Biochemistry

 Volume 32, Number 44 November 9, 1993

Perspectives in Biochemistry

Structural Conservation in the CheY Superfamily†

Karl Volz

Department of Microbiology and Immunology, University of Illinois at Chicago, 835 South Wolcott Avenue, Chicago, Illinois 60612

Received June 15, 1993; Revised Manuscript Received August 9, 1993

Bacteria adaptively respond to a wide variety of stimuli encountered in their unpredictable environment. Their behavioral responses enable establishment and maintainance of infectious states in host organisms, as well as survival in the external milieu. Bacterial adaptation is accomplished through the coordinated activation of many different sensory receptors, signal processing proteins, and an array of molecular machinery designed to execute the appropriate reactions. Key control points for many of these behavioral responses are signal processing pathways based on an archetypal two-component paradigm. The primary sensing component of such a system is an autophosphorylating protein kinase, while the second component—the response regulator—becomes activated after receiving the phosphoryl group from the first. "Two-component" systems have been identified in as many as 20 distinct bacterial signal transduction pathways, ranging over 30 prokaryotic genera. Operations of the individual systems vary widely according to their specific tasks, but for response regulation, it is likely that they all utilize the same activation mechanism of Mg^{2+}-dependent protein phosphorylation. The molecular logic behind preservation of a common signaling mechanism in so many different pathways may lie in the ability to modulate crosstalk, thereby achieving a complex, concerted measure of response.

Although response regulators exhibit great diversity in their domain organization and full-length primary sequences, they all share similarity in their regulatory domain. One member of this superfamily is CheY, the single domain regulator of the bacterial chemotaxis system. CheY is the only response regulator for which a detailed three-dimensional structure is known. In this report, the CheY molecule is presented as the structural model for the regulatory domains of all bacterial two-component systems. A multiple alignment of 79 regulatory domain sequences is analyzed. The correlation of the conserved amino acid sites with specific structural features of CheY provides new insights into the structure and function of all members of the "CheY superfamily".

MOLECULAR STRUCTURE OF CHEY

Overall Folding and Secondary Structural Elements. The three-dimensional structure of CheY from both the *Salmonella typhimurium* (Stock et al., 1989a) and *Escherichia coli* species (Volz & Matsumura, 1991) has been solved. CheY is a small, globular protein of 128 amino acid residues. The molecule folds in a straightforward $(\beta/\alpha)_5$ manner, where the five β strands are arranged in a parallel sheet with the topology $\beta2$–$\beta1$–$\beta3$–$\beta4$–$\beta5$. CheY is ~40% α helix and ~20% β strand, while ~25% of the molecule is in turns bracketing the regions of repetitive secondary structure. The remaining ~15% of unclassified structure appears in loops of varying size. The three internal β strands, $\beta1$, $\beta3$, and $\beta4$, comprise the structural core of CheY. The five amphiphilic α helices cluster into two groups lying on opposite sides of the central β sheet, so that much of the two outermost strands ($\beta2$ and $\beta5$) is solvent accessible. CheY's topology is reminiscent of the dinucleotide binding fold but lacks one β/α segment in the first half of that well-known structural class.

Active Site Structure. The phosphorylation region of CheY lies near the carboxy terminus of the central β sheet. It includes five important amino acid side chains from the strands $\beta1$, $\beta3$, $\beta4$, and $\beta5$. At one side of the active site, D12 and D13 are grouped on the same face of the turn following $\beta1$ and project toward the active site. The central strand, $\beta3$, contributes D57, the residue phosphorylated in the activation of CheY (Sanders et al., 1989). The adjacent strand, $\beta4$, holds T87 in a position such that its γ-hydroxyl group participates in the hydrogen-bonding network of the active site solvent molecules. Lastly, K109, at the end of the outermost $\beta5$ strand, has its side chain fully extending back into the active site, with a firm hydrogen bond to the carboxyl group of D57. The D57–

† Supported by National Institutes of Health Grant GM47522.

