THEOREM 0.6. *Suppose that ρ_0 is irreducible and satisfies the hypotheses of the conjecture, including* (I) *above. Suppose further that*

(i) $\rho_0 = \mathrm{Ind}_L^{\mathbf{Q}} \kappa_0$ *for a character κ_0 of an imaginary quadratic extension L of $\mathbf{Q}$ which is unramified at p.*

(ii) $\det \rho_0|_{I_p} = \omega$.

Then a representation ρ as in the conjecture does indeed come from a modular form.

This theorem can also be used to prove that certain families of elliptic curves are modular. In this summary we have only described the principal theorems associated to Galois representations and elliptic curves. Our results concerning generalized class groups are described in Theorem 3.3.

Start Here --> The following is an account of the origins of this work and of the more specialized developments of the 1980's that affected it. I began working on these problems in the late summer of 1986 immediately on learning of Ribet's result. For several years I had been working on the Iwasawa conjecture for totally real fields and some applications of it. In the process, I had been using and developing results on ℓ-adic representations associated to Hilbert modular forms. It was therefore natural for me to consider the problem of modularity from the point of view of ℓ-adic representations. I began with the assumption that the reduction of a given ordinary ℓ-adic representation was reducible and tried to prove under this hypothesis that the representation itself would have to be modular. I hoped rather naively that in this situation I could apply the techniques of Iwasawa theory. Even more optimistically I hoped that the case $\ell = 2$ would be tractable as this would suffice for the study of the curves used by Frey. From now on and in the main text, we write p for ℓ because of the connections with Iwasawa theory.

After several months studying the 2-adic representation, I made the first real breakthrough in realizing that I could use the 3-adic representation instead: the Langlands-Tunnell theorem meant that ρ_3, the mod 3 representation of any given elliptic curve over $\mathbf{Q}$, would necessarily be modular. This enabled me to try inductively to prove that the $\mathrm{GL}_2(\mathbf{Z}/3^n\mathbf{Z})$ representation would be modular for each n. At this time I considered only the ordinary case. This led quickly to the study of $H^i(\mathrm{Gal}(F_\infty/\mathbf{Q}), W_f)$ for $i = 1$ and 2, where F_∞ is the splitting field of the $\mathfrak{m}$-adic torsion on the Jacobian of a suitable modular curve, $\mathfrak{m}$ being the maximal ideal of a Hecke ring associated to ρ_3 and W_f the module associated to a modular form f described in Chapter 1. More specifically, I needed to compare this cohomology with the cohomology of $\mathrm{Gal}(\mathbf{Q}_\Sigma/\mathbf{Q})$ acting on the same module.

I tried to apply some ideas from Iwasawa theory to this problem. In my solution to the Iwasawa conjecture for totally real fields [Wi4], I had introduced

a new technique in order to deal with the trivial zeroes. It involved replacing the standard Iwasawa theory method of considering the fields in the cyclotomic $\mathbf{Z}_p$-extension by a similar analysis based on a choice of infinitely many distinct primes $q_i \equiv 1 \bmod p^{n_i}$ with $n_i \to \infty$ as $i \to \infty$. Some aspects of this method suggested that an alternative to the standard technique of Iwasawa theory, which seemed problematic in the study of W_f, might be to make a comparison between the cohomology groups as Σ varies but with the field $\mathbf{Q}$ fixed. The new principle said roughly that the unramified cohomology classes are trapped by the tamely ramified ones. After reading the paper [Gre1], I realized that the duality theorems in Galois cohomology of Poitou and Tate would be useful for this. The crucial extract from this latter theory is in Section 2 of Chapter 1.

In order to put these ideas into practice I developed in a naive form the techniques of the first two sections of Chapter 2. This drew in particular on a detailed study of all the congruences between f and other modular forms of differing levels, a theory that had been initiated by Hida and Ribet. The outcome was that I could estimate the first cohomology group well under two assumptions, first that a certain subgroup of the second cohomology group vanished and second that the form f was chosen at the minimal level for $\mathfrak{m}$. These assumptions were much too restrictive to be really effective but at least they pointed in the right direction. Some of these arguments are to be found in the second section of Chapter 1 and some form the first weak approximation to the argument in Chapter 3. At that time, however, I used auxiliary primes $q \equiv -1 \bmod p$ when varying Σ as the geometric techniques I worked with did not apply in general for primes $q \equiv 1 \bmod p$. (This was for much the same reason that the reduction of level argument in [Ri1] is much more difficult when $q \equiv 1 \bmod p$.) In all this work I used the more general assumption that ρ_p was modular rather than the assumption that $p = 3$.

