An investigation of the role of Glu-842, Glu-844 and His-846 in the function of the cytoplasmic domain of the epidermal growth factor receptor

John F. TIMMS, Martin E. M. NOBLE and Mary GREGORIOU*

Laboratory of Molecular Biophysics, Department of Biochemistry, University of Oxford, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, U.K.

Activation of several protein kinases is mediated, at least in part, by phosphorylation of conserved Thr or Tyr residues located in a variable loop region, near the active site. In certain kinases, this activation loop also controls access of peptide substrates to the active site. In the corresponding region of the epidermal growth factor (EGF) receptor, a potential phosphorylation site, Tyr-845, does not appear to have a major regulatory role. In order to find out whether this variable loop can modulate the peptide phosphorylation and self-phosphorylation activities of the EGF receptor kinase, we investigated the role of residues around Tyr-845, using site-directed mutagenesis. Multiple sequence alignment showed that residues Glu-842, Glu-844 and His-846 are conserved or nearly conserved in eight members of the EGF receptor family. Mutants Glu-842→Ser, Glu-844→Gln and His-846→Ala were expressed in the baculovirus/insect cell system, purified to near-homogeneity and characterized with respect to their peptide phosphorylation and self-phosphorylation activities. All three mutants were active, and these changes did not affect ATP binding directly. However, all mutations increased the $K_m^{app.}$ for peptide substrates and MnATP in peptide phosphorylation reactions. The $V_{max.}$ for the phosphorylation of peptide RREELQDDYEDD was unaltered, but the $V_{max.}$ for self-phosphorylation (with variable [MnATP]) decreased 4-, 2- and 7-fold for mutants Glu-842→Ser, Glu-844→Gln and His-846→Ala respectively, compared with the wild-type. These results suggest that binding of this peptide restored an optimal conformation at the active site that might be impaired by the mutations. A study of the dependence of initial rates of self-phosphorylation on cytoplasmic domain concentration showed that the order of reaction increased with the progress of self-phosphorylation. Both pre-phosphorylation and high concentrations of ammonium sulphate restored maximal or near-maximal levels of self-phosphorylation in the mutants, possibly through compensating conformational changes. A plausible homology model, based on the cyclic AMP-dependent protein kinase catalytic subunit, accommodated the sequence Glu-841-Glu-Lys-Glu as an insertion in the peptide binding loop at the edge of the active site cleft. The model suggests that Glu-844 and His-846 may participate in H-bonding interactions, thus stabilizing the active site region, while Glu-842 does not appear to interact with regions of the catalytic core.

INTRODUCTION

Growth factors trigger intracellular signalling by allosteric activation of the cytoplasmic tyrosine kinase domains of their respective receptors. Activation leads to increased self-phosphorylation of the receptor and phosphorylation of intracellular proteins, thereby activating pathways of intracellular communication. Receptor tyrosine kinases exhibit specificity with respect to their interactions with signalling proteins, substrate phosphorylation and mechanisms of regulation of their catalytic domains. The mechanism of activation of the epidermal growth factor (EGF) receptor kinase involves oligomerization of the receptor, which is induced and stabilized by EGF binding to the extracellular domain [1–3]. Self-phosphorylation of the EGF receptor cytoplasmic domain occurs at five sites located in the C-terminal domain outside the catalytic core. Self-phosphorylation serves to connect and trigger downstream signalling, but does not appear to cause conspicuous activation of the kinase (i.e. increase in $V_{max.}$) towards protein substrates [4]. The C-terminal self-phosphorylation sites cause auto-inhibition, which is relieved after their phosphorylation, as indicated by a decrease in the K_m for peptide substrates [5].

Several kinases are activated by phosphorylation of the catalytic core at conserved Thr or Tyr sites in the activation loop [6]; this loop is otherwise variable in length and has poor homology among protein kinases. Phosphorylation of Thr-197 has been shown to be required for the activation of the cyclic AMP-dependent protein kinase (cAPK) [7,8], and phosphorylation of Thr-160 and Thr-161 is required for activation of human cdk2 and cdc2 kinases respectively [9,10]. The crystallographic models of cAPK suggest that the phosphate group on Thr-197 forms a rich network of hydrogen bonds which may stabilize an active conformation near the active site [11,12]. Self-phosphorylation of the insulin receptor kinase domain at Tyr-1162 and Tyr-1163 was clearly shown to stimulate kinase catalytic activity [13], and phosphorylation of Tyr-416 is essential for the activation of src kinase [14]. In the EGF receptor the corresponding site is Tyr-845, which appears not to perform this regulatory function [15]. Multiple sequence alignment of protein kinases showed conservation of charged residues around the conserved Tyr-845 in members of the EGF receptor family [16]. The apparent lack of regulation of the EGF receptor kinase by phosphorylation at this site suggests that stabilization of an active conformation might be provided, at least in part, by interactions of charged and polar side chains at one or several positions in this region.

In this paper we aimed to investigate the role of residues

Abbreviations used: EGF, epidermal growth factor; cAPK, cAMP-dependent protein kinase A; peptide SA-12, RREELQDDYEDD; m.o.i., multiplicity of infection; [V^5]angiotensin, [Val^5]angiotensin (Asp-Arg-Val-Tyr-Val-His-Pro-Phe); ERK-2, extracellular signal-related kinase 2; DCIC, 3,4-dichloroisocoumarin.

* To whom correspondence and reprint requests should be addressed.

around Tyr-845 in the regulation of the kinase activity of the EGF receptor cytoplasmic domain. We used site-directed mutagenesis to make two conservative and one alanine substitution at positions 842, 844 and 846. We also built a plausible homology model of the catalytic core, in order to test our conclusions.

EXPERIMENTAL

Materials

Insect *Spodoptera frugiperda* cell line Sf9, *Autographa californica* nuclear polyhedrosis virus AcUW1-lacZ [17] and the transfer vector pAcCL29.1 [18] were all obtained from the Institute of Virology and Environmental Medicine, Oxford, U.K. Radiolabelled nucleotides [α-^{32}P]dCTP and [γ-^{32}P]ATP were purchased from Amersham International. Sequenase version 2.0 was obtained from USB, and other DNA modifying and restriction enzymes were from Boehringer. Cell culture reagents were from Gibco BRL. Oligonucleotides and the src peptide analogue RRLIEDAEYAARG were synthesized at the Oxford Centre for Molecular Sciences, Oxford, U.K. [V^5]Angiotensin II ([Val^5]angiotensin; Asp-Arg-Val-Tyr-Val-His-Pro-Phe) was from Bachem, Switzerland, and the peptide substrate analogue corresponding to the erythrocyte band 3 phosphorylation site RREELQDDYEDD (here abbreviated to SA-12) [19] was generously provided by Tim Page (Dyson Perrins Laboratory, Oxford, U.K.).

Isolation of mutant cDNA clones

A 1.8 kb cDNA encoding the cytoplasmic domain of the EGF receptor (EGF receptor residues 640–1186) were subcloned into the baculovirus transfer vector pAcCL29.1 [18] at the *Bam*HI cloning site adjacent to the polyhedrin promoter. Clones were isolated in *Escherichia coli* strain JM109, and a sense clone, pAcCL29.TK, was selected after restriction enzyme analysis. The pAcCL29.1 vector contains the M13 intragenic region, enabling production of single-stranded DNA in the presence of the helper phage M13KO7 [20]. Using single-stranded pAcCL29.TK, site-directed mutagenesis was carried out as described [21] after uracil incorporation by passage of the phagemid through *E. coli* strain RZ1032 (*ung*$^-$ *dut*$^-$ *F'*). The oligonucleotides 5'-GGTATTC-TTTCGATTCCGCACC-3', 5'-GCATGGTATTGTTTCTC-TTCC-3' and 5'-CCTCCTTCTGCAGCGTATTCTTTCTC-3' were used to make the amino acid substitutions Glu-842→Ser, Glu-844→Gln and His-846→Ala respectively, and mutant clones were selected from *E. coli* JM109 (*ung*$^+$) after transformation with *in vitro* synthesized double-stranded DNA. Mutant clones were screened by dideoxy sequencing of their plasmid DNA (prepared according to [22]). Mutant plasmid cDNA was purified on CsCl gradients [23] prior to lipofectin-mediated co-transfection with engineered baculovirus strain AcUW1-lacZ into Sf9 insect cells [24,25].

Isolation of recombinant baculovirus expression clones

Recombinants were harvested from the medium 3 days after co-transfection and cloned by limiting dilution and DNA hybridization [26]. Recombinant viruses were detected using a purified 1.3 kb *Pst*IEGF receptor cDNA probe, labelled with [α-^{32}P]dCTP. Pure recombinant baculoviruses were isolated from dilutions of 10^{-8} after at least three cycles of cloning. Clones were amplified in three consecutive passages before use for protein production.