0006-2960/93/0432-11741$04.00/0

Table I: Multiple Sequence Alignment of CheY Superfamily[a-d]

```
                          10        20        30                  40        50        60
EcoCheY          MADKELKFLVVDDFSTMRRIVRNLLKELGF--------NNVEEAEDGVDALNKLQAGGYGFVISDWNMPN---MD
PaeCheY               MKILIVDDFSTMRRIIKNLLRDLGF--------TNTAEADDGTTALPMLHSGNFDFLVTDWNMPG---MT
BsuCheY            MAHRILIVDDAAFMRMMIKDILVKNGF--------EVVAEAENGAQAVEKYKEHSPDLVTMDITMPE---MD
BsuSpo0F           MNEKILIVDDQYGIRILLNEVFNKEGY---------QTFQAANGLQALDIVTKERPDLVLLDMKIPG---MD
BjaORF138   ...PSSTKPTVFVVDDDAAVLGSLRFLLETDGF---------AVRTFRSGTALLNAGGAPGADCYVIDYKMPD---IN
XcaXcc1     ...ESRPDRTLLLLDDEENVLRSLVRLFRRDGY---------RILAAGNVRDAFDLLATNDVQVILSDQRMSD---MS
RcaRegA     ...ELGSDRSLLLVDDDNAFLTRLARAMEKRGF---------QTEIAETVSAGKAIVQNRAPAYAVIDLRLED---GN

RmeNtrC            MTGATILVADDDAAIRTVLNQALSRAGY---------DVRITSNAATLWRWIAAGDGDLVVTDVVMPD---EN
RcaNtrC            MDGTVLVADDDRTIRTVLTQALTRAGC---------KVHATASLMTLMRWVEEGKGDLVISDVVMPD---GN
KpnNtrC            MQRGIAWIVDDDSSIRWVLERALTGAGL---------SCTTFESGNEVLDALTTKTPDVLLSDIRMPG---MD
AcaNtrX            MAHDILIVDDEPDISGLVAGILEDEGY---------SARTARDADGALAEIAARRPNLIFLDIWLQGSR-LD
PaePilR            MSRQKALIVDDEPDIRELLEITLGRMKL---------DTRSARNVKEARELLAREPFDLCLTDMRLPD---GS
EcoHydG          MTHDNIDILVVDDDISHCTILQALLRGWGY---------NVALANSGRQALEQVREQVFDLVLCDVRMAE---MD
PaeAlgB       METTSEKQGRILLVDDESAILRTFRYCLEDEGY---------SVATASSAPQAEALLQRQVFDLCFLDLRLGE---DN
RleDctD            MDTLMPVALIDDDKDLRRATAQTLELAGF---------SVSAYDGAKAALADLPADFAGPVVTDIRMPE---ID
StyPgtA          MLNDECSILLIDDDVDVLDAYTQMLEQAGY---------RVRGFTHPFEAKEWVKADWEGIVLSDVCMPG---CS
AeuHoxA          MSDKQATVLVVDDETRSQDALRRTLDEEF----------DVLTAQGAEAAIAILEEEWVQVIICDQRMPG---LT
RcaHupRI         MAASAPAILLVDDEPHSLAAMKLALEDDF----------DVLTAQGAEAAIAILEEEWVQVIICDQRMPG---RT
BthRteB            MDKTKIIVVEDNIVYCEYVCNMLSREAY---------RNMKAYHLSTAKKHLQQATDNDIVVADLRLPD---GS
CcrFlbD              MRLLVVGKLNGQLSVAVKMAMNAGA---------KVSHVETTEQATNALRAGQGADLLMVDYVLD----IA

BjaFixJ            MTTKGHIYVIDDDAAMRDSLNFLLDSAGF---------GVTLFDDAQAFLDALPGLSFGCVVSDVRMPG---LD
AcaFixJ            MPESLPVHVIDDDDAVRESLAFLLESSGL---------AVTQHTSAAAFLDAGVPLDRGCIVTDVRMPG---TS
RmeFixJ            MTDYTVHIVDDEEPVRKSLAFMLTMNGF---------AVKMHQSAEAFLAFAPDVRNGVLVTDLRMPD---MS
BjaNodW     ...EASTKAIVFVVEDDISMRRSLTNLFRSVGL---------EVVAFGSAREMLQSTMPDVTSCLVLDVRLPG---LS
EcoUvrC2           MINVLLVDDHELVRAGIRRILEDIKG--------IKVVGEASCGEDAVKWCRTNAVDVVLMDMSMPG---IG
