Chapter 5

VASOSPASM AND GENERAL ARTERIAL DISEASE

It is vain to do with more what can be done with less.

William of Occam, 1300-1349

INTRODUCTION

Having made the case that vasospasm might be important in the pathogenesis of certain specific circulatory disorders, I should now like to go on and examine how it may also help understand the pathophysiological basis of a broader spectrum of vascular disease. In doing so, I shall try to make the analysis more than merely speculative by keeping close to the data in the various fields, so as to allow an assessment of where vasospasm might indeed explain what traditional theories cannot. In such areas, any mechanism like vasospasm which carries a potential for reversibility must be worth considering even by those who think it unlikely to explain very much. The present chapter will be devoted to an analysis of its possible role in arterial disease per se. Much of the discussion will centre around pathophysiology, because these are the arena from which I believe most understanding will come. A serious decline in physiology as an academic discipline over recent years has been due largely to two factions: societal pressure for a reduction in animal experimentation, and the modern drive towards molecular biology. Yet it seems to me that, to paraphrase Dobzhansky: Nothing in molecular biology makes sense except in the light of physiology.¹

ARTERIES WITH VASA VASORA

These are the large and medium-sized arteries. A case has already been made out for an important role of vasospasm in the pathogenesis of atherosclerosis in the previous chapter, where it was suggested that focal narrowing of large and medium-sized arteries by vasospasm could lead to local increases in blood flow velocity at the site, and the consequent damage to the arterial intima this could set the stage for the development of local atherosclerosis. It was also postulated that some of the complications of the atherosclerotic plaque might be related to its eventually becoming dependent on a vasa vasorum blood supply. The vasa vasora, or vessels of the vessels, normally enter the large and medium-sized arteries through their adventitial or external coats, and provide nutrition only as far into the wall as the middle media.² Now, atherosclerotic plaques are intimal in position, and derive their blood supply from these vasa vasora being extended beyond their normal (see Ch. 4). What I should like to do here, however, is discuss the possibility that vasospasm of large and medium-sized arteries may also contribute to diseases of the arterial media through restriction of the more normal distribution of the vasa vasorum circulation. Recurrent, and especially prolonged arterial vasospasm could have several important
DISEASES OF THE LARGE ARTERIES

1. Vasa vasora and atherosclerotic arterial aneurysms.

Even quite minor episodes of vasospasm in large and medium-sized arteries could seriously compromise the vasa vasorum circulation to their outer layers, because any pressure generated in an artery by vasospasm would only need to be enough to balance the luminal blood pressure in order to seriously impede the vasa vasorum circulation within its walls. Indeed, even a constriction which did not narrow the major arterial lumen very much at all might still interfere with outer wall nutrition. Such a mechanism might well explain several otherwise puzzling aspects of large arterial disease. Firstly, it could account for the thinning of the arterial media over areas severely affected by atheroma. Wherever this medial thinning became extensive through repeated vasospastic episodes, it could eventually weaken the arterial wall, and so provide the basis for the formation of ‘atherosclerotic’ aneurysms, for example of the lower abdominal aorta. This process only needs a start, because once any dilatation has begun, the forces tending to stretch and dilate it still further would, from Laplace’s laws, increase, so that the aneurysmal dilatation could eventually become to some extent self-perpetuating. Then, a thin-walled aneurysm might perhaps only require a slight further episode of vasospasm to produce the final ischaemic necrotic event which led to its rupture.

2. Medial degeneration of the aorta.

Even in areas not affected by atherosclerosis, impairment of arterial wall nutrition, from repetitive and prolonged vasospasm indirectly causing restriction of the vasa vasorum circulation, could lead to atrophy and degeneration of the media of the larger arteries, and this might be very relevant to the condition of cystic medial degeneration and/or cystic arterial necrosis. This is largely an atrophic condition, and in a mild form is seen not uncommonly with advancing years even in normal people. It is not unusually of great consequence, but may become so in the background to a condition that most decidedly is, viz. dissecting aneurysm of the aorta.

3. Dissecting aortic aneurysm.

This is a most peculiar condition, where the aortic blood stream suddenly begins to track down the length of the aortic wall, splitting its medial smooth muscle cell layer in two in the process. The usual entry site for this blood is through an intimal tear, so why the aortic stream should suddenly diverge down through a medial coat pathway of much higher resistance is very uncertain. Some features are recognised as important in setting the stage for dissection, yet none seem sufficient to explain it alone. Thus, intimal tears can be deliberately produced in the aorta during the study of experimental atherosclerosis without, as far as I am aware, causing any dissection of the aortic wall at all. Cystic medial degeneration is known to facilitate dissection, but that is not sufficient either. Hypertension also contributes as a factor predisposing to the dissection process, but even when added to others, this does not explain its whole basis.

Given this general background, the situation might be very different indeed if acute severe aortic vasospasm were to be superimposed. At least it is worth analysing the possible consequences of vasospasm in this respect. Dissecting aneurysms often begin in the ascending aorta, where there are no significant side-branches between it and the heart that can act as “blow off” points to provide a safety valve for the diversion of the systolic blood stream away from any area of segmental obstruction to flow. Severe vasospasm sufficient to constrict this region of the aorta, especially one involving any appreciable length, might therefore cause a sharp increase in resistance to aortic flow, and in the absence of any effective alternate pathway for blood diversion, the pressure generated could become very high indeed. Under such circumstances, the resistance
offered by the degenerated and atrophic media might very well be less than that through the constricted aortic lumen, and by this means the process of intimal tearing and aortic dissection could be set in train.