Expression and preliminary characterization of wild-type and mutant proteins

Protein for preliminary characterization was produced from 1.5×10^7 Sf9 cells that had been infected with recombinant virus at a multiplicity of infection (m.o.i.) of 10, and grown at 28 °C. At 48 h post-infection, cell cultures were spun very gently and cell pellets were resuspended in 2 ml of homogenization buffer (10 mM triethanolamine/HCl, pH 7.6, 5 mM EDTA, 5 mM dithiothreitol, 1 % glycerol, 0.05 % Triton X-100, 0.5 mM PMSF and 10 μg/ml leupeptin; final concentrations) and frozen at −70 °C. Cells were lysed by thawing and incubation at 0 °C for 15 min followed by homogenization. Homogenates were supplemented with 150 mM NaCl and clarified by ultracentrifugation (61 700 *g* for 10 min at 4 °C). Recombinant protein was detected by SDS/PAGE [27] after Coomassie Blue staining, silver staining, immunostaining or autoradiography after self-phosphorylation with [γ-^{32}P]ATP (see below).

Protein production and purification

High-titre (10^8 plaque-forming units/ml) recombinant baculovirus stock was harvested 4 days post-infection from the medium of 1×10^8 Sf9 cells, infected at an m.o.i. < 1 in four 50 ml stirred cultures. This stock was used to infect Sf9 cells (1.2 litres; 2×10^9 cells) in 300 ml cultures, at an m.o.i. > 10. Cells were harvested after 48 h at 4 °C, and frozen in homogenization buffer at −70 °C. Cells were thawed and homogenized in 20 ml of buffer as described above, in the presence of the proteinase inhibitors leupeptin (12.5 μg/ml), aprotinin (6.25 μg/ml), E-64 (2.5 μg/ml), pepstatin (1 μg/ml) and 3,4-dichloroisocoumarin (DCIC) (50 μg/ml). The homogenate was clarified twice by spinning at 110000 *g* for 30 min. Protein was precipitated from the supernatant with 0.8 M $(NH_4)_2SO_4$ and pelleted at 16000 *g* for 15 min. The pellet was washed in 1.14 M ammonium sulphate, and dissolved in 10 ml of buffer A [50 mM triethanolamine/HCl, 25 mM NaEDTA, 3 mM NaN_3, 1 mM dithiothreitol, leupeptin (5 μg/ml), aprotinin (2.5 μg/ml), E-64 (2.5 μg/ml), pepstatin (1 μg/ml) and DCIC (2.5 μg/ml)]. The protein was clarified by spinning at 16000 *g* for 15 min, and purified by FPLC using a Mono Q HR10/10 column (Pharmacia) equilibrated in Buffer A, without proteinase inhibitors. The column was developed with a NaCl gradient in the same buffer, and the cytoplasmic domain was recovered at 0.15–0.2 M NaCl as two closely eluting protein peaks, I and II. Fractions in these peaks were pooled separately, diluted with 2 vol. of buffer A with proteinase inhibitors, and concentrated in Centricon 30 devices. Prior to storage in 50 % glycerol at −70 °C, the self-phosphorylation and [V^5]angiotensin II peptide phosphorylation activities of the two protein peaks were measured. The purity of the preparations was determined from 1 μg protein-loaded polyacrylamide gels after Coomassie Blue staining, by scanning in an LKB Ultrascan XL Laser Densitometer. Percentage purity was calculated from the ratio of the area under the kinase peak to the total area of all protein peaks.

Protein analysis and quantification

Protein was detected by SDS/PAGE [27] after Coomassie Blue or silver staining, and by immunostaining using a rabbit anti-(EGF receptor) antiserum (rHER) at 1:2000 dilution, followed by anti-(rabbit IgG)–alkaline phosphatase conjugate, also at 1:2000 dilution. Total protein in extracts was estimated using the micro BCA protein assay reagent system (Pierce) and BSA as a

standard. For determination of catalytic-centre activity, purified protein was quantified using the Bradford assay and BSA as standard.

Assays of kinase activity

Self-phosphorylation activity was determined by following the initial rate of incorporation of [^{32}P]phosphate from [γ-^{32}P]ATP into the cytoplasmic domain at 0 °C. Clarified cell extracts (20 μl) or purified protein (0.1–0.3 μg) were assayed in a final volume of 40 μl in 30 mM Hepes/NaOH, pH 7.4, and 10 mM $MnCl_2$. Reactions were initiated by addition of 20 μM of [γ-^{32}P]ATP, and terminated with 14 μl of 4 × Laemmli sample buffer. Phosphorylated protein was resolved by SDS/PAGE, and radiolabelled phosphoprotein was visualized by autoradiography of the dried gels and corresponding immunoblots, where it was also identified by reactivity with a specific anti-(EGF receptor) antiserum (rHER). Specific enzyme activity was determined by Čerenkov scintillation counting of the cytoplasmic domain bands cut out of the dried gels, and expressing the amount of radioactivity in pmol of [^{32}P]phosphate/min per mg of total protein.

Peptide phosphorylation was assayed essentially as described [28], using [V^5]angiotensin II (5 mM) in 30 mM Hepes/NaOH buffer, pH 7.4, in the presence of 10 mM $MnCl_2$ at 0 °C. Reactions were initiated with 20 μM [γ-^{32}P]ATP and terminated with trichloroacetic acid (7% final concentration). BSA (0.7%, w/v) was added as a carrier protein and, following a 20 min incubation on ice, samples were clarified at 13000 ***g*** for 20 min at 4 °C. [^{32}P]Phosphopeptide was quantified by spotting 10 μl portions of the supernatant on to gridded Whatman DE 81 phosphocellulose paper. This was washed four times in phosphoric acid (0.5%, v/v), followed by consecutive washes in ethanol, acetone and ether. The paper was dried, and bound [^{32}P]peptide determined by Čerenkov scintillation counting. Specific enzyme activities were calculated from the initial rates of reaction, which were determined from time-course studies.

Kinetic analysis of peptide phosphorylation

Kinetic parameters for peptide phosphorylation were determined from initial rates, measured in triplicate, at 0 °C; values were the means of at least two separate determinations. Parameters were determined using both the direct linear plot and the Michaelis–Menten equation. For the latter, a kinetics program was used (LEONORA; written by A. J. Cornish-Bowden). For the [V^5]angiotensin II phosphorylation and src peptide phosphorylation, parameters were determined from initial rates measured over 15 min, at 20 μM [γ-^{32}P]ATP, without enzyme pre-incubation. The ranges of peptide concentrations were 0–10 mM and 0–20 mM for [V^5]angiotensin II and src peptide respectively. For the peptide substrate analogue SA-12, rate measurements were made in the range 0.02–1 mM peptide. In this case the enzyme was pre-incubated with 20 μM [γ-^{32}P]ATP for 10 min to allow enzyme pre-phosphorylation, and rates were measured over 10 min following the addition of peptide. The kinetic parameters for MnATP as the variable substrate were determined from initial rates measured over 10 min, following a 15 min pre-incubation with [γ-^{32}P]ATP (1.1 μM for wild-type and 4.5 μM for the mutants), after which additional ATP was added in the range 0–50 μM, together with SA-12 peptide at a final saturating concentration of 400 μM. Residual [γ-^{32}P]ATP concentration was measured directly from control reactions which did not receive additional MnATP and which were terminated with trichloroacetic acid prior to addition of peptide. The amounts of residual ATP were then used to calculate the total amount of ATP for determination of kinetic parameters.

Kinetic analysis of autophosphorylation

All assays were carried out at 0 °C. Kinetic parameters for self-phosphorylation were determined from triplicate initial rates measured over 30 s, at 0–20 μM [γ-^{32}P]ATP. Parameters were calculated using the direct linear plot and the program LEONORA, and the values given were averaged from two to four separate experiments.

Sequence alignment

A sequence alignment of 76 protein kinases was derived using the program AMPS [29]. The sequences were obtained from the SWISSPROT database (version 22) [30] by scanning for homologues of the catalytic core of the EGF receptor (residues Phe-688–Phe-944), as defined by sequence similarity to the protein kinase family [16]. The alignment of eight members of the EGF receptor family of protein tyrosine kinases with the catalytic Cα subunit of murine cAPK was adjusted, where necessary, to take account of the secondary structure of cAPK, indicated in the crystallographic model of Knighton et al. [11]. Alignment of the EGF receptor catalytic core with the corresponding sequence of cAPK (Phe-43–Ile-291; 249 residues) required insertion of a total of 12 residues and deletion of a total of four residues in the cAPK sequence. These adjustments were made so as to maximize not only sequence identity and secondary structure, but also conservation of the positions of essential and structurally important residues. Insertions and deletions were thus made in the variable sections connecting elements of secondary structure, all of which were taken to be conserved with the exception of α-helix B.