PflGacA            MIRVLVVDDHDLVRTGITRMLADIDG--------LQVVGQAESGEESLLKARELKPYVVLMDVKMPG---IG
EcoNarL          MSNQEPATILLIDDHPMLRTGVKQLISMAPD--------ITVVGEASNGEQGIELAESLDPDLILLDLNMPG---MN
BsuDegU            MTKVNIVIIDDHQLFREGVKRILDFEPT--------FEVVAEGDDGDEAARIVEHYHPDVVIMDINMPN---VN
EcoUhpA            MITVALIDDHLIVRSGFAQLLGLEPD--------LQVVAEFGSGREALAGLPGRGVQVCICDISMPD---IS
PaeAgmR            MYKILIADDHPLFREAIHNVIADGFPG-------SEVMETADLDSALGLTQEHDDLDLILLDLNMPG---MH
BpeBvgA            MYNKVLIIDDHPVLRFAVRVLMEKEGF---------EVIGETDNGIDGLKIAREKIPNLVVLDIGIPK---LD
PdeMoxX     ...DDARPLQILIVDDHPVVAEGWGRIIRTKTA---------CEIASAPSASEGWRAWRQARPDLMVVDLSIGRNK-IA
EcoEvgA            MNAIIIDDHPLAIAAIRNLLIKNDI---------EILAELTEGGSAVQRVETLKPDIVIIDVDIPG---VN
PaeTrpX            MSKVLIVDDHPAIRLAVRLLFERDGF---------TMSREADNGAEALQVARKKSPDLAILDIGIPK---ID
BmeGdhR          MNKKAWTVLLIEDDPMVQEVNRQFIEQVEGF--------TVIAAASNGLEGVQLIKQHQPDLTIIDMYMPS---QD
EcoRcsB            MNNMNVIIADDHPIVLFGIRKSLEQIEW--------VNVVGEFEDSTALINNLPKLDAHVLITDLSMPGDKYGD
BsuComA            MKKILVIDDHPAVMEGTKTILETDSNL-------SVDCLSPEPSEQFIKQHDFSSYDLILMDLNLGGE--VN
EcoFimZ            MKPTSVIIMDTHPIIRMSIEVLLQKNSE--------LQIVLKTDDYRITIDYLRTRPVDLIIMDIDLPG---TD

EcoOmpR          MQENYKNLVVDDDMRLRALLERYLTEQGF---------QVRSVANAEQMDRLLTRESFHLMVLDLMLPG---ED
PaerOmpR           MQKEKILVIDDEASIRRILETRLSIIGY---------DVISAADGEEALSIFKREHPNLVVLDLMMPK---LD
StyPhoP              MMRVLVVEDNALLRHHLKVQLQDSGH---------QVDAAEDAREADYYLNEHLPDIAIVDLGLPD---ED
BsuPhoP            MNKKILVVDDEESIVTLLQYNLERSGY---------DVITASDGEEALKKAETEKPDLIVLDVMLPK---LD
EcoPhoB            MARRILVVEDEAPIREMVCFVLEQNGF---------QPVEAEDYDSAVNQLNEPWPDLILLDWMLPG---GS
PaePhoB            MVGKTILIVDDEAPIREMIAVAVEMAGY---------ECLEAENTQQAHAVIVDRKPHLILLDWMLPG---TS
BsuORFX17       MDQTNETKILVVDDEARIRRLLRMYLERENY---------AIDEAENGDEAIAKGLEANYDLIILDLMMPG---TD
RcaPetR          MMSASPPHLLIVDDDERIRGLLQKFLIRNGF---------LVTAGRDAAHARRLLSGLEFNLIVLDVMMPG---ED
ScoAfsQ1           MPSLLLIEDDEAERTALELSLTRQGH---------RVATAASGEDGLKLLREQRPDLIVLDVMLPG---ID
EcoKdpE            MTNVLIVEDEQAIRRFLRTALEGDGM---------RVFEAETLQRGLLEAATRKPDLIILDLGLPD---GD
Ssp7942SphR ...AASPASRILVVEDEAVIRDMVALVLQQEGFTVDVAADGRTALNYFRSDSPEAGSVTENPDLVVLDLMLPA---VN
FdiRcaC1           MKILLVEDDDVLIKVLTKNLTTHHY---------IVDVVKDGEMGWTYGSTFEYDLIILDIMLPN---LD
StyTctD            MRLLLAEDNRELAHWLEKALVQNGF---------AVDCVFDGLAADHLLHSEMYALAVLDINMPG---MD
SliCutR            MRVLVVEDEQLLADAVATGLRREAM---------AVDVVYDGAAALERIGVNDYDVVVLDRDLPL---VH
PsyCopR            MKLLVAEDEPKTGIYLQQGLREAGF---------NVDRVVTGTDAVDQALNEAYDLLILDVMMPG---LD