Even where side branches do happen to exist proximal to any region of vasospasm (e.g. in the descending thoracic aorta), the alternative pathway they offered for blood diversion might not always be enough to relieve the situation, so that high proximal pressures could again be generated. This could be especially so in cases associated with chronic hypertension, because this tends to cause structural arteriolar changes which would limit the capacity of any organ to decrease its flow resistance and so provide a high conductance pathway for blood flow diversion in these circumstances.

What then is the evidence for this mechanism in aortic dissection? Here, as elsewhere, evidence is often influenced by interpretation, and because aortic dissection is not usually thought of in terms of vasospasm, the data have not really been analysed in its light. Nonetheless, much of the evidence is at least consistent with the view. The clinical observation that a marked heaving of the upper sternum is often seen in patients with aortic dissection is quite in keeping with an aortic obstruction at some more distal site, for example. Radiologically, too, the aorta often appears narrowed, consistent with local obstruction to aortic blood flow at the site. Whether that obstruction is due to vasospasm, as proposed here, many might doubt, but actually aortic dissection is difficult to explain in any other way. It is certainly hard to account for merely on the basis of the blood found within the arterial media. Just on common sense grounds, it seems highly unlikely that any haemorrhage into the aortic wall could generate sufficient force per se to overcome the very pressure within the underlying aortic lumen that caused it. The pressure propelling any dissection would have to be at least as high as that in the aorta before there would be even a tendency for the luminal aspect of the related aortic wall to collapse, and this would seem impossible without some additional factor. Even if an intimal tear communicating with the lumen acted as a sort of one-way valve, it would only allow blood to enter the aortic wall with each heart beat if the systolic pulse pressure wave though the area of dissection were somehow out of phase with that traversing the aorta. Since the source of blood driving the dissection down the layers of the aortic wall is thought by most to stem from the adjacent aortic lumen itself this, too, seems unlikely — unless, of course, a segment of aortic vasospasm existed between the region of the intimal tear and the distal area of radiological aortic narrowing, which would make my point about the importance of vasospasm in any case.

In general, then, the evidence on dissecting aneurysm seems perfectly in keeping with the present thesis, namely that one way or another, vasospasm is important in pathogenesis. And the simplest way of seeing this is to say that the narrowed area observed radiologically is itself entirely due to aortic vasospasm. Testing this will depend, among other things, on demonstrating a regression of the area of aortic narrowing with vasodilators, but the collection of such information will itself be dependent on the idea being entertained in the first place, and this is what the present discussion is aimed to stimulate.

Of course, many will think this suggestion for the aetiology of aortic dissection far fetched, even absurd, but there it is. That is my interpretation of the evidence.

**DISEASES OF MEDIUM-SIZED ARTERIES**

Many of the processes which affect the larger arteries no doubt lead to similar disease in the middle-sized ones, especially those with vasa vasora, but in addition there may be rather special consequences at this level. In particular, any process, either structural or functional, that tends to narrow arterial lumen size — be it vasospasm, thrombosis, atherosclerosis, or a combination of all three — will produce much more encroachment on lumen diameter in the medium-sized arteries, and therefore a much greater tendency to occlude the artery concerned. This has important implications for ischaemia in organs of supply. The medium-sized arteries may differ in other ways. For one thing they are more muscular, and so can probably constrict to a greater degree for any given stimulus. We certainly know from the studies in migraine, subarachnoid haemorrhage, Raynaud’s disease, and Prinzmetal’s angina that quite severe vasospasm can
and does occur in vessels of the cerebral, digital, and coronary artery size. Severe vasospasm may of course be unusual, but lesser, perhaps subclinical, degrees occurring, say, in response to psychological or physiological stresses, might be much commoner, and could be just as serious in the long run through insidiously contributing to the gradual build-up of an atherosclerotic obstruction at the site in the way outlined in chapter 4. And once the atherosclerotic plaque had begun to encroach significantly on lumen diameter, any even minor subsequent bout of vasospasm there could seriously compromise whatever of its patency remained, and so interfere with organ blood flow. Also, if as well as being severe, any such episode of spasm were prolonged, two other factors might determine that the lumen did not return to its original size after the episode had passed. The first is thrombosis occurring at the site of vasospasm as a result of focal endothelial damage at the point of narrowing (see chapter 3). The second is ischaemic plaque damage and swelling from haemorrhage, necrosis, oedema, etc., as a result of a prolonged restriction/distortion of its vasa vasorum blood supply during the vasospasm (see chapter 4). By either of these means an episode of more long-lasting vasospasm might well produce a much less reversible ischaemia in the organ of supply than the initiating event would have determined.

1. Transient focal cerebral ischaemic attacks (TIAs)

Transient focal cerebral ischaemia, or TIA, is a syndrome where recurrent reversible, usually brief focal cerebral symptoms occur from episodic cerebral ischaemia, with a marked tendency to involve the same region of the brain on each occasion producing, for example, recurrent attacks of monocular blindness (from involvement of the ophthalmic artery), or recurrent dysphasia and/or hemiparesis (from ischaemia in the middle cerebral artery territory). The particular form of this general syndrome I wish to consider here is recurrent, episodic, focal cerebral cortical ischaemia occurring within the distribution of the internal carotid arterial supply, and related to atherosclerotic disease in it just above its point of origin in the neck. This is a common form of the condition whose importance lies in a real potential for reversibility after removal of the carotid artery plaque. Briefly stated, the traditional view is that the attacks arise largely when thrombi, formed on a roughened atherosclerotic plaque narrowing at the internal carotid artery origin, are swept off from time to time by the passing blood stream and so embolise distally to its intracerebral branches. But there are many aspects difficult to account for on this basis. In particular, as with coronary artery disease, there is a very poor clinico-pathological correlation between the degree of narrowing at the internal carotid origin and the amount of symptomatic disease, with some patients having gross yet completely asymptomatic narrowing, whilst others suffer severe, frequent clinical attacks without much reduction in lumen diameter at this point at all. Be that as it may, there is little doubt that atheroma at the internal carotid origin is somehow related to the pathogenesis of symptoms, even in cases where very little structural narrowing is observed, because even there patients may still benefit from surgical removal of the carotid plaque. The most accepted explanation for this clinico-pathological discrepancy is that the amount of stenosis itself is less important than the degree to which the plaque is ulcerated, and therefore able to provide a focus for thrombus formation. When this is the case, we need to know how the plaques become ulcerated in the first place. And whilst it is true that ulcerated atheromatous plaques are often found at operation, the area is not infrequently entirely smooth, and not at all what one might expect if this mechanism were to hold in any general way. We then must look further to explain how thrombosis occurs in the first place, and why at one particular point in time.