Homology modelling

A model of the EGF receptor was built from the ternary complex of cAPK with inhibitory peptide and MgATP [31] using the program 'O' [32]. Between regions containing insertions and deletions, main-chain and side-chain co-ordinates were adopted directly from the parent structure. For non-conserved residues, side-chain conformations were chosen from a database of preferred rotamers, using the most commonly observed torsion angle combination, where this did not produce unacceptable steric clashes. The multiple sequence alignment was modified as described above, in order to move insertions or deletions into loop regions between secondary structural elements. All of the insertions and deletions involved single residues except for two four-residue insertions, which extended two existing loops. The first insertion extended the loop connecting β-strands 2 and 3 without altering secondary structure or the conserved features of the ATP binding site. The second insertion extended the loop connecting β-sheet 9 with conserved residues surrounding the active site. The structure of this region was chosen to maintain the secondary structure and the positions of these conserved residues. Main chain co-ordinates for these loops were selected using program 'O' from a collection of well refined structures, with side-chain rotamers selected from the rotamer database. A ternary complex model was produced by modelling into the catalytic core a five-residue peptide, Ala-Glu-Tyr-Leu-Arg. This sequence corresponds to that of the major autophosphorylation site, Tyr-1173, also used by Knighton et al. [33] in their model of the EGF receptor catalytic core. Co-ordinates for this substrate peptide were loosely based on the conformation of the inhibitory peptide seen in the ternary cAPK complex [31], although significant changes in main-chain conformation were required to allow for the relatively larger tyrosine side chain. The peptide was modelled so as to allow in-line attack on the γ-phosphate of

the ATP molecule, while keeping the substrate tyrosine in the most preferred side-chain conformation.

The complex was subjected to energy minimization by simulated annealing using the program X-PLOR [34]. A distance-dependent dielectric constant was used, with all full atomic charges (e.g. acidic or basic side-chain residues) set to zero. A cycle of simulated annealing consisted of 200 steps of Powell energy minimization, followed by cooling the structure from 500 K to 100 K in steps of 25 K, performing 100 steps of 1 fs dynamics at each temperature. Finally, the annealed structure was subjected to a further 100 steps of energy minimization. Harmonic restraints to the parent structure were applied to conserved side chains and to main-chain atoms away from regions of insertion/deletion. Additional harmonic restraints were applied to the MgATP atoms, and the position of the OH moiety of the substrate tyrosine residue was constrained to maintain a position suitable for in-line attack.

The structure was analysed using the PROCHECK [35] package to confirm stereochemical feasibility, and by the Eisenberg package to identify whether a 'native-like' distribution of exposed and buried side chains had been produced [36].

RESULTS

Sequence alignment of the catalytic cores of EGF receptor homologues with cAPK

The protein kinase family is characterized by eight invariant residues or groups of residues of functional or structural significance. These are in conserved subdomains, separated by sections of variable sequence and length [16]. Classes and subclasses of protein kinases are further characterized by conserved residues determining substrate specificity or regulation of activity. In the catalytic domains of protein kinases, regulatory phosphorylation sites may be present in the activation loop. This is a variable region between β-sheet 9 and α-helix F, and connects the DFG and P--W---E motifs that line the active site. This region is shorter by five residues in cAPK compared with the EGF receptor tyrosine kinase family.

In order to identify loop residues which may have a role in stabilizing an active conformation around the active site, we carried out a multiple sequence alignment of eight EGF receptor homologues and cAPK (Figure 1). The alignment was derived using the programme AMPS [29], as described in the Experimental section. Four single-residue deletions in cAPK were made, namely Asp-75, Leu-82, Asp-112 and Phe-281. Four single-residue insertions in cAPK accommodated residues Lys-782, Ile-854, Phe-886 and Glu-907 of the EGF receptor. Two four-residue insertions in cAPK accommodated the EGF receptor sequences Pro-709-Glu-Gly-Glu, following β-sheet 2, and Glu-841-Glu-Lys-Glu, following β-sheet 9 (Figure 1). The position of the latter insertion was necessitated by aligning residues Pro-853 and Trp-856 in the EGF receptor with Pro-202 and Tyr-204 in cAPK respectively. Residues Glu-842, Glu-844 and His-846 of the EGF receptor kinase appeared to be significantly

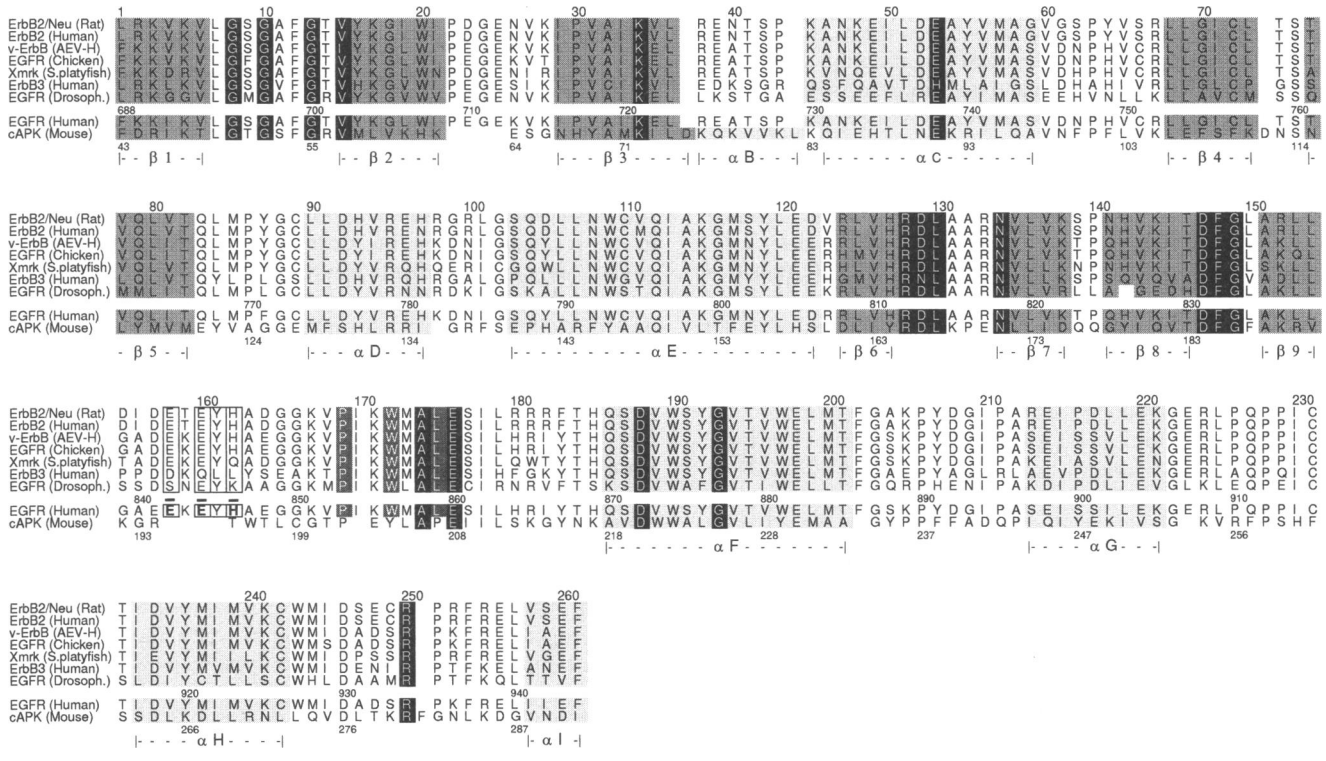

Figure 1 Multiple sequence alignment of eight homologous EGF receptor kinase cores with the kinase core of cAPK

The alignment was presented using the program ALSCRIPT [51]. The secondary structure of the cAPK crystallographic model [11] is indicated under the sequence and denoted by shading; light shading refers to α-helices and darker shading to β-strands. Invariable or nearly invariable residues conserved across the protein kinase family are shown in white against black background. Residues Pro-853, Trp-856 and Leu-859 conserved in the tyrosine kinase subfamily are shown in white letters against a dark grey background. Residues around Tyr-845, which are conserved in the EGF receptor family, are shown in vertical boxes. The EGF receptor numbering [52] is indicated on top of the EGF receptor sequence, and cAPK numbering below the cAPK sequence. The three sites of mutation, His-846, Glu-844 and Glu-842, are indicated.

Table 1 Purification of the wild-type and mutant cytoplasmic domains

The 0.8 M $(NH_4)_2SO_4$ fractions of soluble extracts of infected Sf9 cells were purified by FPLC on a Mono Q column. Data for the major Mono Q peak are shown. Total activity refers to phosphorylation of angiotensin II, and was determined as described in the Experimental section, in the presence of 0.5 M $(NH_4)_2SO_4$. Percentage yield was calculated by reference to the $(NH_4)_2SO_4$ fraction, which was taken to represent a value of 100%. Purity was estimated by scanning densitometry of SDS gels, as described in the Experimental section.

Mutation	Total activity $(NH_4)_2SO_4$ fraction (pmol/min)	Total activity Mono Q fraction (pmol/min)	Specific activity (pmol/min per mg)	Total protein (mg)	Yield (%)	Purity (%)
Wild-type	3296.85	401.35	836.15	0.48	12.17	91
Glu-842→Ser	1290.50	799.14	1452.98	0.55	61.92	88
Glu-844→Gln	2657.50	736.40	1067.24	0.69	27.71	90
His-846→Ala	951.90	331.13	1103.77	0.30	34.70	83

conserved within the subfamily, and we therefore investigated their role using site-directed mutagenesis.