EcoPcoR          MQRILIVEDEQKTGRYLQQGLVEEGY---------QADLFNNGRDGLGAASKGQYDLIILDVMLPF---LD
EfaVanR            MSDKILIVDDEHEIADLVELYLKNENY---------TVFKYYTAKEALECIDKSEIDLAILDIMLPG---TS
EcoCreB          MQRETVWLVEDEQGIADTLVYMLQQEGF---------AVEVFERGLPVLDKARKQVPDVMILDVGLPD---IS
EcoArcA            MQTPHILIVEDELVTRNTLKSIFEAEGY---------DVFEATDGAEMHQILSEYDINLVIMDINLPG---KN
RspLcrB            MPHLIIVDDDPRIRSMLSRYLEDEGF---------RVRLAENISQLRRVLSPSVDLVLLDIGLPD---GN
BfrRprY          MDEKLRILLCEDDENLGMLLREYLQAKGY---------SAELYPDGEAGFKAFLKNKYDLCVFDVMMPK---KD
AtuVirG            MKHVLLVDDDVAMRHLIIEYLTIHAF---------KVTAVADSTQFTRVLSSATVDVVVVDLNLGR---ED
EcoHnr          MTQPLVGKQILIVEDEQVFRSLLDSWFSSLGA---------TTVLAADGVDALELLGGFTPDLMICDIAMPR---MN

PaeAlgR1           MNVLIVDDEPLARERLARLVGQLDGY------RVLEPSASNGEEALTLIDSLKPDIVLLDIRMPG---LD
EcoBarA     ...ESKLAMTVMAVDDNPANLKLIGALLEDMVQ---------HVELCDSGHQAVERAKQMPFDLILMDIQMPD---MD
PsyLemA     ...LSSRAPRVLCVDDNPANLLLVQTLLEDMGA---------EVVAVEGGYAAVNAVQQEAFDLVLMDVQMPG---MD
EcoArcB     ...MPLPALNVLLVEDIELNVIVARSVLEKLGN---------SVDVAMTGKAALEMFKPGEYDLVLLDIQLPD---MT
MxaFrzE     ...PAAKRLRVLLVDDSPIARATEGALVKALGH---------SVEEAQDGEEAYVKVQNNTYDLILTDVQMPK---LD
EcoRcsC     ...SDNDDMMILVVDDHPINRRLLADQLGSLGY---------QCKTANDGVDALNVLSKNHIDIVLSDVNMPN---MD
BpeBvgS     ...KPQVSLRVLVVDDHKPNLMLLRQQLDYLGQ---------RVIAADSGEAALALWREHAFDVVITDCNMPG---IS
BsuSpo0A           MEKIKVCVADDNRELVSLLSEYIEGQED--------MEVIGVAYNGQECLSLFKEKDPDVLVLDIIMPH---LD
XcaRpfC     ...ARVRSMRMLVADDHEANRMVLQRLLEKAGH---------KVLCVNGAEQVLDAMAEEDYDAVIVDLHMPG---MN
CacORF3     ...LLNEYSKNLVVDDEFIMRQGITHMIDWEKEGF------QIIGQASNGQEALEMIKKDIPDIIISDVVMPQ---IN
KpnMrkE     ...MSGEKMKVIIVEDEFLAQQELSWLINTHSQ--------MEIVGSFDDGLDVLKFLQHNKVDAIFLDININPS---LD
StyCheB            MSKIRVLSVDDSALMRQIMTEIINSHSD--------MEMVATAPDPLVARDLIKKFNPDVLTLDVEMPR---MD
AspPatA     ...TDKKIYTIFCIDENPIVLNNIKNFLDDQIF---------AVIGVTDSLKALMEILCTKPDIILINVDMPD---LD
BthRteA     ...HNNKFHDVVAIDNDEVLLLMLKEMYSQEGI---------HCDTCTDAAALMEMIRQKEYSLLLTDLNMPG---IN
LinORF2     ...PNGRPYQVLIAENSRFQAKQLAQILESEGY--------QVIGFAENGKELVKLYDEHRLVDLITLDLNLPV---MD
SauAgrA            MKIFICEDDPKQRENMVTIIKNYIM-IEEKPMEIALATDNPYEVLEQAKNMNDIGCYFLDIQLSTD--IN
AtuVirA     ...PRGNGEVALIEPDPVLREVYEDKIAALGY---------EPVGFKTCADLCNWISKGKQADLVLVDQSSLP---EN
AtuVirAL    ...PLGSGEIVAVVEPDPVTLERYEETIAAFGY---------EPVGFNTFKGLTDWVLAGKEADIVLIDHSSFL---DG

Consensus                L VDD     R       L   G             V         A          DL   D  MP
```