The clinico-pathological disparity between the degree of atherosclerotic narrowing at the internal carotid origin and the severity of clinical symptoms could be accounted for by a role for vasospasm at the site. And as in myocardial infarction, such vasospasm could also initiate wall injury and secondary thrombosis on the atherosclerotic plaque, so as to provide the source for the emboli which we know do occur in at least some transient ischaemic attacks. By this view, thrombus formation on relatively smooth plaques would be a consequence of vasospasm-induced haemodynamic forces causing intimal damage at the site (see Ch. 4).
plaques, vasospasm of the arterial smooth muscle coat would have the effect of choking their vasa vasorum blood supply, so as to cause ischaemic damage or actual necrosis (see Ch. 4), depending on the duration of the vasospastic episode.

As yet, there is very little direct evidence on which to evaluate the mechanism proposed, because it has not been seriously considered so far. But the data available are quite consistent with it. The internal carotid is a muscular arterial system, well-supplied with sympathetic nerves, and is well-recognised to constrict under appropriate stimulation. Indeed, we have known since the time of the early angiographic studies in subarachnoid haemorrhage and migraine that the major branches of the internal carotid artery in man are quite capable of active constriction. Yet only in relatively recent years has it been recognised that their propensity for doing so may be more than uncommon. It has been stressed that migraine should be high on the list of differential diagnosis in patients presenting with transient recurrent cerebral ischaemic attacks, but perhaps the reason for this is that their pathophysiological basis, as well as their clinical features, are much more related than we think.

Another difficulty in assessing the present vasospastic theory is the absence of direct data comparing the appearance of the internal carotid artery origin during attacks versus in between. This is true both for angiography and for the less invasive methods of investigation such as ultrasound Doppler scanning (Weaver et al., 1980). But as the resolution of imaging of these non-invasive techniques improve and are developed further, repeated studies of this type should become relatively simple. It will then be important to bear the current vasospastic mechanism in mind, particularly because of the potential it carries for reversibility and prevention, not only of the transient ischaemic attacks themselves, but of the subsequent strokes which all too often follow in their wake.

2. Stroke.

I want to confine attention here to cerebral infarction and, moreover, to that form of it which occurs in relation to atherosclerotic disease at the internal carotid origin, and follows reversible transient clinical episodes of the type discussed. One striking feature here is a remarkable similarity in pattern to that form of myocardial infarction which follows recurrent episodes of unstable angina (Ch. 3). Anatomically, the final irreversible episode in both stroke and heart attack tends to affect the very same region of the organ as the more transient attacks which preceded them. Pathologically, the final outcome in both events is infarction, most often occurring in relation to atherosclerotic disease in their artery of supply, and usually associated with superimposed thrombosis during the final episode. This clinico-pathological parallel must prompt one to ask whether there is also an analogy of mechanism, and this is indeed what I propose. Just as with myocardial infarction, it seems to me that the most likely cause of stroke occurring in this context is a final and less reversible episode of vascular obstruction precipitated by vasospasm at the site of internal carotid artery origin already narrowed by atherosclerosis. Again, as in myocardial infarction, such vasospasm may not be solely responsible for the eventual degree of internal carotid occlusion, because secondary plaque swelling (from vasospasm throttling the vasa vasorum supply to the plaque), the super-imposition of thrombus, and even subsequent distal thromboembolism could all play a role in that. Because the brain is so exquisitely sensitive to even transient deprivation of blood supply, it is even possible on occasions that irreversible neuronal damage or infarction might occur merely from a more prolonged and severe episode of (reversible) vasospasm without the superimposition of secondary plaque complications or thromboembolism at all. The point to be stressed though, is the possible importance of vasospasm as a potential initiator of the less reversible episodes of arterial occlusion, and therefore of irreversible cerebral infarction in the territory beyond.

What is the evidence to support this view? Although neither collected in this context nor interpreted in its light, much of the data is consistent with it. The clinico-pathological parallel with myocardial ischaemia is also very close, and in that area the evidence favouring a vasospastic basis is, I believe, strong (see Ch. 3). Clinically, one of the only real differences in pattern between these two syndromes is the much later average age of onset of stroke. This may relate to the
larger diameter of the internal carotid artery necessitating a longer time for subclinical vasospasm to build up atheroma to a critical level of occlusion, so that even this difference may not be as important to mechanism as may first seem. Radiologically, too, the evidence in the acute phase of stroke is perfectly consistent with the mechanism proposed. Thus, Bladin found a relatively high incidence of internal carotid artery narrowing and/or occlusion in the first 24 hours after the onset of acute stroke.25. Primary vasospasm with super-imposed thrombosis could account for this, and vasospasm of the internal carotid artery undoubtedly occurs in at least some circumstance, such as ergot overdosage.26

If internal carotid artery vasospasm is important in TIA/stroke in the way suggested, we should ask just why it should occur so commonly at the internal carotid origin. What may be important in this context is that this is the area of the carotid sinus, known to be concerned with the reflex sympathetic regulation of cerebral blood flow through an effect on general systemic blood pressure. It will therefore be of interest in future to learn whether this area can undergo constriction directly by, say, efferent sympathetic stimulation during baroreceptor adjustments to postural change. Such a mechanism would certainly help regulate cerebral blood flow directly in response to posture, independent of general systemic arterial pressure.