In the late 1980's, I translated these ideas into ring-theoretic language. A few years previously Hida had constructed some explicit one-parameter families of Galois representations. In an attempt to understand this, Mazur had been developing the language of deformations of Galois representations. Moreover, Mazur realized that the universal deformation rings he found should be given by Hecke rings, at least in certain special cases. This critical conjecture refined the expectation that all ordinary liftings of modular representations should be modular. In making the translation to this ring-theoretic language I realized that the vanishing assumption on the subgroup of H^2 which I had needed should be replaced by the stronger condition that the Hecke rings were complete intersections. This fitted well with their being deformation rings where one could estimate the number of generators and relations and so made the original assumption more plausible.

To be of use, the deformation theory required some development. Apart from some special examples examined by Boston and Mazur there had been

little work on it. I checked that one could make the appropriate adjustments to the theory in order to describe deformation theories at the minimal level. In the fall of 1989, I set Ramakrishna, then a student of mine at Princeton, the task of proving the existence of a deformation theory associated to representations arising from finite flat group schemes over $\mathbf{Z}_p$. This was needed in order to remove the restriction to the ordinary case. These developments are described in the first section of Chapter 1 although the work of Ramakrishna was not completed until the fall of 1991. For a long time the ring-theoretic version of the problem, although more natural, did not look any simpler. The usual methods of Iwasawa theory when translated into the ring-theoretic language seemed to require unknown principles of base change. One needed to know the exact relations between the Hecke rings for different fields in the cyclotomic $\mathbf{Z}_p$-extension of $\mathbf{Q}$, and not just the relations up to torsion.

The turning point in this and indeed in the whole proof came in the spring of 1991. In searching for a clue from commutative algebra I had been particularly struck some years earlier by a paper of Kunz [Ku2]. I had already needed to verify that the Hecke rings were Gorenstein in order to compute the congruences developed in Chapter 2. This property had first been proved by Mazur in the case of prime level and his argument had already been extended by other authors as the need arose. Kunz's paper suggested the use of an invariant (the η-invariant of the appendix) which I saw could be used to test for isomorphisms between Gorenstein rings. A different invariant (the $\mathfrak{p}/\mathfrak{p}^2$-invariant of the appendix) I had already observed could be used to test for isomorphisms between complete intersections. It was only on reading Section 6 of [Ti2] that I learned that it followed from Tate's account of Grothendieck duality theory for complete intersections that these two invariants were equal for such rings. Not long afterwards I realized that, unlikely though it seemed at first, the equality of these invariants was actually a criterion for a Gorenstein ring to be a complete intersection. These arguments are given in the appendix.

The impact of this result on the main problem was enormous. Firstly, the relationship between the Hecke rings and the deformation rings could be tested just using these two invariants. In particular I could provide the inductive argument of Section 3 of Chapter 2 to show that if all liftings with restricted ramification are modular then all liftings are modular. This I had been trying to do for a long time but without success until the breakthrough in commutative algebra. Secondly, by means of a calculation of Hida summarized in [Hi2] the main problem could be transformed into a problem about class numbers of a type well-known in Iwasawa theory. In particular, I could check this in the ordinary CM case using the recent theorems of Rubin and Kolyvagin. This is the content of Chapter 4. Thirdly, it meant that for the first time it could be verified that infinitely many j-invariants were modular. Finally, it meant that I could focus on the minimal level where the estimates given by my earlier

Galois cohomology calculations looked more promising. Here I was also using the work of Ribet and others on Serre's conjecture (the same work of Ribet that had linked Fermat's Last Theorem to modular forms in the first place) to know that there was a minimal level.

The class number problem was of a type well-known in Iwasawa theory and in the ordinary case had already been conjectured by Coates and Schmidt. However, the traditional methods of Iwasawa theory did not seem quite sufficient in this case and, as explained earlier, when translated into the ring-theoretic language seemed to require unknown principles of base change. So instead I developed further the idea of using auxiliary primes to replace the change of field that is used in Iwasawa theory. The Galois cohomology estimates described in Chapter 3 were now much stronger, although at that time I was still using primes $q \equiv -1 \bmod p$ for the argument. The main difficulty was that although I knew how the η-invariant changed as one passed to an auxiliary level from the results of Chapter 2, I did not know how to estimate the change in the $\mathfrak{p}/\mathfrak{p}^2$-invariant precisely. However, the method did give the right bound for the generalised class group, or Selmer group as it is often called in this context, under the additional assumption that the minimal Hecke ring was a complete intersection.

I had earlier realized that ideally what I needed in this method of auxiliary primes was a replacement for the power series ring construction one obtains in the more natural approach based on Iwasawa theory. In this more usual setting, the projective limit of the Hecke rings for the varying fields in a cyclotomic tower would be expected to be a power series ring, at least if one assumed the vanishing of the μ-invariant. However, in the setting with auxiliary primes where one would change the level but not the field, the natural limiting process did not appear to be helpful, with the exception of the closely related and very important construction of Hida [Hi1]. This method of Hida often gave one step towards a power series ring in the ordinary case. There were also tenuous hints of a patching argument in Iwasawa theory ([Scho], [Wi4, §10]), but I searched without success for the key.