Isolation of mutant cDNA clones and expression and preliminary comparison of mutant and wild-type proteins

Mutants having the single amino acid substitutions Glu-842→Ser, Glu-844→Gln and His-846→Ala were made. The first two substitutions generally occur with high natural frequency [37], as exemplified in the cases of the *Drosophila* EGF receptor (Glu-842→Ser) and the human Erb B3 (Glu-844→Gln). The His-846→Ala substitution occurs naturally, though at relatively lower frequency at accessible positions, and at very low frequency at buried sites. This mutation was selected over slightly more frequently occurring natural substitutions (e.g. Lys, Asn) in order to test the possible involvement of His in hydrogen-bonding interactions.

Mutant cDNA clones were constructed using the Kunkel method for oligonucleotide-directed mutagenesis, and the corresponding cytoplasmic domains were designated Glu-842→Ser, Glu-844→Gln and His-846→Ala. These were expressed as soluble proteins in insect Sf9 cells, using the baculovirus expression system. Recombinant protein in clarified soluble cell extracts was detected by SDS/PAGE after Coomassie Blue and silver staining and by immunoblot analysis. The cytoplasmic domain was visible as a doublet, running with an apparent molecular mass of 62–65 kDa. Immunoblotting suggested that similar amounts of mutant and wild-type cytoplasmic domain were expressed in these preliminary tests. The rates of self-phosphorylation in clarified soluble extracts were reduced by approx. 7- and 9-fold for the mutants Glu-842→Ser and His-846→Ala respectively, and 2-fold for Glu-844→Gln, compared with the wild-type kinase.

Purification of wild-type and mutant cytoplasmic domains

Cytoplasmic domains were purified from 1.2 litres (2×10^9 cells) of infected cells using a two-step procedure, as described in the Experimental section. The proteins were recovered in two closely eluting peaks of self-phosphorylation activity from the FPLC Mono Q column, at approx. 150 mM NaCl. These peaks were not related by tyrosine self-phosphorylation [38]. This purification procedure yielded approx. 0.30–0.69 mg of protein, which was 83–91% pure (Table 1; Figure 2). The proteins were stored in the presence of 50% glycerol at −70 °C for at least 1 year without detectable loss of activity.

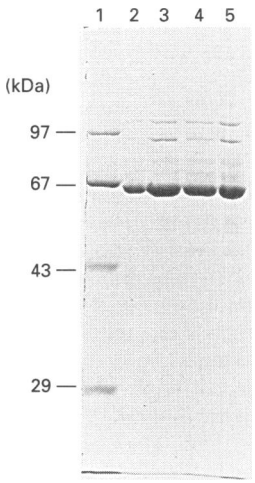

Figure 2 SDS/PAGE analysis of the purified wild-type and mutant proteins

Wild-type and mutant proteins were purified by $(NH_4)_2SO_4$ fractionation and Mono Q chromatography as described. Proteins (1–2 μg) were analysed by SDS/PAGE and stained with Coomassie Blue. Protein markers included (from top to bottom): phosphorylase, BSA, ovalbumin and carboxypeptidase. Lanes: 1, markers; 2, wild-type; 3, Glu-842→Ser; 4, Glu-844→Gln; 5, His-846→Ala.

Mutant kinase domains were analysed by gel filtration using an FPLC Superose 12 column (Pharmacia). All mutants and the wild-type cytoplasmic domain eluted as single peaks of active protein with Stoke's radii similar to that of BSA. This suggested that the mutants, like the wild-type protein, were monomeric.

Kinetic analysis of the purified cytoplasmic domain mutants

Initially, we compared the purified mutants with the wild-type cytoplasmic domain with respect to their $V_{max.}^{app.}$ and $K_m^{app.}$ values using the [V^5]angiotensin II phosphorylation assay, in the presence of 0.5 M $(NH_4)_2SO_4$. Ammonium sulphate at high concentrations stimulates kinase activity towards this peptide [39]. The His-846→Ala mutant was not saturated with peptide in the range 0–10 mM angiotensin II, suggesting a large increase in the $K_m^{app.}$ (Table 2). The $K_{m(src)}^{app.}$ of His-846→Ala was increased at least 13-fold compared with that of the wild-type, whilst the $V_{max.}^{app.}$ was relatively unaffected (Table 2). This $K_{m(src)}^{app.}$ (21.96 ± 2.73 mM), for His-846→Ala, represents a minimum

Table 2 Apparent kinetic parameters determined for the phosphorylation of peptide substrates angiotensin II and src peptide by the wild-type and His-846→Ala mutant cytoplasmic domain

Reaction rates were measured over 15 min, in the presence of 0.375 M $(NH_4)_2SO_4$ at 0 °C. Values for the kinetic parameters were averages of 2–4 experiments, and were determined using the Michaelis–Menten equation. Values are means ± S.D. No saturation was observed for His-846→Ala in the angiotensin II assay.

	$K_m^{app.}$ (mM)		$V_{max.}$ (pmol/min per mg)	
Substrate	Wild-type	His-846→Ala	Wild-type	His-846→Ala
Angiotensin II	5.32 ± 1.57	–	15450 ± 3274	–
src peptide	1.63 ± 0.93	21.96 ± 2.73	1162 ± 486	791 ± 262

Table 4 Apparent kinetic parameters for the autophosphorylation of wild-type and mutant EGF receptor cytoplasmic domains

Initial rates of autophosphorylation were measured over 30 s, using 0.4 μg of protein, in the MnATP concentration range 0–20 μM, at 0 °C. Values for the kinetic constants were determined as for Table 3 and were averaged from 2–4 separate experiments. Values are means ± S.D.

Cytoplasmic domain mutant	$K_{m(MnATP)}^{app.}$ (μM)	$V_{max.(MnATP)}^{app.}$ (pmol/min per mg)
Wild-type	1.04 ± 0.13	13365 ± 374
Glu-842→Ser	0.82 ± 0.13	3350 ± 233
Glu-844→Gln	0.79 ± 0.16	5079 ± 274
His-846→Ala	0.79 ± 0.10	1993 ± 181

value, because the solubility of the src peptide precluded use of higher concentrations. Therefore an alternative substrate, RREELQDDYEDD (SA-12), for which the wild-type kinase had a higher affinity, was used to compare the kinetic properties of the mutant and wild-type cytoplasmic domains.

The time course of [^{32}P]phosphate incorporation into peptide SA-12 (20–1000 μM) was non-linear, indicating that the rate increased with time. This suggested progressive relief from autoinhibition as a result of self-phosphorylation of the cytoplasmic domain, as observed previously for the phosphorylation of other peptide substrates [5]. Linearity was achieved after preincubation with 20 μM ATP for 10 min. Ammonium sulphate was found to cause a 50% inhibition of the SA-12 phosphorylation reaction, and was therefore omitted. In all cases, saturating Michaelis–Menten kinetics were observed from plots of rate versus substrate concentration. The His-846→Ala mutation caused a 6.6-fold increase in the $K_m^{app.}$ for this peptide and a 1.8-fold decrease in the $V_{max.}^{app.}$. The Glu-844→Gln mutation caused a less dramatic change in the affinity for SA-12 (2.8-fold increase in $K_m^{app.}$), and the $V_{max.}^{app.}$ was unaltered. The Glu-842→Ser mutation caused a 2.3-fold increase in the $K_m^{app.}$, and a 1.8-fold decrease in $V_{max.}$, as was observed for the His-846→Ala mutation (Table 3).

Comparison of the effects of the mutations on the affinity for MnATP showed that all mutations increased the $K_{m(ATP)}^{app.}$ by 2.5-, 4.9- and 4.2-fold in the mutants Glu-842→Ser, Glu-844→Gln and His-846→Ala respectively. The values of $V_{max.(ATP)}^{app.}$ were relatively unchanged (Table 3). To summarize, these mutations decreased affinity for peptide and MnATP substrates by 2–7-fold, the largest effect resulting from the His-846→Ala substitution.

Self-phosphorylation in mutant and wild-type cytoplasmic domains

Previous characterization of preparations of the cytoplasmic domain produced by the baculovirus expression system showed that the protein was not detectably phosphorylated, and was monomeric [38]; therefore it was possible to compare self-phosphorylation activities of the mutant and wild-type cytoplasmic domains. In the present study, the extents and rates of self-phosphorylation of wild-type and mutant cytoplasmic domains, the rate-dependences on MnATP and protein concentrations and their stimulation by $(NH_4)_2SO_4$, were compared. The kinetic parameters for self-phosphorylation as MnATP concentration was varied are shown in Table 4. All three mutants and the wild-type cytoplasmic domain had a K_m of approx. 0.8–1 μM, indicating that these mutations had no effect on ATP binding. However, comparison of the $V_{max.}$ values for self-phosphorylation showed that Glu-842→Ser, Glu-844→Gln and His-846→Ala had lower activities (4.1-fold, 2.7-fold and 7.3-fold respectively) compared with the wild-type protein. Comparison of time courses of self-phosphorylation at saturating ATP concentrations (19 μM) and at the same protein concentration showed that incorporation of 1 mol of phosphate/mol of cytoplasmic domain required approx. 7 min for the wild-type, 15 min for Glu-844→Gln, and 30 min for Glu-842→Ser; in the case of the His-846→Ala mutant, less than 1 phosphate was incorporated in 30 min. Phosphate incorporation began to plateau between 2 and 3 mol of phosphate/mol of protein in the wild-type protein, followed by a slower incorporation of an additional phosphate; slower rates of additional phosphate incorporation were observed in the mutants.