3. Intermittent claudication

This might also be due to constriction of the femoral artery - as originally suggested by Osler27 - particularly perhaps in those patients who develop pallor of the leg and absent foot pulses only during exercise. The traditional explanation for this phenomenon is that a decreased arteriolar muscle bed resistance occurs during exercise and causes shunting of blood away from the superficial vessels. But this is by no means established as a cause, and is hard to reconcile with the view that pain in intermittent claudication is due to the development of muscle ischaemia during exercise. Vasospasm of the femoral artery might explain a great deal in this situation, whether superimposed on focal atherosclerotic narrowing of the femoral artery or not. Certainly there is enough evidence to warrant further investigation of this as a possible mechanism. Lorelius and colleagues, for example, have noted a high incidence of femoral artery constriction at angiography in patients with intermittent claudication, even in the apparently healthy limb,28 suggesting that there may well be some underlying predisposition to vasospasm in this condition.

As it happens, the tendency for femoral and other arteries to constrict has been known since the early days of angiography.29 The appearances were initially interpreted as being "stationary waves", but we now know that they are reversible with vasodilator drugs,30 so there can be little doubt as to their constrictive nature.

4. Other arterial thrombo-embolic phenomena

Arterial vasospasm might also be responsible for many of the other so-called thrombo-embolic phenomena occurring in larger arteries, and for those 'spontaneous' embolic events where there is no obvious source for the embolus. Further, even where an origin for embolism is found, such as in atrial fibrillation, the completeness of obstruction at the site of clot lodgement might also be related to vasospasm, both from a local irritant effect as with catheter-induced spasm31 and from the local release of vasoconstrictor prostaglandins such as thromboxane A2 from the embolus.32 Obviously, one could see ischaemic events in many other organs from this same standpoint, for example abdominal angina/infarction in relation to atherosclerotic narrowing and/or vasospasm of the superior mesenteric artery. My intention here, though, is not to carry the iteration of this mechanism to the extreme, but merely to establish the need for its further investigation.

5. Cerebral berry aneurysms.
As with aneurysms of larger arteries, the pathogenesis of the small so-called ‘berry’ aneurysms along the distribution of the internal carotid artery branches is ill-understood. They are most often found at major branch points, where the medial coat is often relatively defective, but this is certainly not enough to account for them alone, because similar defects often occur in normal subjects. In my view, as with larger arteries, the pathogenesis of these smaller aneurysms may be related to vasospasm. The reason for raising this is that there is indeed some evidence for it. The relevant observations have been made at the time these aneurysms rupture and cause subarachnoid haemorrhage, when arterial spasm is not infrequently observed at angiography near the aneurysmal site. It has been noted in two separate phases. The more delayed one is generally considered to result from the accumulation of some product of altered blood on the adjacent blood vessel wall, but the low incidence of vasospasm in other causes of subarachnoid haemorrhage such as trauma raises serious doubts about this explanation. Also, it is very difficult to explain the initial phase of vasospasm on this basis.

It seems to me that a great deal more might be explained if we viewed the relationship between aneurysm and local vasospasm the other way around, viz. that repeated bouts of local segmental vasoconstriction could well set the stage for aneurysm formation by causing ischaemic atrophy and weakening of the media, through compression of its vasa vasorum circulation. Eventually, rupture might well occur when a vasospastic episode caused complete ischaemic breakdown of an already weakened and attenuated arterial wall, so as to produce the final subarachnoid haemorrhage which brings the whole matter to medical attention.

6. Fibromuscular dysplasia of the renal and other arteries.

Fibromuscular dysplasia, or hyperplasia, occurs in a number of arteries including the renal, superior mesenteric, and carotid. In the renal artery, it may give rise to renal artery stenosis or narrowing, the importance of which lies in the fact that it causes a surgically remediable form of secondary hypertension. Despite inroads made into understanding mechanisms in this form of hypertension however, and despite the good documentation of its various pathological forms, we know remarkably little about the pathogenesis of the underlying dysplastic process. It seems to be an acquired condition, more common in females, and when it occurs in the carotid arteries, may do so on a background of high intake of the vasoconstrictor, ergot, and/or oral contraceptive therapy.

The aspect which particularly struck me in the context of the present theory was the close similarity between the somewhat ‘beaded’ appearance of fibromuscular dysplasia at angiography and the appearances of segmental arterial vasospasm. In view of the clinical history of high ergot intake in some patients with carotid fibromuscular dysplasia, it is of particular interest that acute ergotism also may have an angiographic appearance similar to fibromuscular dysplasia in both the carotid and renal arteries. This parallel has also been noted by others and raises the question of a possible casual relationship, but so far there has been no clear exposition of mechanism. Let us now examine this.