Table I (Continued)

```
     70              80              90             100             110             120
GLELLKTIRADGAMS-----ALPVLMVTA--EAKKENIIAAAQAGA-SGYVVKPFTAATLEEKLNKIFEKLGM          1
GIDLLRAVRADERLK-----HLPVLMVTA--EAKRDQIIEAAQAGV-NGYVVKPFTAQVLKEKIEKIFERVNG         2
GITALKEIKQIDA-------QARIIMCSA--MGQQSMVIDAIQAGA-KDFIVKPFQADRVLEAINKTLN              3
GIEILKRMKVIDE-------NIRVIIMTA--YGELDMIQESKELGA-LTHFAKPFDIDEIRDAVKKYLPLKSN         4
GIELASRLRKSDG-------ETPVILITG--YPDENISTRAAAAGV-KDVVLKPLLDENLLKRIRRAIQDRPRA        5
GTEFLGRVKMLYP-------DTVRLVLSG--YTDLATVTEAINRGAIYRFLTKPWNDDELREHIRQAFRTHDELRNGRE   6
GLEVVEALRERRP-------EARIVVLTG--YGAIATAVAAVKMGA-TDYLSKPADANDITNALLAKGEALPP...     7

AFDLLPRIKKARP-------DLPVLVMSA--QNTFMTAIKASEKGA-YDYLPKPFDLTELIGIIGRALAEPKR...     8
GLEALARIARERP-------GLPVIVISA--QNTIMTAIQAAEADA-YDYLPKPFDLPDLMKRAARALELTRR...     9
GLALLKQIKQRHP-------MLPVIIMTA--HSDLDAAVSAYQQGA-FDYLPKPFDIDEAVALVDRAISHYQE...    10
GLELLDIIKREHP-------EVPVVMISG--HGNIETAVAAIKRGA-YDFIEKPFNADRLVVITERALETLRL...    11
GLDLVQYIQQRHP-------QTPVAMITA--YGSLDTAIQALKAGA-FDFLTKPVDFDFRLRELVATALRLRN...    12
GIATLKEIKALNP-------AIPVLIMTA--YSSVETAVEALKTGA-LDYLIKPLDFDNLQATWKKRSHTHSI...    13
GLDVLAQMRVQAP-------WMRVVIVTA--HSAVDTAVDAMQAGA-VDYLVKPCSPDQLRLAAAKQLEVRQL...    14
GLQLFATLQGMDV-------DLPVILMTG--HGDIPMAVQAIQDGA-YDFIAKPFAADRLVQSVRRASEKRRL...    15
GIDLMTLFHQDDD-------QLPILLITG--HGDVPMAVDAVKKGA-WDFLQKPSIRAKLLILIEDALRQRRS...    16
GVEFLKEVRERWP-------EIVRIVISG--YTDSEDIIAGVNEAGIYQYILKPWVPDHLIDTVRQAVEAQGL...    17
GVDFLTEVRERWP-------ETVRIIITG--YTDSASMMAAINDAGIHQFLTKPWHPEQLLSSARNAARMFTL...    18
GIDLLCWMRKEGK-------MQPFIIMTD--YAEVNTAVESMKLGS-IDYIPKQLVEDKLVPLIRSILKERQA...    19
GLIAANEAERM---------RVPVVACGV--DADPMRAANAIKAGA-KEFIPLPPDAELIAAVLAAVTDDEKP...    20

GIELLKRMKAQQS-------PFPILIMTG--HGDVPLAVEAMKLGA-VDFLEKPFEDDRLTAMIESAIRQAEP...    21
GLELLKELNARGA-------HMAVIVMTG--HGDVPLAVEAMKLGA-ADFLEKPFDDAAIIAAVRASLGRSAE...    22
GVELLRNLGDLKI-------NIPSIVITG--HGDVPMAVEAMKAGA-VDFIEKPFEDTVIIEAIERASEHLVA...    23
GLDYQTELARLNI-------HIPIIFITG--HGDIPMTVRAMKGGA-VDFLSKPFRDQELLDAVVAATERDRK...    24
GLEATRKIARSTA-------DVKIIMLTV--HTENPLPAKVMQAGA-AGYLSKGAAPQEVVSAIRSVYSGQRY...    25
GLEATRKLLRSHP-------DIKVVAVTV--CEEDPFPTRLLQAGA-AGYLTKGAGLNEMVQAIRLVFAGQRY...    26
GLETLDKLREKSL-------SGRIVVFSV--SNHEEDVVTALKRGA-DGYLLKDMEPEDLLKALHQAAAGEMV...    27
GVEATKQLVELYP-------ESKVIILSI--HDDENYVTHALKTGA-RGYLLKEMDADTLIEAVKVVAEGGSY...    28
GLELLSQLPK----------GMATIMLSV--HDSPALVEQALNAGA-RGFLSKRCSPDELIAAVHTVATGGCY...    29
GLNGLMNLRNEAP-------TIPVVIVSA--EQDKQVVLQAITYGA-VGFITKSSPRAQMTEAIEQILNGNVY...    30
GLEVIARLQSLGL-------PLRVLVLTG--QPPSLFARRCLNSGA-AGFVCKHENLHEVINAAKAVMAGYTY...    31
GIRLIERLRRVDP-------DLPILVFTM--HRSPVLARRALMFGA-NGIINKDSPPAEICAAFTEVARGGNY...    32
GIQVLETLRKRQY-------SGIIIIVSA--KNDHFYGKHCADAGA-NGFVSKKEGMNNIIAAIEAAKNGYCY...    33
GLEVIARLKSLKL-------DTKVLVLTR-QNRSQFARRLQAGPWA-SSAKGKPLRAAARRQGVLAGGALRSI...    34
GLTTLQQIRANGY-------KTDVIAVTA--ASDIETVRKVLQYGA-VDYIMKPFKFERMKQALEQYRSFQVK...    35
GITLIKYIKRHFP-------SLSIIVLTM--NNNPAILSAVLDLDI-EGIVLKQGAPTDLPKALAALQKGKKF...    36
GMELSKQILQENP-------HCKIIVYTG--YEVEDYFEEAIRAGL-HGAISKTESKEKITQYIYHVLNGEIL...    37
GFTFLKRIKQIQS-------TVKVLFLSS--KSECFYAGRAIQAGA-NGFVSKCNDQNDIFHAVQMILSGYTF...    38

GLSICRRLRSQSN-------PMPIIMVTA--KGEEVDRIVGLEIGA-DDYIPKPFNPRELLARIRAVLRRQAN...    39