Table 3 Apparent kinetic parameters of wild-type and mutant EGF receptor cytoplasmic domains for the phosphorylation of peptide substrate analogue SA-12

To determine parameters with peptide as the variable substrate, [γ-^{32}P]MnATP was kept at a saturating concentration (20 μM). To determine parameters for MnATP, a saturating concentration (400 μM) of peptide was used. Wild-type and mutant cytoplasmic domains Glu-842→Ser, Glu-844→Gln and His-846→Ala were used at final concentrations of 0.16, 0.39, 0.58 and 0.60 mM respectively, as described in the Experimental section. Kinetic parameters were determined using the Michaelis–Menten equation. Values are means ± S.D.

Cytoplasmic domain mutant	$K_{m(SA-12)}^{app.}$ (μM)	$V_{max.(SA-12)}^{app.}$ (pmol/min per mg)	$K_{m(MnATP)}^{app.}$ (μM)	$V_{max.(MnATP)}^{app.}$ (pmol/min per mg)
Wild-type	22.05 ± 1.96	1150 ± 487	2.63 ± 0.13	2820 ± 166
Glu-842→Ser	50.50 ± 7.81	678 ± 52	6.75 ± 1.18	2500 ± 195
Glu-844→Gln	62.67 ± 7.64	1373 ± 170	12.94 ± 2.80	4415 ± 793
His-846→Ala	146.55 ± 45.55	718 ± 82	10.96 ± 1.10	2073 ± 82

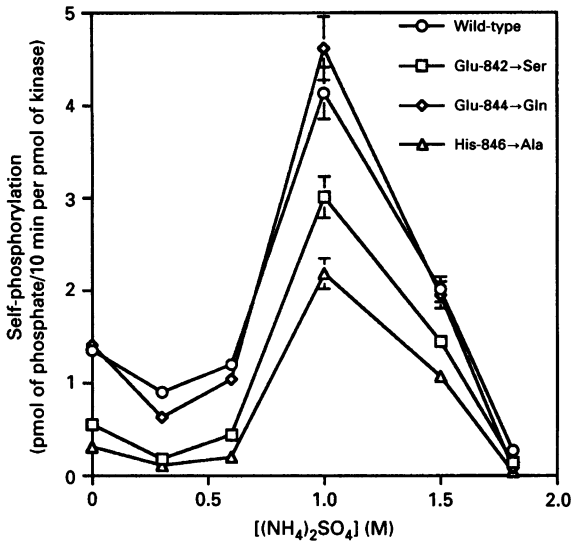

Figure 3 Effect of ammonium sulphate concentration on the autophosphorylation of wild-type and mutant EGF receptor cytoplasmic domains

Autophosphorylation assays were carried out in 30 mM Hepes/NaOH, pH 7.5, 10 mM $MnCl_2$, 10 μM MnATP and 5% glycerol with various concentrations of ammonium sulphate. Reactions were carried out on ice using 3.3 pmol of enzyme per assay, in a final reaction volume of 20 μl. Phosphate incorporation was measured over 10 min. The data points shown are the means of determinations from reactions performed in triplicate. Error bars represent % average error. Similar results were obtained in two other experiments.

Mutants were also compared with wild-type cytoplasmic domain with respect to their specificity of self-phosphorylation after either 2 min or 30 min of incubation with MnATP at 0 °C. Extensive digestion with trypsin and analysis by reverse-phase HPLC showed that all three mutants were indistinguishable from the wild-type in terms of the pattern of phosphopeptides produced. However, the His-846→Ala mutation yielded a strikingly low level of phosphopeptides compared with the wild-type, even after 30 min of reaction at 0 °C; the level of Glu-842→Ser phosphopeptides was also significantly reduced (results not shown).

Self-phosphorylation of all three mutants and wild-type cytoplasmic domain was stimulated to a similar extent by $(NH_4)_2SO_4$ at 1 M. However, the maximum level of activity was lower for Glu-842→Ser and His-846→Ala (Figure 3). Resolution of the phosphorylated proteins by SDS/PAGE, followed by autoradiography, revealed two retarded major bands of phosphorylated protein in all cases (results not shown). In the His-846→Ala mutant, the bands were less intensely labelled. Thus the His-846→Ala mutant could not be maximally phosphorylated, as also revealed by the tryptic phosphopeptide maps. These results showed conclusively that the His-846→Ala mutation caused a large decrease in the extent and rate of self-phosphorylation, though it did not eliminate it completely.

Kinetic analysis of self-phosphorylation is complicated by the existence of at least five phospho-acceptor sites, and a lack of detailed knowledge of the hierarchy of their phosphorylation. We examined the dependence of the stoichiometry of self-phosphorylation on the concentration of cytoplasmic domain, both under initial rate conditions and after pre-phosphorylation. Initial rates were achieved by using low concentrations of cytoplasmic domain and short and linear time courses (40 s–2 min, according to the level of intrinsic kinase activity),

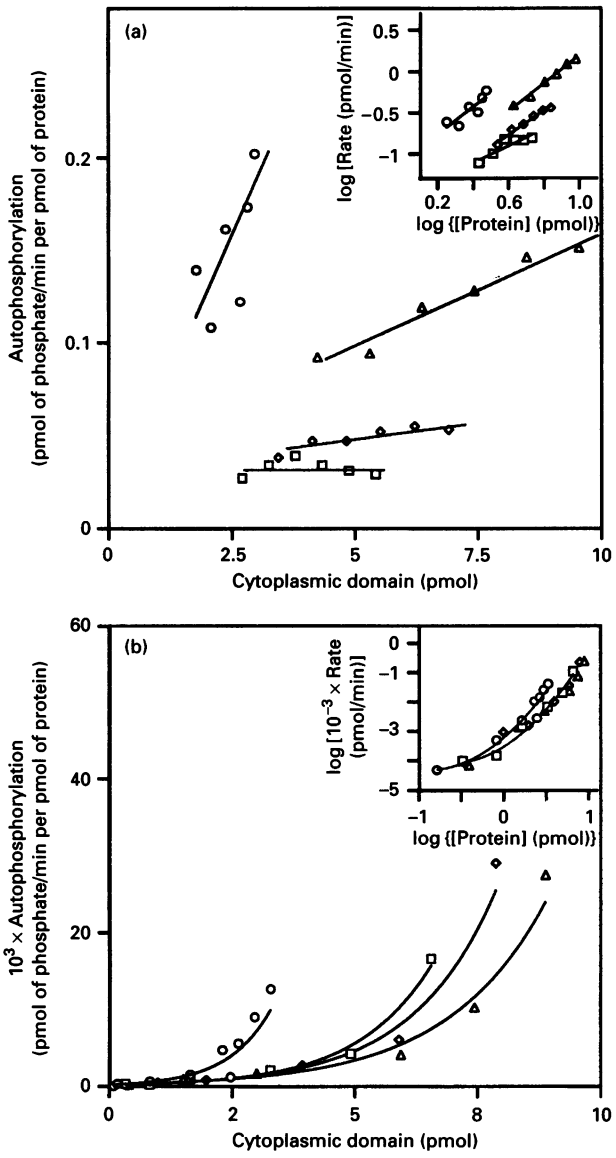

Figure 4 Effect of enzyme concentration on the autophosphorylation of wild-type and mutant EGF receptor cytoplasmic domains

Autophosphorylation assays were carried out as described, at 10 μM MnATP, in the presence of 5% glycerol. Initial rates (**a**) and 30 min reaction rates (**b**) were measured. The data points shown are the average values of three measurements. ○, Wild-type; ◇, Glu-842→Ser; △, Glu-844→Gln; □, His-846→Ala.

and by measuring activities at 0 °C. The dependence of initial rates of self-phosphorylation on the concentration of the cytoplasmic domain increased with increasing intrinsic kinase activity (Figure 4a). In the case of the His-846→Ala mutant, the stoichiometry of self-phosphorylation was independent of protein concentration, and the corresponding logarithmic plot had a slope of 1, indicating a first-order process (Figure 4; inset). This suggests an intramolecular self-phosphorylation process, or cross-phosphorylation within a dimer (E·S complex), provided that its formation is not rate-limiting. Logarithmic plots of the data for the Glu-842→Ser and Glu-844→Gln mutants and for the wild-type had slopes of 1.5, 1.7 and 1.7 respectively. A previous study of the wild-type cytoplasmic domain in which 10 s initial rates were measured at protein concentrations over a 50-