Fibromuscular dysplasia is a disease which predominantly affects the medial and adventitial layers of arteries and so, as with aneurysm formation, it is possible that the vasa vasorum blood supply to these layers is somehow involved in the process. Indeed, fibromuscular dysplasia is often associated with aneurysm formation at some of the sites, so a similarity of mechanism is well worth considering. Perhaps the odd thing is that whereas aneurysms normally occur at areas of relative atrophy or weakening of the arterial wall, the process we are dealing with here is more often associated with hyperplasia or local tissue overgrowth. Yet this may merely represent a different reaction on the part of the renal artery to vasospasm, as well perhaps as a different time-sequence to the vasospasm itself. Thus, whilst a prolonged and severe episode of vasospasm might tend to cause atrophy of the adventitia and outer media from severe ischaemia interfering with the vasa vasorum blood supply, less-severe and less-prolonged vasospasm could still result in damage to the wall short of cell atrophy, and this could lead to hyperplasia during the regenerative phase of cell repair, i.e. after the vasa vasorum blood flow had been restored. In this
way, the overgrowth of tissue occurring in the outer layers of the arterial wall might be explained.

Fibromuscular dysplasia sometimes involves the inner media and intima of the artery as well, and we know that this is beyond the normal reaches of the vasa vasorum circulation. However, as with atherosclerosis, these inner wall changes might well be the more direct result of local increases in arterial blood flow velocity and secondary intimal damage during vasospastic episodes (see also Ch. 4).

The present discussion has raised the possibility that vasospasm may lead to both atrophic and hyperplastic conditions, a point I shall take up again below. As to pathogenesis of the vasospasm, I view it likely due to sympathetic nervous activation from the physiological and psychological stresses of everyday life. Of course, nothing in this respect is proven.

I turn now to the smallest branches of the arterial tree, namely, the arteriolar resistance vessels.

**DISEASES OF THE SMALL ARTERIES AND ARTERIOLES**

**The Pathogenesis of Hypertension**

Here, I want to put forward the idea that what arterial vasospasm is to atheroma in larger arteries, so may constriction of arterioles be to the chronic arteriolosclerotic process of hypertension in smaller ones.

Whatever the case for vasospasm in large arteries, there is no doubt that the arterioles constrict. Indeed, it is well-established that their constriction is the main determinant of resistance to blood flow within the circulation, and therefore the major factor responsible for variation in both local blood flow within different vascular beds, and the overall level of systemic arterial blood pressure. Because they constrict, it is certainly worth examining the proposition that, as with larger arteries, they may do themselves some haemodynamic harm in the process.

**Background**

Functional arteriolar constriction leads to an elevation of arterial pressure, and this continues for as long as the constriction is maintained. This simple mechanism has provided the whole basis for our understanding of the pathogenesis of hypertension both in man and experimental animals. Because of this, we normally expect to find a clear correlation between whatever nervous or humoral activity is thought to be driving the arteriolar constriction, and the degree of blood pressure elevation observed. This certainly works well enough in some circumstances, but there are problems in others. For example, in my own special area of research interest, experimental renal hypertension, it has emerged fairly clearly over recent years that elevated levels of the vasoconstrictor renin-angiotensin system may well be enough to account for the high blood pressure in the acute phase of hypertension, but the same cannot be said of the chronic phase, where all evidence for involvement of the system may eventually become lost. And this enigma has not been resolved by the extensive search for other renal pressor materials, nor by the discovery of more subtle ‘inactive’ forms of rennin. I believe that an important clue to solving this riddle lies in looking at the different forms of chronic hypertension as a group, because what is striking from the literature there is that when high blood pressure from any cause has been maintained for long enough, it tends to remain so, and to a large extent become independent of that cause, whatever its original nature. This has been observed with hypertension as diverse as experimental mineralocorticoid (DOCA) hypertension, primary aldosteronism in man, experimental and human renal hypertension, as well as psychologically-induced hypertension in mice. It is difficult to believe that some common humoral factor has taken over the major driving role in pathogenesis in all of these very different forms of
hypertension, particularly when blood pressure may remain elevated even after the basic underlying abnormality (e.g. adrenal tumour in Conn’s syndrome) has been removed entirely.  

If a relative autonomy of the chronic phase of hypertension is common to all or most of its various forms, there would seem likely to be some common factor underlying it. Now, the only evident candidate for this role is the high blood pressure itself. Because of this, it is reasonable to suggest that the elevated blood pressure per se may well be responsible in some way for changes which eventually contribute to its own maintenance. This view that acute hypertension begets chronic may sound a little odd, but there is indeed some evidence in favour of it. Thus, keeping blood pressure relatively normal with drugs in both steroid hypertension and spontaneously hypertensive rats, delays the onset of the whole hypertensive process, including hypertensive vascular change, until well after the treatment has been withdrawn. This finding is unlikely to be due merely to some side effect of drug therapy, because the same observation has been made with many different drugs, and even where blood pressure has been reduced by non-drug means. We are left to conclude, then, that the high blood pressure itself, or at the very least some factor closely associated with it, is likely responsible for the very changes which come to drive the chronic maintenance phase of the whole hypertensive process, regardless of its origins.

Folkow saw this clearly from his experiments on the spontaneously hypertensive rat, and went on to postulate that the hypertension there might well contribute to its own eventual maintenance by gradually causing a structural arteriolar narrowing that eventually came to replace the earlier functional arteriolar constriction as the major cause of the increased peripheral resistance to blood flow. If we extend this view to the more general field of chronic hypertension, it becomes apparent that the relatively ‘fixed’ organic arteriolar narrowing or arteriolosclerosis, which is such a striking feature of chronic hypertension, might well become the major determinant of the high blood pressure in any chronic situation. Then, we can conclude that hypertension begets hypertension, i.e. initial hypertension leads to a structural arteriolar narrowing which eventually becomes responsible for maintaining the elevated peripheral resistance in chronic hypertension phase in its own right. This approach helps resolve a number of problems, for example the long-standing argument between clinicians and pathologists on whether arteriolosclerosis is the cause or the effect of hypertension. From the present viewpoint, both would be right, because hypertension initially due to functional arteriolar constriction from any cause would gradually lead to secondary structural arteriolar narrowing, which would slowly take over as the main cause of the increased peripheral resistance to blood flow, and hence as the major determinant of the chronic hypertensive state. In the past, this argument has become bogged down because of our tendency to view cause and effect in a much too static way, and the more dynamic view helps to resolve, it, viz. what is initially the effect of the hypertension later becomes its cause, as the process of arteriolosclerotic structural narrowing gradually comes to replace earlier functional arteriolar constriction as the factor determining increased circulatory resistance to blood flow.

