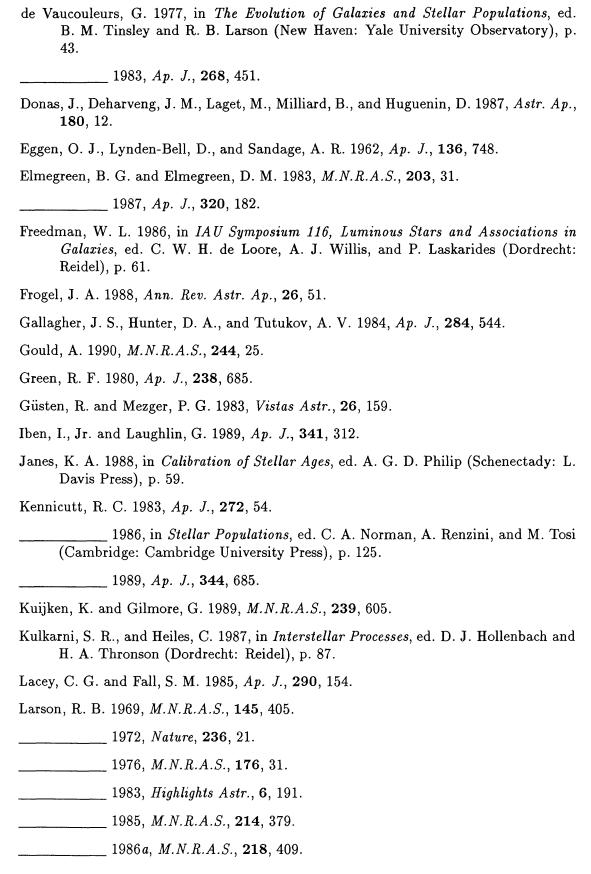
Buat, V., Deharveng, J. M., and Donas, J. 1989, Astr. Ap., 223, 42.

Carlberg, R. G., Dawson, P. C., Hsu, T., and VandenBerg, D. A. 1985, Ap. J., 294, 674.

Chiosi, C. 1980, Astr. Ap., 83, 206.

Clarke, C. J. 1989, M.N.R.A.S., 238, 283.

Demarque, P., Lee, Y.-W., Zinn, R., and Green, E. M. 1989, in *The Abundance Spread Within Globular Clusters: Spectroscopy of Individual Stars*, ed. G. Cayrel de Strobel, M. Spite, and J. Lloyd Evans (Paris: Observatoire de Paris), p. 97.



- ______ 1986b, in Stellar Populations, ed. C. A. Norman, A. Renzini, and M. Tosi (Cambridge: Cambridge University Press), p. 101.
- ______1987, in Starbursts and Galaxy Evolution, ed. T. X. Thuan, T. Montmerle, and J. Tran Thanh Van (Gif sur Yvette: Editions Frontières), p. 467.
- _____ 1988a, in IAU Symposium 126, Globular Cluster Systems in Galaxies, ed. J. E. Grindlay and A. G. D. Philip (Dordrecht: Kluwer), p. 311.
- and M. Fich (Dordrecht: Kluwer), p. 459.
- ______1989, in *Dynamics and Interactions of Galaxies*, ed. R. Wielen (Berlin: Springer-Verlag), p. 48.
- ______1990, in *Physical Processes in Fragmentation and Star Formation*, ed. R. Capuzzo-Dolcetta, C. Chiosi, and A. Di Fazio (Dordrecht: Kluwer), p. 389.
- Larson, R. B., Tinsley, B. M., and Caldwell, C. N. 1980, Ap. J., 237, 692.
- Lee, Y.-W., Demarque, P., and Zinn, R. 1988, in *Calibration of Stellar Ages*, ed. A. G. D. Philip (Schenectady: L. Davis Press), p. 149.
- _____ 1990, Ap. J., **350**, 155.

Lynden-Bell, D., 1975, Vistas Astr., 19, 299.

Malaney, R. A. and Fowler, W. A. 1989, M.N.R.A.S., 237, 67.

Matteucci, F. and François, P. 1989, M.N.R.A.S., 239, 885.