GYGVCQELRKESD--------VPIIMLTA--LSDVSDRITGLELGA-DDYIVKPFSPKELEARIRSVLRRVDK...    40
GLSLIRRWRSSDV-------SLPVLVLTA--REGWQDKVEVLSSGA-DDYVTKPFHIEEVMARMQALMRRNSG...    41
GIEVCKQLRQQKL-------MFPILMLTA--KDEEFDKVLGLELGA-DDYMTKPFSPREVNARVKAILRRSEI...    42
GIQFIKHLKRESMTR-----DIPVVMLTA--RGEEEDRVRGLETGA-DDYITKPFSPKELVARIKAVMRRISP...    43
GIELARRLKRDELTL-----DIPIIMLTA--KGEEDNKIQGLEAGA-DDYITKPFSPRELVARLKAVLRRTGP...    44
GIEVCRQIREKKA--------TPIIMLTA--KGEEANRVQGFEAGT-DDYIVKPFSPREVVLRVKALLRRASQ...    45
GLSLTRDLRTKMA--------TPILLLTA--RGETRERIEGLEAGA-DDYLPKPFEPKELLLRINAILRRVPE...    46
GFEVCRRIRRTDQ--------LPIILLTA--RNDDIDVVVGLESGA-DDYVVKPVQGRVLDARIRAVLRRGER...    47
GIEFIRDLRQWSR--------VPVIVLSA--RSEESDKIAALDAGA-DDYLSKPFGIGELQARLRVALRRHSA...    48
GLDFCRLLRRQGV-------TVPILMLSA--KDTETDRVVGLEIGA-DDYLTKPFGTRELVARCRALLRRSQN...    49
GISLCKRFRAQGY-------TVPILLLTA--QDNITAKVQGLDAGA-DDYVVKPFDPIELIARIRALLRRGSN...    50
GLEVVQRLRKRGQ-------TLPVLLLTA--RSAVADRVKGLNVGA-DDYLPKPFELEELDARLRALLRRSAG...    51
GDDVCRKIVELGM-------PTRVLMLTA--SGDVSDRVEGLEIGA-DDYLPKPFAFSELIARVRALGRRTSV...    52
GWEVIRRLRTAGQ-------AVPVLFLTA--RDGVDDRVKGLELGA-DDYLVKPFALSELLARVRTLLRRGSS...    53
GWQIISALRESGH-------EEPVLFLTA--KDNVRDKVKGLELGA-DDYLIKPFDFTELVARVRTLLRRARS...    54
GLTICQKIRDKHT--------YPIIMLTG--KDTEVDKITGLTIGA-DDYITKPFRPLELIARVKAQLRRYKK...    55
GFELCRQLLALHP-------ALPVLFLTA--RSEEVDRLLGLEIGA-DDYVAKPFSPREVCARVRTLLRRVKK...    56
GLLLARELREQA--------NVALMFLTG--RDNEVDKILGLEIGA-DDYITKPFNPRELTIRARNLLSRTMN...    57
GLELAREIDANFR--------VPTIIVSG--RDDDVDRIIGLEMGA-DDYVSKPFNLRELLARVRSVLRRSQR...    58
GFTLAQEVRAANA-------EIPIIFLTA--KTLKEDILEGFKIGA-DDYITKPFSMEELTFRIEAILRRVRG...    59
GLEIVRNLAAKS--------DIPIIISG-DRLEETDKVVALELGA-SDFIAKPFSIREFLARIRVALRVRPN...    60
GLKLLEHIRNRGD-------QTPVLVISA--TENMADIAKALRLGV-EDVLLKPVKDLNRLREMVFACLYPSM...    61

GLQVAARLCEREA-------PPAVIFCTA----HDEFALEAFQVSA-VGYLVKPVRSEDLAEALKKASRPNRV...    62
GIRACELIHQLPHQQ------QTPVIAVTA--HAMAGQKEKLLGAGM-SDYLAKPIEEERLHNLLLRYKPGSGI...    63
GRQATEAIRAWEAERNQS--SLPIVALTA--HAMANEKRSLLQSGM-DDYLTKPISERQLAQVVLKWTGLALR...    64
GLDISRELTKRYPRED-----LPPLVALTA---NVLKDKQEYLNAGM-DDVLSKPLSVPALTAMIKFWDTQDDE...    65
GFSLARRLKSTPAVA------RIPVIILSS--LASPEDKRRGLDAGA-DAYLVKGELGVEVLAQAIDRLT          66
GYRLTQRIRQLGL--------TLPVIGVTA--NALAEEKQRCLESGM-DSCLSKPVTLDVIKQSLTLYAERVRKSRDS  67
GYELARRIRAAEAAPGYGRTRCILFGFTA--SAQMDEAQACRAAGM-DDCLFKPIGVDALRQRLNEAVARAAL...    68
GLAVLERLRESDLKK-----QPNVIMLTA--FGQEDVTKKAVDLGA-SYFILKPFDMENLVGHIRQVSGNASS...    69
GLDMLKQLRVMQASGMR---YTPVVVLSA--DVTPEAIRACEQAGA-RAFLAKPVLAAKLLDNPGRSGSEHPA...    70
GVELAKFVQDKYP--------KIRIIILSS--YSDFEYVKSSFQNGA-VDYILKPSLNPSELLSTLKKISIKMK...    71
GVLLAQNISQFAH-------KPFIVFITA----WKEHAVEAFELEA-FDYILKPYQESRIINMLQKLTTAWEQ...    72
GLDFLEKLMRLR--------PMPVVMVSSLTGKGSEVTLRALELGA-IDFVTKPQLGIREGMLAYSEMIAEKV...    73
GYELCSLLRKHSYFK-----NTPVIMVTE--KAGLVDRARAKIVRA-SGHLTKPFNQGDLLKVIFKHIT          74
GFELLELLRSSNVGNSP---TIPVVVATA---SGSCNKGELLAKGF-AGCLFKPFSISELMEVSDRCAIKATP...  75
GYATFFEIKGKGV--------LPRIVIVSE--ENTPAVLKNLIDEGA-MDYIPKPIKREKILEKVNAAIKKVPKV   76
GIKLGSEIRKHDP--------VGNIIFVTS--HSELTYLTFVYKVAA-MDFIFKDDPAELRTRIIDCLETAHTR...  77
GSATALHAAF----------KTASIIIGG---SDLKMSLSSDDMTS-ALFLPKPISSRTMAYAIRTKIKA        78
GRVGSWVATL----------EKVPIIMIG--QHQKNVSISADGEAS-NHFLQKPVSSKALAYVVRANIRTE       79

G                    PVI  TA              GA  DY  KP     L
```