Then, in August, 1991, I learned of a new construction of Flach [Fl] and quickly became convinced that an extension of his method was more plausible. Flach's approach seemed to be the first step towards the construction of an Euler system, an approach which would give the precise upper bound for the size of the Selmer group if it could be completed. By the fall of 1992, I believed I had achieved this and began then to consider the remaining case where the mod 3 representation was assumed reducible. For several months I tried simply to repeat the methods using deformation rings and Hecke rings. Then unexpectedly in May 1993, on reading of a construction of twisted forms of modular curves in a paper of Mazur [Ma3], I made a crucial and surprising breakthrough: I found the argument using families of elliptic curves with a

common ρ_5 which is given in Chapter 5. Believing now that the proof was complete, I sketched the whole theory in three lectures in Cambridge, England on June 21–23. However, it became clear to me in the fall of 1993 that the construction of the Euler system used to extend Flach's method was incomplete and possibly flawed.

Chapter 3 follows the original approach I had taken to the problem of bounding the Selmer group but had abandoned on learning of Flach's paper. Darmon encouraged me in February, 1994, to explain the reduction to the complete intersection property, as it gave a quick way to exhibit infinite families of modular j-invariants. In presenting it in a lecture at Princeton, I made, almost unconsciously, a critical switch to the special primes used in Chapter 3 as auxiliary primes. I had only observed the existence and importance of these primes in the fall of 1992 while trying to extend Flach's work. Previously, I had only used primes $q \equiv -1 \bmod p$ as auxiliary primes. In hindsight this change was crucial because of a development due to de Shalit. As explained before, I had realized earlier that Hida's theory often provided one step towards a power series ring at least in the ordinary case. At the Cambridge conference de Shalit had explained to me that for primes $q \equiv 1 \bmod p$ he had obtained a version of Hida's results. But except for explaining the complete intersection argument in the lecture at Princeton, I still did not give any thought to my initial approach, which I had put aside since the summer of 1991, since I continued to believe that the Euler system approach was the correct one.

Meanwhile in January, 1994, R. Taylor had joined me in the attempt to repair the Euler system argument. Then in the spring of 1994, frustrated in the efforts to repair the Euler system argument, I began to work with Taylor on an attempt to devise a new argument using $p = 2$. The attempt to use $p = 2$ reached an impasse at the end of August. As Taylor was still not convinced that the Euler system argument was irreparable, I decided in September to take one last look at my attempt to generalise Flach, if only to formulate more precisely the obstruction. In doing this I came suddenly to a marvelous revelation: I saw in a flash on September 19th, 1994, that de Shalit's theory, if generalised, could be used together with duality to glue the Hecke rings at suitable auxiliary levels into a power series ring. I had unexpectedly found the missing key to my old abandoned approach. It was the old idea of picking q_i's with $q_i \equiv 1 \bmod p^{n_i}$ and $n_i \to \infty$ as $i \to \infty$ that I used to achieve the limiting process. The switch to the special primes of Chapter 3 had made all this possible.

After I communicated the argument to Taylor, we spent the next few days making sure of the details. The full argument, together with the deduction of the complete intersection property, is given in [TW].

In conclusion the key breakthrough in the proof had been the realization in the spring of 1991 that the two invariants introduced in the appendix could be used to relate the deformation rings and the Hecke rings. In effect the η-

invariant could be used to count Galois representations. The last step after the June, 1993, announcement, though elusive, was but the conclusion of a long process whose purpose was to replace, in the ring-theoretic setting, the methods based on Iwasawa theory by methods based on the use of auxiliary primes.

One improvement that I have not included but which might be used to simplify some of Chapter 2 is the observation of Lenstra that the criterion for Gorenstein rings to be complete intersections can be extended to more general rings which are finite and free as $\mathbf{Z}_p$-modules. Faltings has pointed out an improvement, also not included, which simplifies the argument in Chapter 3 and [TW]. This is however explained in the appendix to [TW].

It is a pleasure to thank those who read carefully a first draft of some of this paper after the Cambridge conference and particularly N. Katz who patiently answered many questions in the course of my work on Euler systems, and together with Illusie read critically the Euler system argument. Their questions led to my discovery of the problem with it. Katz also listened critically to my first attempts to correct it in the fall of 1993. I am grateful also to Taylor for his assistance in analyzing in depth the Euler system argument. I am indebted to F. Diamond for his generous assistance in the preparation of the final version of this paper. In addition to his many valuable suggestions, several others also made helpful comments and suggestions especially Conrad, de Shalit, Faltings, Ribet, Rubin, Skinner and Taylor. Finally, I am most grateful to H. Darmon for his encouragement to reconsider my old argument. Although I paid no heed to his advice at the time, it surely left its mark.

Table of Contents