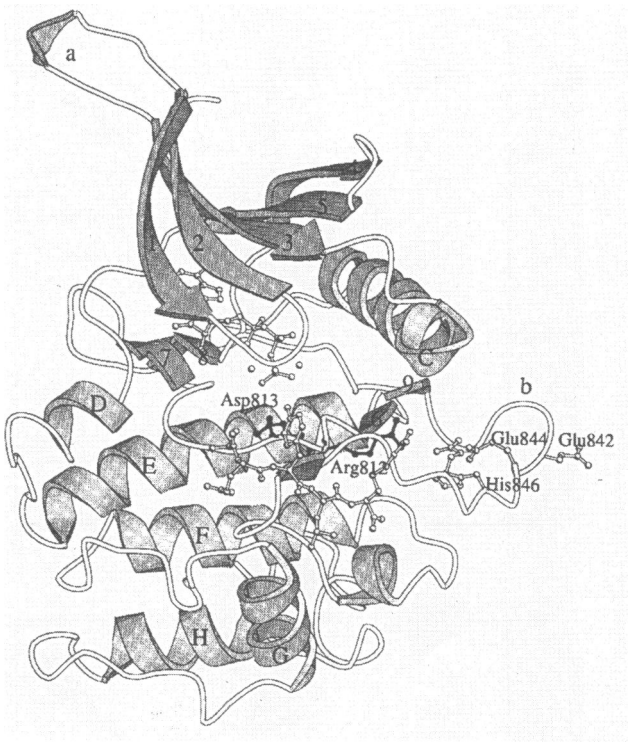

Figure 5 Model of the kinase catalytic core of the EGF receptor

The model was derived by homology modelling using the crystallographic structure of cAPK [11] as a template. The diagram was drawn using the program XOBJECTS (M. Noble, unpublished work). Arrows and spirals represent β-sheets and α-helices respectively, and their labels refer to the secondary structure of cAPK. Two four-residue insertions are indicated by the letters **a** (Pro-Glu-Gly-Glu) and **b** (Glu-Glu-Lys-Glu); remaining single-residue insertions and deletions are indicated in Figure 1 and in the text. The residue Arg-812 and the presumed catalytic base Asp-813 are indicated. The substrate sequence Ala-Glu-Tyr-Leu-Arg, and ATP, are shown; for clarity, the Arg at P + 2 is omitted from the drawing. The side chains of the mutated residues are indicated (labelled with large letters) along with Glu-848 (unlabelled).

fold range gave a slope of 1.5. Under the present experimental conditions, mutants Glu-842→Ser, Glu-844→Gln and the wild-type underwent 10–15% phosphorylation (mol of phosphate/mol of protein) during initial-rate measurements at the highest protein concentration investigated, while phosphorylation of the His-846→Ala mutant was much less. Therefore the order of reaction appeared to depend on the phosphorylation state of the kinase. The proportion of phosphorylated cytoplasmic domain increased with time and with protein concentration for all mutants and the wild-type protein. When 30 min reactions were measured (instead of initial rates) the apparent order of reaction increased for all mutants and the wild-type (Figure 4b). Plots of rates of self-phosphorylation against protein concentration were exponential, and the corresponding logarithmic plots fitted a second-order polynomial, indicating that the dependence of self-phosphorylation on kinase concentration increased with kinase concentration. Maximum slopes of 3.7 were measured for the wild-type and all mutants (Figure 4b, inset), suggesting complex higher-order reactions.

These results are consistent with alternative mechanisms coming into operation as self-phosphorylation progresses: an initial first-order mechanism involving phosphorylation of the first site(s), and a subsequent higher-order mechanism, involving phosphorylation at several multiple sites, in the C-terminal domain. It is reasonable to hypothesize that initial phosphorylation exposes additional phospho-acceptor sites in the C-terminal domain, and increases accessibility of the active site in the catalytic core through relief of auto-inhibition, thus promoting bimolecular and larger complexes. Supporting experimental evidence is provided by the report that phosphorylated cytoplasmic domain adopts an extended conformation [40], which we also observe using our recombinant preparations [38]. Both the cross-phosphorylation mechanism and the auto-inhibitory properties of the C-terminal domain are well documented [5,41].

Homology modelling

The sequence Phe-688–Phe-944 of the EGF receptor was modelled on the crystallographic structure of cAPK [31], using the sequence alignment shown in Figure 1. A diagram of the model, with spirals and arrows representing α-helices and β-sheets respectively, is shown in Figure 5. After energy minimization, the modelled structure had good stereochemistry with respect to standard bond lengths (r.m.s. = 0.01 Å) and bond angles (r.m.s. = 2.7°). There are three residues just outside the allowed regions of the Ramachandran plot, and the agreement with the preferred conformational parameters analysed by PROCHECK [35] is comparable with a structure of 2.5 Å resolution. Comparison of the core main chain (N, Cα, C, O) atoms between the model and the parent structure gave an r.m.s. value of 0.33 Å, which is acceptable given the low degree of identity (22%) between the two sequences. Eisenberg analysis [36] of the structure gave a summed value (85.8) falling within the acceptable range for a 'correct' structure of this size.

The structure of the modelled EGF receptor catalytic core differs in three regions from that of the cAPK template. Two of these are extended surface loops, further extended in the EGF receptor by four residues each (a and b; Figure 5). The third is conversion of α-helix B in cAPK into a four-residue strand in the EGF receptor, necessitated by the presence of Pro-729 in EGF receptor and deletion of cAPK residues Asp-75 and Leu-82 (Figures 1 and 5).

The structure of the model close to the active site was found to depend critically on the alignment of the activation loop (residues 840–856) with that in cAPK. Owing to the poor conservation of this region, this loop was built principally by hand (as described in the Experimental section) to give the alignment shown in Figure 1. This alignment maintains equivalence of a proline residue (Pro-853 in EGF receptor and Pro-202 in cAPK), preceded by a β-branched residue (Val-852 in EGF receptor and Thr-201 in cAPK) close to the active site. The insertion Glu-841-Glu-Lys-Glu increased the length of the loop, forming an extended bend at the surface (b in Figure 5). This insertion and that of Ile-854 maintained equivalence of the position of Trp-856 with that of Trp-204 (in cAPK). An aromatic residue in this position can fill a hydrophobic pocket (formed by the side chains of Val-878, Lys-889, Arg-779, Leu-775, Pro-853 and Ala-815) adjacent to the active site [11,33].

The model appears to be largely in agreement with the two published models for an EGF receptor kinase catalytic core–substrate complex [33,42], although differences in the alignments of the surface loop caused a significant change in the interactions available to the peptide substrate. As has been observed for the models of Knighton et al. [33] and Singh [42], the guanidino group of Arg-817 is in close proximity with the γ-phosphate of ATP, and so presumably compensates for the loss of the amino group of Lys-168 in cAPK (equivalent to Ala-815 in the EGF receptor). In addition to these observations, the

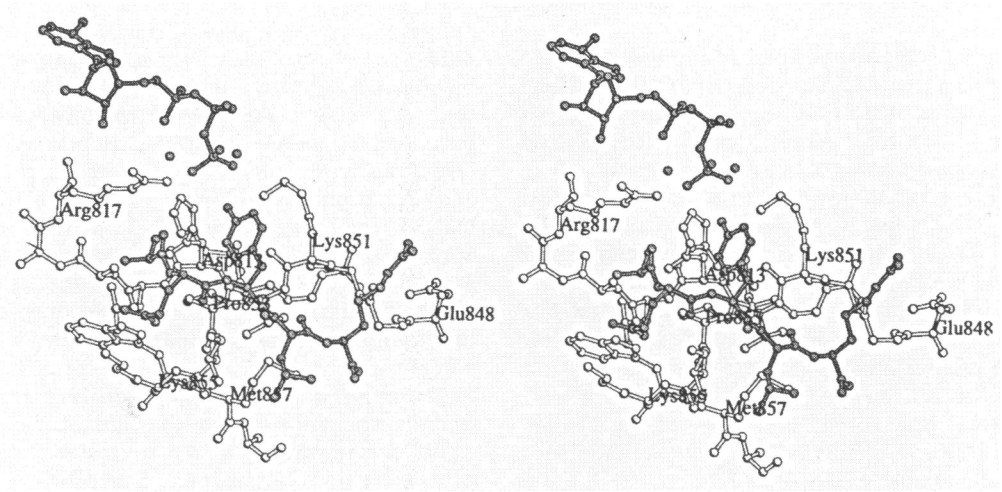

Figure 6 Stereo view of the modelled active site

The kinase atoms are shown in white while these of the peptide and ATP substrates and Mg^{2+} are shown in grey. The presumed catalytic base, Asp-813, and the adjacent Arg-812, are in the background, with side chains extending forward and to the right respectively. The guanidino group of Arg-812 is poised to interact with the carboxyl group of Glu-848. The carboxyl group of Asp-813 is in the vicinity of the γ-phosphate of ATP and the hydroxyl group of substrate tyrosine. The tyrosine aromatic ring is flanked on the right by the side chain of Lys-851 and underneath by Pro-853. The hydrophobic side chain in the substrate P+1 position is flanked by Met-857 at the back and Ile-854 on the left, which lies next to the aliphatic side chain of Lys-855. The amino group of Lys-855 may interact with and stabilize the carboxyl group at the P−1 position of the substrate. The side chain of Trp-856 extends to the left of Pro-853, extending the hydrophobic environment at the base of the active site. At the top, the γ-phosphate is flanked by two Mg ions, and by the guanidino and amino groups of Arg-817 and Lys-851 respectively.