Largely because of Folkow’s work, there is little reason now to suppose that this view should be especially controversial. But less well-established is the mechanism by which an early elevation of blood pressure brings about these chronic structural arteriolar changes. I would like now to examine this.

Folkow’s suggestion is that structural arteriolar changes result from a process of physiological ‘adaptation’ or hypertrophy of the smooth muscle cells within the medial coat of the arteriolar walls to the higher pressure, the stimulus for this being an increase in circumferential wall tension. However, there are a number of difficulties with this view.

First, in the usual case where the early phase of hypertension is associated with arteriolar constriction, the relationship between elevated blood pressure and arteriolar circumferential wall tension must be somewhat ambiguous. This is because wall tension is determined Laplace’s law, and is therefore not only directly proportional degree of elevation of the blood pressure itself, but inversely proportional to arteriolar radius. Elevated blood pressure per se will indeed increase wall tension, but the associated arteriolar constriction will, in its own right, have just the opposite effect. In the proximal arterioles, it may well be that the high blood pressure effect overcomes that of any reduction in arteriolar radius, but this will become less so as we move down the length of these resistance arterioles towards the capillaries, because intraluminal pressure will gradually become lower and lower in the process. As a consequence, blood pressure in the more distal arterioles
may not be much elevated at all, and if these distal vessels, too, are constricted, the stimulus proposed by Folkow becomes very weak there indeed. And because chronic arteriolosclerosis in hypertension affects arterioles right down to those of even smallest size, this does raise a significant difficulty with Folkow’s view.

Second, chronic hypertension in spontaneously hypertensive rats probably arises on the basis of intermittent elevations of blood pressure, so that the average 24 hour wall tension load even on the proximal arterioles is unlikely to be very high.

Third, at the more theoretical level, there is a difficulty in understanding just how any truly physiological response in the arterioles could become irreversible after the initiating stimulus had been removed.

Structural arteriolar changes in hypertension

The most important objection to increased circumferential wall tension as the stimulus to structural arteriolar narrowing concerns the histological nature of the arteriolar changes themselves. These include not only medial smooth muscle cell hypertrophy (and hyperplasia) which might fit as a physiological response to an increased pressure, but increased medial (and adventitial) connective tissue, as well as intimal thickening from smooth muscle cell proliferation and hyaline deposition there, and these are much more difficult to reconcile with Folkow’s views. The intimal changes are often very prominent indeed, and may, moreover, be markedly eccentric in location around the arteriolar lumen, a feature not at all explained by any alteration of circumferential wall tension, which should be distributed uniformly around the lumen.

As I see it, the structural arteriolar changes of chronic hypertension look much more like a pathological response to damage in the hypertensive arterioles than any physiological process, and by analogy with the mechanism postulated in chapter 4 relating atherosclerosis to vasospasm in larger vessels, I suggest that arteriolosclerosis results from endothelial damage occurring during bouts of arteriolar constriction because of a similar increase in shearing stress on their walls. There are certainly histological similarities between atheroma and arteriolosclerosis, including an intimal thickening from increased connective tissue and smooth muscle cell proliferation, often eccentrically located around the vessel lumen. Such changes would fit well with an arteriolar reaction to endothelial damage, and I suggest that this indeed occurs as a consequence of an increased arteriolar blood flow velocity, and therefore wall shearing stress, at sites of arteriolar constriction. Once endothelial integrity has been breached, it is further postulated that plasma and platelet factors insude into the underlying arteriolar wall and thereby stimulate the proliferation of smooth muscle cells and the general increase in intimal ground substance, collagen, elastin etc. which is so characteristic of the chronic hypertensive state. Of course in larger arteries the atheromatous changes which follow endothelial damage are largely confined to the intima without much involvement of the media at all, but in the smaller arterioles the media layer is neither as precisely demarcated from the intima, nor as far from any breach in the endothelium, so that medial smooth muscle cell proliferation and connective tissue overgrowth there might also be explained on the basis proposed.