Footnotes to Table I

[a] Organism abbreviations: Aca, *Azorhizobium caulinodans*; Aeu, *Alcaligenes eutrophus*; Arh, *Agrobacterium rhizogenes*; Asp, *Anabaena* sp. 7120; Atu, *Agrobacterium tumefaciens*; Bfr, *Bacteroides fragilis*; Bja, *Bradyrhizobium japonicum*; Bme, *Bacillus megaterium*; Bpa, *Bradyrhizobium parasponiae*; Bpar, *Bordetella parapertussis*; Bpe, *Bordetella pertussis*; Bsu, *Bacillus subtilis*; Bth, *Bacteroides thetaiotaomicron*; Cac, *Clostridium acetobutylicum*; Cca, *Cyanidium caldarium*; Ccr, *Caulobacter crescentus*; Cje, *Campylobacter jejuni*; Eco, *Escherichia coli*; Efa, *Enterococcus faecium*; Fdi, *Fremyella diplosiphon*; Kpn, *Klebsiella pneumoniae*; Lin, *Leptospira interrogans*; Lmo, *Listeria monocytogenes*; Mtu, *Mycobacterium tuberculosis*; Mxa, *Myxococcus xanthus*; Pae, *Pseudomonas aeruginosa*; Paer, *Porphyridium aeruginium*; Pde, *Paracoccus denitrificans*; Pfl, *Pseudomonas fluorescens*; Psy, *Pseudomonas syringae*; Pvu, *Proteus vulgaris*; Rca, *Rhodobacter capsulatus*; Rle, *Rhizobium leguminosarum*; Rme, *Rhizobium meliloti*; Rsp, *Rhizobium* sp.; Sau, *Staphylococcus aureus*; Sce, *Saccharomyces cerevisiae*; Sco, *Streptomyces coelicolor*; Sdy, *Shigella dysenteriae*; Sli, *Streptomyces lividans*; Ssp7002, *Synechococcus* sp. PCC 7002; Ssp7942, *Synechococcus* sp. PCC 7942; Ssp6803, *Synechocystis* sp. PCC 6803; Sty, *Salmonella typhimurium*; Wsu, *Wolinella succinogenes*; Xca, *Xanthomonas campestris*; Yen, *Yersinia enterocolitica*; Yps, *Yersinia pseudotuberculosis*. [b] Sources of sequences—EcoCheY: Matsumura, P., Rydel, J. J., Linzmeier, R., & Vacante, D. (1984) *J. Bacteriol. 160*, 36–41; Mutoh, N., & Simon, M. I. (1986) *J. Bacteriol. 165*, 161–166. PaeCheY: Starnbach, M. N., & Lory, S. (1992) *Mol. Microbiol. 6*, 459–469. BsuCheY: Bischoff, D., & Ordal, G. W. (1991) *J. Biol. Chem. 266*, 12301–12305. BsuSpo0F: Trach, K. A., Chapman, J. W., Piggot, P. J., & Hoch, J. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7260–7264; Yoshikawa, H., Kazami, J., Yamashita, S., Chibazakura, T., Sone, H., Kawamura, F., Oda, M., Isaka, M., Kobayashi, Y., & Saito, H. (1986) *Nucleic Acids Res. 14*, 1063–1072; Trach, K., Chapman, J. W., Piggot, P., leCoq, D., & Hoch, J. A. (1988) *J. Bacteriol. 170*, 4194–4208. BjaORF138: Anthamatten, D., & Hennecke, H. (1991) *Mol. Gen. Genet. 225*, 38–48. XcaXcc1: Osbourn, A. E., Clarke, B. R., Stevens, B. J. H., & Daniels, M. J. (1990) *Mol. Gen. Genet. 222*, 145–151. RcaRegA: Sganga, M. W., & Bauer, C. E. (1992) *Cell 68*, 945–954. RmeNtrC: Szeto, W. W., Nixon, B. T., Ronson, C. W., & Ausubel, F. M. (1987) *J. Bacteriol. 169*, 1423–1432. RcaNtrC: Jones, R., & Haselkorn, R. (1989) *Mol. Gen. Genet. 215*, 507–516. KpnNtrC: Buikema, W. J., Szeto, W. W., Lemley, P. V., Orme-Johnson, W. H., & Ausubel, F. M. (1985) *Nucleic Acids Res. 13*, 4539–4555; Drummond, M., Whitty, P., & Wootton, J. (1986) *EMBO J. 5*, 441–447. AcaNtrX: Pawlowski, K., Klosse, U., & de Bruijn, F. J. (1991) *Mol. Gen. Genet. 231*, 124–138. PaePilR: Ishimoto, K. S., & Lory, S. (1992) *J. Bacteriol. 