Meusinger, H. 1989, Astr. Nachr., 310, 29 and 115.

Miller, G. E. and Scalo, J. M. 1979, Ap. J. Suppl., 41, 513.

- Mirabel, I. F., 1989, in IAU Colloquium 120, Structure and Dynamics of the Interstellar Medium, ed. M. Moles, G. Tenorio-Tagle, and J. Melnick (Berlin: Springer-Verlag), p. 396.
- Mould, J. R. and Da Costa, G. S. 1988, in *Progress and Opportunities in Southern Hemisphere Optical Astronomy*, ed. V. M. Blanco and M. M. Phillips (Provo: Astronomical Society of the Pacific), p. 197.

Noh, H.-R. and Scalo, J. M. 1990, Ap. J., 352, 605.

Olive, K. A. 1986, Ap. J., 309, 210.

Oort, J. H. 1970, Astr. Ap., 7, 381.

Quirk, W. J. 1972, Ap. J. (Letters), 176, L9.

Quirk, W. J. and Tinsley, B. M. 1973, Ap. J., 179, 69.

Rana, N. C. 1987, Astr. Ap., 184, 104.

Salpeter, E. E. 1955, Ap. J., 121, 161.

Sandage, A. 1986, Astr. Ap., 161, 89.

Sarajedini, A. and King, C. R. 1989, A.J., 98, 1624.

Scalo, J. M. 1986, Fund. Cosmic Phys., 11, 1.

______ 1987a, in Starbursts and Galaxy Evolution, ed. T. X. Thuan, T. Montmerle, and J. Tran Thanh Van (Gif sur Yvette: Editions Frontières), p. 445.

______ 1987b, in *Evolution of Galaxies*, ed. J. Palous (Prague: Czechoslovak Academy of Sciences), p. 101.

Schmidt, M. 1959, Ap. J., 129, 243.

_____ 1963, Ap. J., **137**, 758.

Schwarzschild, M. and Spitzer, L. 1953, Observatory, 73, 77.

Searle, L. 1977, in *The Evolution of Galaxies and Stellar Populations*, ed. B. M. Tinsley and R. B. Larson (New Haven: Yale University Observatory), p. 219.

Searle, L. and Zinn, R. 1978, Ap. J., 225, 357.

Skillman, E. D. 1987, in *Star Formation in Galaxies*, ed. C. Lonsdale Persson (Washington: NASA Conference Pub. 2466), p. 263.

Stetson, P. B., VandenBerg, D. A., Bolte, M., Hesser, J. E., and Smith, G. H. 1989, A.J., 97, 1360.

Tinsley, B. M. 1977, Ap. J., 216, 548.

1980, Fund. Cosmic Phys., 5, 287.

_____ 1981, Ap. J., **250**, 758.

Tinsley, B. M. and Larson, R. B. 1978, Ap. J., 221, 554.

_____ 1979, M.N.R.A.S., **186**, 503.

Toomre, A. 1977, in *The Evolution of Galaxies and Stellar Populations*, ed. B. M. Tinsley and R. B. Larson (New Haven: Yale University Observatory), p. 401.

______ 1981, in *The Structure and Evolution of Normal Galaxies*, ed. S. M. Fall and D. Lynden-Bell (Cambridge: Cambridge University Press), p. 111.

Twarog, B. A. 1980, Ap. J., 242, 242.

Vader, J. P. and de Jong, T. 1981, Astr. Ap., 100, 124.

van den Bergh, S. 1957, Zs. Ap., 43, 236.

Viallefond, F. 1988, in *Galactic and Extragalactic Star Formation*, ed. R. E. Pudritz and M. Fich (Dordrecht: Kluwer), p. 439.

Wielen, R. 1977, Astr. Ap., 60, 263.

Yamanaka, J. M. 1987, Ph.D. thesis, Yale University.

Zinn, R. 1985, Ap. J., 293, 424.

Zinn, R. 1986, in *Stellar Populations*, ed. C. A. Norman, A. Renzini, and M. Tosi (Cambridge: Cambridge University Press), p. 73.