The view that chronic arteriolosclerosis could be a pathological response to local wall damage is unlikely to be particularly controversial — indeed some might see any distinction from more traditional “structural adaptation” theories as somewhat semantic — but there will be stronger reservations about how this damage is caused, even by those who take the point that Folkow’s theory does leave some discrepancies. In this context, there are two potentially important objections which may be raised against the present view. The first centres on whether arteriolar constriction of a degree we might anticipate in moderate hypertension could cause enough increase in local arteriolar blood flow velocity to damage the endothelial linings in the way proposed. At face value, this is an important objection, because any uniform constriction along the whole arteriolar length would give haemodynamic changes in accordance with Poiseuille’s law, whereby an elevation of mean arterial pressure of, say, 50% (corresponding to a 50% increase in arteriolar resistance) would be brought about by narrowing the arteriolar lumen diameter as little
as 10%, and this corresponds to a mere 20% increase in arteriolar blood flow velocity. This would hardly be enough to damage the arteriolar wall. But this calculation makes the assumption that arteriolar constriction in hypertension is in fact uniform along the length of any arteriole, and this is unlikely to be valid. Certainly, published experimental studies so far have shown that arteriolar constriction is often extremely irregular, for example during the high-dose infusion of various vasoactive substances\(^59,60\) and in the malignant phase of acute experimental hypertension\(^61\) where localised irregular, even "sausage-string" constrictions are frequently seen along arterioles in a manner not dissimilar to acute segmental vasoconstriction in larger arteries in man.\(^28\) Now if, as seems likely, such irregular arteriolar constriction also occurs in moderate hypertension, including human hypertension, then our whole concept of the haemodynamics of arteriolar blood flow there will have to be altered. This is because resistance in any vascular segment is proportional to its length, so that short local constrictions may make very little difference to resistance overall along the arteriole. In those circumstances, since arteriolar blood flow would be relatively unchanged, flow velocity would increase in inverse proportion to the square of the radius at any constricted point (see Ch. 4). Moreover, to now obtain, by focal constriction, the same 50% increase in overall resistance along any arteriole - as in the example quoted above - would require very severe (localised) narrowings indeed, with even sharper increases in local arteriolar flow velocity and wall shear stress at the more constricted sites. If this mechanism were to apply in human and experimental hypertension, we would expect to find an irregular distribution of arteriolosclerotic changes along the length of arterioles. And the available data do seem to indicate irregularity of arteriolar narrowing in chronic hypertension both in the rat,\(^62\) and in man where one of the hallmarks of long-standing hypertension is variation in lumen calibre along the only arterioles.\(^63,64\)

A second potential difficulty with the present theory (and indeed with any other) is that at least in severe hypertension, any evidence of arteriolar damage is not usually found in constricted segments but dilated ones,\(^59,60\) and this has given rise to the view that damage occurs at sites of arteriolar stretch. If we ignore for the moment the fact that any such dilatation is often more relative than absolute,\(^65\) we might accept this as an explanation of some of the arteriolar changes in the severe or malignant phases of hypertension, but it could hardly account for the problem we are analysing, namely moderate hypertension, because dilatation is not usually found there at all.\(^61\) In any event, the only damage which could be relevant to structural increases in peripheral resistance in chronic hypertension is that which occurred in narrowed or constricted segments. I suggest the following explanation for this apparent paradox. By analogy, again with the studies in experimental atherosclerosis,\(^58\) it seems likely that any arteriolar wall reaction to endothelial damage will be mediated by the insudation of plasma and platelet factors through an area of impaired endothelial integrity. It is also likely that, in contrast to the shearing forces acting to cause this endothelial breach, the component of pressure driving plasma etc. through the arteriolar wall will be the lateral pressure perpendicular to the direction of blood flow. Now Bernouilli’s principle determines that, at any given rate of volume flow, this lateral pressure will be at a minimum at the more narrowed sites. This means that damage done at sites of focal constriction during episodes from an increase in shearing stress, may not become manifest as such until after the vessel has subsequently relaxed and allowed the lateral pressure to build up and force plasma etc. through the endothelial breach. According to this view, damage to the arteriole would be done during bouts of constriction right enough, but would not become evident in terms of histological change within its wall until after the vessel had subsequently relaxed.

The present hypothesis, then, provides an explanation for the long-standing observation that chronic hypertension from any cause may eventually become irreversible when widespread arteriolosclerosis has supervened.

This view might also explain why, on occasions, all trace of the initiating cause of hypertension can disappear with time, because when pathological arteriolar damage eventually comes to replace physiological constriction as the cause of the increased peripheral resistance, the degree of arteriolar narrowing will no doubt become much less well attuned to the initiating stimulus, and therefore perhaps ‘inappropriately’ high as a result. In this way, structural arteriolar
changes could eventually come to set blood pressure at a level much higher than would have been determined by the initiating vasoconstrictor stimulus. Since we have no reason to believe that this vasoconstrictor stimulus would then be in any way divorced from usual negative feedback mechanism controlling it, there would now be a tendency for it to be reflexly suppressed. As pointed out by Folkow, there must be powerful negative feedback mechanisms inhibiting the initiating stimulus if stable hypertension is to be maintained. Such a process might well explain the difficulty of finding any evidence for a continuing initial stimulus in the more chronic phase of any hypertensive process.

We have now discussed the case where focal constrictions within an arteriole happen to be severe enough to cause a resistance increase overall along the arteriolar length. But even short of this, quite minor degrees of focal constriction, whilst contributing little to overall arteriolar resistance, could still cause a sharp local increase in arteriolar blood flow velocity at the more constricted points (again inversely proportionate to the square of the radius as the site), and this could perhaps account for two other odd puzzles of hypertension research. The first is to explain just how essential hypertension that begins, in haemodynamic terms, as an increase in cardiac output in early life, ends up being sustained entirely by an elevated peripheral resistance. I suggest that, there, subtle arteriolar constriction does occur early in these circumstances, and that whilst this contributes little in resistance terms to the initial degree of hypertension overall, it nonetheless leads to important alterations of local haemodynamics and secondary arteriolar structural changes which eventually do. Thus, irregular focal constrictions along any such arteriole, even if too minor to change its resistance overall, could still cause a sharp increase in local blood flow velocity at the more constricted points proportionate to the inverse square of radius of constriction. Furthermore, in the absence of any resistance change along its length overall, such an arteriole would still take its share of the increased cardiac output, and this would mean that blood flow velocity at constricted sites would be even higher. Under these circumstances, damaging local increases in arteriolar blood flow velocity might still be achieved, so that despite the absence of any overall change in total peripheral resistance, structural arteriolar change would again begin to develop. As blood pressure gradually began to rise to perhaps 'inappropriately' high levels in consequence, reflex mechanisms would be brought into play tending to reduce cardiac output back towards normal in compensation. In this way, a largely structural increase in peripheral resistance could gradually come to replace high cardiac output as the haemodynamic factor driving the hypertension, and the whole time-course of the haemodynamic profile of essential hypertension would be explained.