model and alignment we describe here places Lys-851 close to the active site, so that its amino group is able to interact with the γ-phosphate of ATP, while its aliphatic side chain can form a hydrophobic interaction with tyrosine in the peptide substrate (Figure 6). Other residues close to the substrate tyrosine include Asp-813 and Pro-853. Asp-813 is the proposed catalytic base, and in this model it is well placed to polarize the proton of the substrate tyrosine. Pro-853 is conserved between serine/threonine and tyrosine kinases, but in the context of the tyrosine kinase active site, it might be expected to provide a hydrophobic platform for the aromatic ring of the substrate tyrosine. As well as these side-chain interactions, the modelled complex features main-chain to main-chain hydrogen bonds between residues Val-852–Ile-854 and the backbone of residues P_{-1} to P_{+1} of the peptide substrate.

The model presented here predicts that Glu-848 is the preferred candidate for forming an ion-pair with Arg-812 adjacent to the presumed catalytic base Asp-813 (Figure 6). This interaction is likely to stabilize the conformation of the active site. In cAPK, Arg-165 (adjacent to the catalytic base) is stabilized through a network of hydrogen-bonding interactions centring on the phosphorylated Thr-197 side chain. Glu-848 is also likely to hydrogen-bond with Lys-836, in a manner analogous with the interactions between the phosphorylated Thr-197 and Lys-189 in cAPK (Figures 5 and 6). His-846 may interact with both Glu-848 and Glu-844, the carboxylate of which (Glu-848) may also interact with Lys-836. Thus it would appear that both Glu-848 and Glu-844 might perform the stabilizing function of phosphothreonine in cAPK, while His-846 might affect the relative positions of these carboxylates for optimal interactions with Lys-836. The imidazole of His-846 may also interact with the main-chain carbonyl oxygen of Tyr-867, the hydroxyl of which maintains an interaction with the guanidino group of Arg-812, as in cAPK.

In the sequence Glu-842-Lys-Glu-Tyr, the model shows the side chains of Glu-842 (Figure 5) and Tyr-845 pointing towards the solvent, apparently not interacting with any parts of the catalytic core. However, it is not known whether these side chains make important interactions with the C-terminal domain or with protein substrates.

DISCUSSION

Crystal structures of several protein kinase catalytic cores suggest that protein kinases share a similar framework for substrate binding and catalysis. Two distinguishing features have emerged from known kinase structures. One concerns the relative position of the small (ATP binding) and large (peptide binding) lobes, which can rotate relative to each other to generate open and closed active site clefts. Another characteristic, observed in open conformations, is the accessibility of the active site. This seems freely accessible in the open structure of porcine cAPK, but is blocked by the activation loop in the open structures of human cdk2 (blocking the protein substrate binding site), of the mitogen-activated protein (MAP) kinase, extracellular signal-related kinase 2 (ERK2) (by Tyr-185, blocking the peptide binding site) and by the C-terminal pseudosubstrate polypeptide in twitchin kinase (blocking the active site) [43–46]. Thus variability in the structure of the loop at the edge of the active site in protein kinase subclasses may provide diversity in the mechanism of regulation of their activities. Furthermore, catalytic activity often depends on the phosphorylation state of critical Thr or Tyr residues on these loops (e.g. in cAPK, ERK2, cdk2, cdc2, src and insulin receptor kinases). Although in most kinases these phosphorylation sites are conserved, it is not always clear whether they actually perform a regulatory function. In the EGF receptor kinase, phosphorylation of the corresponding conserved Tyr-845 is not necessary for biological activity [15]. Tyr-845 is surrounded by charged and polar residues, which might perform equivalent stabilizing functions. It has been suggested, on the basis of an earlier homology model, that Glu-842 and Glu-844 might play

important roles in stabilizing an active conformation around the active site [33].

In this paper we investigated the roles of residues Glu-842, Glu-844 and His-846 in the peptide phosphorylation and self-phosphorylation activities of the EGF receptor. Sequence alignment of members of the EGF receptor family showed that residue Glu-842 is conserved in at least six proteins, and is replaced by Asp in Erb B3 and by Ser in the *Drosophila* EGF receptor. Glu-844 is more highly conserved, being replaced by Gln in human Erb B3. His-846 is conserved in five proteins and is replaced by Gln in Xmrk, Leu in human Erb B3, and Lys in the *Drosophila* EGF receptor. Little is known about the relative activities and regulation of these homologues. We replaced Glu-842 and Glu-844 by conservative amino acids, Ser and Gln, and His-846 by Ala. Kinetic analysis of the single amino acid mutants showed that these substitutions affected the peptide phosphorylation and self-phosphorylation activities of the kinase to a moderate and variable extent. The $K_{m(MnATP)}^{app.}$ for the peptide phosphorylation reaction catalysed by the mutants Glu-842→Ser, Glu-844→Gln and His-846→Ala increased 2.5-, 4.9- and 4.2-fold respectively compared with the wild-type, while the values of $K_{m(MnATP)}$ in the absence of peptide (self-phosphorylation) were unaltered. Thus the ATP binding site was not affected directly, but only as a result of peptide binding. The affinity of the mutants for peptide substrates decreased, and comparison of changes in the value of $K_{m(SA\text{-}12)}^{app.}$ for Glu-842→Ser, Glu-844→Gln and His-846→Ala showed increases of 2.3-, 2.8- and 6.6-fold respectively. It would thus appear that the non-conservative mutation His-846→Ala had a significant effect on peptide substrate binding. The finding that all three mutants had similar $V_{max.(SA\text{-}12)}^{app.}$ values for peptide phosphorylation suggests that this peptide induces an optimal conformation at the active site, overcoming the effects of these mutations.

Although the mutations did not alter the peptide phosphorylation activity significantly, they caused 4.1-, 2.7.- and 7.3-fold decreases (for Glu-842→Ser, Glu-844→Gln and His-846→Ala respectively) in self-phosphorylation activity, as determined from initial reaction rates. Ammonium sulphate activated self-phosphorylation of all mutants to similar levels, possibly through stabilizing conformations that compensated for the amino acid changes. There was a striking difference among mutant and wild-type cytoplasmic domains with respect to the dependence of their initial rates of self-phosphorylation on protein concentration. Initial-rate dependences were linear, and in the case of His-846→Ala there was no measurable dependence of the initial rate on protein concentration. Logarithmic plots were also linear and had a slope of 1 for His-846→Ala (suggesting a first-order process) and slopes of 1.5–1.7 for the other mutants and the wild-type. Reaction orders of > 1 are in agreement with the well documented cross-phosphorylation (intermolecular) mechanism [47]. However, a reaction order of 1 is consistent with both intramolecular and intermolecular mechanisms (provided that intermolecular association is not rate-limiting). In view of the well documented intermolecular mechanism, the observed changes in the order of reaction may therefore not necessarily suggest changes in the molecularity of self-phosphorylation with reaction progress, though changes in molecularity are not excluded by our results or by the demonstration of the cross-phosphorylation mechanism. Changes in the reaction order and molecularity (if applicable) may conceivably result from extension of the conformation of the molecule upon phosphorylation [40], through increased accessibilities of multiple phospho-acceptor sites.

The reason for the difference in the susceptibilities of the peptide phosphorylation and self-phosphorylation activities to mutation is not clear. It is possible that peptide SA-12 and the autophosphorylation sites interact in different ways or induce different conformational changes around the active site upon binding. Large protein substrates may have restricted accessibility to the active site compared with small peptides; moreover, they may interact more extensively, e.g. at secondary sites, thus displaying more complex specificity requirements for binding. In addition, peptide SA-12, being highly negatively charged, may make different interactions with the active site compared with the autophosphorylation sites, which carry less negative charge. Mutations in the loop at the edge of the active-site cleft might affect sites involved in substrate binding directly, or cause unfavourable conformational changes that might impair autophosphorylation. A precedent for induction of different local and global conformational changes around the active-site cleft upon peptide substrate and inhibitor binding is provided by cAPK, where changes in the structure of peptide ligands can cause different conformational changes [48–50]. Crystallographic studies of a binary cAPK–peptide inhibitor complex suggested that peptide binding to this enzyme stabilizes the conformation of the large subunit [44]. Our present results suggest that mutants Glu-842→Ser, Glu-844→Gln and His-846→Ala have reduced affinities for peptide substrates, but peptide SA-12 can restore optimal catalytic activity for peptide phosphorylation. It is possible that peptide binding may stabilize the structure of the large lobe at the active-site cleft, restoring optimal peptide phosphorylation activity in the mutants.