Essential Hypertension in Man

Spontaneous hypertension in the rat and essential hypertension in man both have a familial background. In both, it is also becoming clearer that the early phase of hypertension is of a labile nature, where blood pressure is significantly elevated only in episodes. In this respect, one of the early haemodynamic abnormalities in the spontaneously hypertensive rat is an increased blood pressure response to 'alerting' stimuli, or an exaggerated 'defence' reaction. There is similar evidence of an exaggerated response to various stresses, both physiological and psychological, in early human hypertension.

The general hypothesis of this chapter could well explain how structural arteriolar changes might come about from intermittent pressor episodes. Certainly, in larger vessels, local constriction for less than an hour is enough to produce local endothelial injury, and there is no reason why the same should not hold for the smaller arteries and arterioles. Episodes of blood pressure elevation in early human hypertension are most often related to increases in cardiac output, with only subtle changes in peripheral arteriolar resistance. But, as have seen, even-minor resistance changes from focal arteriolar constriction could have very large local haemodynamic consequences, and could account well for the gradual onset of arteriolosclerosis, the steady rise in 'basal' blood pressure and the eventual increase in peripheral resistance so characteristic of the more chronic hypertensive state. Of course, the amount of arteriolar damage (e.g. wall oedema) resulting from each pressor episode might not be great, and even that which did occur might in
large part be reversible once the episode had passed. But, allowing for the long time-course of development of human hypertension, even the most minimal persistent changes would begin to add up significantly over the years. Thus, a rise in baseline blood pressure of no more than .01 mm Hg/day from damage secondary to labile episodes like this would produce an elevation of 35 mm Hg over a ten year period — a relatively short time in the life history of essential hypertension in man.

The accumulating data is that human essential hypertension is primarily neurogenic in nature, via an increased sympathetic nervous system activity. There is also increasing evidence that this, in turn, is driven by psychological stress, though this is of course controversial as is any discussion about the relationship between stress and disease. Part of the problem may again be one of looking too late for the initiating mechanism, and at a stage when secondary arteriolosclerotic changes have supervened and largely come to suppress the original cause. In keeping with this, when patients have been studied in the earlier phases of essential hypertension, evidence for increased sympathetic nervous activity and increased circulating catecholamines has indeed been obtained, particularly when studies have been performed during provocative stress-testing.

Severe Essential Hypertension

In the less severe phases of hypertension, constriction is probably fairly well confined to arterioles. But when the driving stimulus becomes more severe, it could well begin to involve the small arteries as well. In that case, one consequence would be a reduction in blood pressure to the more distal arteriolar resistance sites in various organs. In the kidney, this would tend to cause a reflex arteriolar dilatation, determined by the requirements of autoregulation of renal blood flow, and we know from the work of Eide and colleagues that this would stimulate renin release from the afferent arteriole. Now, once in general circulation, this renin would tend to enhance the degree of general peripheral constriction and so raise blood pressure still further, thus contributing to the pathogenesis of this phase. It is of interest in this regard that Mohring and colleagues (1976) found that suppression of the vasoconstrictor renin-angiotensin system in the ‘accelerated’ phase of hypertension in the rat by the administration of salt was often associated with a striking improvement in the ‘malignant’ nature of this condition.

Whatever the case, there is now little doubt that the renal sympathetic nerves are activated in essential hypertension, probably especially so in its more severe forms, where it has recently been demonstrated that renal denervation can result in a significant decrease in blood pressure.

Salt and essential hypertension

A debate has raged for decades over whether or not high salt intake is important in the pathogenesis of essential hypertension, with Pickering, for example, taking a strong stance on behalf of the negative. However, based on studies in epidemiological studies, Sever and Poulter have proposed that the increased sympathetic nervous activity in hypertension causes, in the kidney, a problem in salt excretion, so that, in contrast with normal subjects, salt may indeed contribute importantly to hypertension pathogenesis.

Arteriolosclerosis and Ageing

Because hypertension is a disease of quantity, i.e. one which overlaps the blood pressure distribution of the normal population as a continuum, it seems likely that the changes we have been discussing will also occur to some extent even in normal people. Indeed there is no doubt that arteriolosclerosis does occur with increasing age in the normal population. Viewed in this light, perhaps even the normal ‘ageing’ process in arterioles has a similar pathogenesis to that
discussed above, albeit at a lesser level, i.e. as a response to episodic focal vasoconstriction from daily physiological or psychological stresses. If this does occur, it would also tend to reduce local blood flow to the organs of supply, and to this extent such a process might be very relevant to the gradual deterioration of organ function or organ 'reserve', with age. What is usual in the population in this respect need not necessarily be physiological, and indeed by this view would be caused by a pathological response to repetitive minor damage of the arterioles. This process could therefore be important in the condition of normal ageing in man and in other species endowed with a vascular circulation (see also Ch. 11).

CONCLUSION

None of the suggestions here can be regarded as proven, and I certainly have no wish to destroy the credibility of the general case by any special pleading beyond reasonable limits. But I do believe that the mechanisms outlined are well worth considering, particularly in view of the hope they offer for reversibility and prevention of much vascular disease. Of course, analysis of the prevalence and the pattern of arterial constriction in man, particularly that which might occur in response to stress, is difficult given current invasive techniques of study. But less invasive methods such as ultrasound Doppler scanning are now in the process of improvement, and repeated imaging of arteries in conscious man under a variety of circumstances should then be possible. My own suspicion is that when we are able to do this, we will be surprised by the degree of arterial constriction that can occur even in normal man. But of course only time and experiment will tell.

References