A plausible homology model of the EGF receptor catalytic core was built using the co-ordinates of a high-resolution, closed conformation, cAPK structure [31] in order to aid interpretation of the effects of mutations around Tyr-845. The model accommodated the sequence Glu-841-Glu-Lys-Glu as an insertion in the surface loop at the edge of the active-site cleft, away from the catalytic base and ATP binding site, with the side chains of Tyr-845 and Glu-842 facing outwards. The guanidino group of Arg-812 seemed to interact with the carboxyl group of Glu-848 and the hydroxyl of Tyr-867. The carboxylate of Glu-848 interacted with the amino group of Lys-836 and the imidazole of His-846. These two residues were also within hydrogen-bonding distance of the carboxylate of Glu-844. Glu-842 faced away from the catalytic core into the solvent, and might interact with other regions of the cytoplasmic domain. The finding that the $V_{max.}^{app.}$ for self-phosphorylation in the mutants decreased 3–7-fold is consistent with a stabilizing role for these residues. The model does not offer a direct explanation for the observed reduction in affinity of the mutants for peptide substrates, since detailed interactions between tyrosine phospho-acceptor peptides and tyrosine kinases are not known. If such interactions are similar to those between cAPK and the protein kinase inhibitor peptide, the model suggests that these mutations might reduce substrate binding by indirect destabilization of the peptide binding site, rather than by direct interaction of residues Glu-842, Glu-844 and His-846 with the substrate.

This study suggests that residues Glu-844 and His-846, adjacent to the conserved Tyr-845 of the EGF receptor, make stabilizing interactions near the active site, and these interactions are important for maximal self-phosphorylation activity of the cytoplasmic kinase domain. These conclusions are supported by a plausible homology model. This model will also be a useful basis for further directed mutagenesis, in order to identify residues important in substrate–active-site interactions.

We thank the MRC, the Edward Penley Abraham Research Fund, and an ICI link grant for funding J. F. T., M. G. and M. E. M. N. We thank Professor L. N. Johnson for useful discussions and support, Dr. Ian Jones for supplying baculovirus vectors, Dr. Robert

Russell and Craig Livingstone for help with the sequence alignment programs, Dr. Athel J. Cornish-Bowden for the program LEONORA, and Mr. Stephen Lee for photography.

REFERENCES

1 Boni, S. M. and Pilch, P. F. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 7832–7836
2 Yarden, Y. and Schlessinger, J. (1987) Biochemistry **26**, 1434–1442
3 Canals, F. (1992) Biochemistry **31**, 4493–4501
4 Downward, J., Waterfield, M. D. and Parker, P. J. (1985) J. Biol. Chem. **260**, 14538–14546
5 Bertics, P. J. and Gill, G. N. (1985) J. Biol. Chem. **260**, 14642–14647
6 Taylor, S. S. and Radzio-Andzelm, E. (1994) Curr. Biol. Structure **2**, 345–355
7 Orellana, S. A. and McKnight, G. S. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 4726–4730
8 Steinberg, R. A., Cauthron, R. D., Symcox, M. M. and Shuntoh, H. (1993) Mol. Cell. Biol. **13**, 2332–2341
9 Solomon, M. J., Lee, T. and Kirschner, M. W. (1992) Mol. Biol. Cell **3**, 13–27
10 Connell, C. L., Solomon, M. J., Wei, N. and Harper, J. W. (1993) Mol. Biol. Cell **4**, 79–92
11 Knighton, D. R., Zheng, J. H., Ten, E. L., Ashford, V. A., Xuong, N. H., Taylor, S. S. and Sowadski, J. M. (1991) Science **253**, 407–414
12 Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H. and Huber, R. (1993) EMBO J **12**, 849–859
13 Ellis, L., Clauser, E., Morgan, D. O., Edery, M., Roth, R. A. and Rutter, W. J. (1986) Cell **45**, 721–732
14 Piwnica, W. H., Saunders, K. B., Roberts, T. M., Smith, A. E. and Cheng, S. H. (1987) Cell **49**, 75–82
15 Gotoh, N., Tojo, A., Hino, M., Yazaki, Y. and Shibuya, M. (1992) Biochem. Biophys. Res. Commun. **186**, 768–774
16 Hanks, S. K. and Quinn, A. M. (1991) Methods Enzymol. **200**, 38–62
17 Weyer, U., Knight, S. and Possee, R. D. (1990) J. Gen. Virol. **71**, 1525–1534
18 Livingstone, C. and Jones, I. (1989) Nucleic Acids Res. **17**, 2366
19 Kemp, B. E. and Pearson, R. B. (1991) Methods Enzymol. **200**, 121–134
20 Vieira, J. and Messing, J. (1987) Methods Enzymol. **153**, 3–11
21 Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 488–492
22 Brun, Y. V., Breton, R. and Lapointe, J. (1991) DNA Seq. **1**, 285–289
23 Garger, S. J., Griffith, O. M. and Grill, L. K. (1983) Biochem. Biophys. Res. Commun. **117**, 835–842
24 Summers, M. D. and Smith, G. E. (1987) Texas Agricultural Experiment Station Bulletin No. 1555
25 Overton, H. A., Fujii, Y., Price, I. R. and Jones, I. M. (1989) Virology **170**, 107–116
26 Fung, M. C., Chiu, K. Y., Weber, T., Chang, T. W. and Chang, N. T. (1988) J. Virol. Methods **19**, 33–42
27 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
28 Pike, L. J. (1987) Methods Enzymol. **146**, 353–362
29 Barton, G. J. and Sternberg, M. J. (1990) J. Mol. Biol. **212**, 389–402
30 Bairoch, A. and Boeckmann, B. (1991) Nucleic Acids Res. **19**, 2247–2249
31 Zheng, J., Knighton, D. R., Ten, E. L., Karlsson, R., Xuong, N., Taylor, S. S. and Sowadski, J. M. (1993) Biochemistry **32**, 2154–2161
32 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Acta Crystallogr. Sect. A **47**, 110–119
33 Knighton, D. R., Cadena, D. L., Zheng, J., Ten, E. L., Taylor, S. S., Sowadski, J. M. and Gill, G. N. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 5001–5005
34 Brünger, A. T. (1992) X-PLOR, A System for X-ray crystallography and NMR, Yale University Press, New Haven
35 Morris, A. L., MacArthur, M. W., Hutchinson, E. G. and Thornton, J. M. (1992) Proteins Structure Function Genet. **12**, 354–364
36 Luthy, R., McLachlan, A. D. and Eisenberg, D. (1991) Proteins Structure Function Genet. **10**, 229–239
37 Overington, J., Donnelly, D., Johnson, M. S., Sali, A. and Blundell, T. L. (1992) Protein Sci. **1**, 216–226
38 Gregoriou, M., Jones, P. F., Timms, J. F., Yang, J. J., Radford, S. E. and Rees, A. R. (1995) Biochem. J., **306**, 667–678
39 Wedegaertner, P. B. and Gill, G. N. (1989) J. Biol. Chem. **264**, 11346–11353
40 Cadena, D. L., Chan, C. L. and Gill, G. N. (1994) J. Biol. Chem. **269**, 260–265
41 Honegger, A. M., Kris, R. M., Ullrich, A. and Schlessinger, J. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 925–929
42 Singh, J. (1994) Protein Eng. **7**, 849–858
43 De Bondt, H., Rosenblatt, J., Jancarik, J., Jones, H. D., Morgan, D. O. and Kim, S. H. (1993) Nature (London) **363**, 595–602
44 Karlsson, R., Zheng, J., Xuong, N.-H., Taylor, S. S. and Sowadski, J. M. (1993) Acta Crystallogr. Sect. D **49**, 381–388
45 Hu, S. H., Parker, M. W., Lei, J. Y., Wilce, M. C., Benian, G. M. and Kemp, B. E. (1994) Nature (London) **369**, 581–584
46 Zhang, F., Strand, A., Robbins, D., Cobb, M. H. and Goldsmith, E. J. (1994) Nature (London) **367**, 704–711
47 Honegger, A. M., Schmidt, A., Ullrich, A. and Schlessinger, J. (1990) Mol. Cell. Biol. **10**, 4035–4044
48 Reed, J., Kinzel, V., Kemp, B. E., Cheng, H. C. and Walsh, D. A. (1985) Biochemistry **24**, 2967–2973
49 Olah, G. A., Mitchell, R. D., Sosnick, T. R., Walsh, D. A. and Trewhella, J. (1993) Biochemistry **32**, 3649–3657
50 Zheng, J., Knighton, D. R., Xuong, N. H., Taylor, S. S., Sowadski, J. M. and Ten, E. L. (1993) Protein Sci. **2**, 1559–1573
51 Barton, G. J. (1993) Protein Eng. **6**, 37–40
52 Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D. and Seeburg, P. H. (1984) Nature (London) **309**, 418–425

Received 4 October 1994/15 December 1994; accepted 4 January 1995