174*, 3514–3521; M. Hobbs, E. S. Collie, P. D. Free, S. P. Livingston, and J. S. Mattick, unpublished work, 1992. EcoHydG: Stoker, K., Reijnders, W. N. M., Oltmann, L. F., & Stouthamer, A. H. (1989) *J. Bacteriol. 171*, 4448–4456. PaeAlgB: Wozniak, D. J., & Ohman, D. E. (1991) *J. Bacteriol. 173*, 1406–1413; Goldberg, J. B., & Dahnke, T. (1992) *Mol. Microbiol. 6*, 59–66. RleDctD: Ronson, C. W., Astwood, P. M., Nixon, B. T., & Ausubel, F. M. (1987) *Nucleic Acids Res. 15*, 7921–7934; Watson, R. J. (1990) *Mol. Plant Microb. Interact. 3*, 174–181. StyPgtA: Yu, G.-Q., & Hong, J.-S. (1986) *Gene 45*, 51–57. AeuHoxA: Eberz, G., & Friedrich, B. (1991) *J. Bacteriol. 173*, 1845–1854. RcaHupR1: Xu, H.-W., & Wall, J. D. (1991) *J. Bacteriol. 73*, 2401–2405; Richaud, P., Colbeau, A., Toussiant, B., & Vignais, P. M. (1991) *J. Bacteriol. 173*, 5928–5932. BthRteB: Stevens, A. M., Sanders, J. M., Shoemaker, N. B., & Salyers, A. A. (1992) *J. Bacteriol. 174*, 2935–2942. CcrFlbD: Ramakrishnan, G., & Newton, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2369–2373. BjaFixJ: Anthamatten, D., & Hennecke, H. (1991) *Mol. Gen. Genet. 225*, 38–48. AcaFixJ: Kaminski, P. A., & Elmerich, C. (1991) *Mol. Microbiol. 5*, 665–673. RmeFixJ: David, M., Daveran, M.-L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P., & Kahn, D. (1988) *Cell 54*, 671–683. BjaNodW: Göttfert, M., Grob, P., & Hennecke, H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2680–2684. EcoUvrC2: Moolenaar, G. F., van Sluis, C. A., Backendorf, C., & van de Putte, P. (1987) *Nucleic Acids Res. 15*, 4273–4289. PflGacA: Laville, J., Voisard, C., Keel, C., Maurhofer, M., Défago, G., & Haas, D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 1562–1566. EcoNarL: Gunsalus, R. P., Kalman, L. V., & Stewart, R. R. (1989) *Nucleic Acids Res. 17*, 1965–1975; Nohno, T., Noji, S., Taniguchi, S., & Saito, T. (1989) *Nucleic Acids Res. 17*, 2947–2957; Stewart, V., Parales, J., Jr., & Merkel, S. M. (1989) *J. Bacteriol. 171*, 2229–2234. BsuDegU: Henner, D. J., Yang, M., & Ferrari, E. (1988) *J. Bacteriol. 170*, 5102–5109; Kunst, F., Débarbouillé, M., Msadek, T., Young, M., Mauel, C., Karamata, D., Klier, A., Rapoport, G., & Dedonder, R. (1988) *J. Bacteriol. 170*, 5093–5101; Tanaka, T., & Kawata, M. (1988) *J. Bacteriol. 170*, 3593–3600; Weston, L. A., & Kadner, R. J. (1988) *J. Bacteriol. 170*, 3375–3383. EcoUhpA: Friedrich, M. J., & Kadner, R. J. (1987) *J. Bacteriol. 169*, 3556–3563. PaeAgmR: Schweizer, H., (1991) *J. Bacteriol. 173*, 6798–6806. BpeBvgA: Aricó, B., Miller, J. F., Roy, C., Stibitz, S., Monack, D., Falkow, S., Gross, R., & Rappuoli, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6671–6675. PdeMoxX: Harms, N., Reijnders, W. N. M., Anazawa, H., van der Palen, C. J. N. M., van Spanning, R. J. M., Oltmann, L. F., & Stouthamer, A. H. (1993) *Mol. Microbiol. 8*, 457–470. EcoEvgA: R. Utsumi, unpublished work, 1992. PaeTrpX: Hadero, A., & Crawford, I. P. (1986) *Mol. Biol. Evol. 3*, 191–204; Henikoff, S., Wallace, J. C., & Brown, J. P. (1990) *Methods Enzymol. 183*, 111–132. BmeGdhR: Mitamura, T., Ebora, R. V., Nakai, T., Makino, Y., Negoro, S., Urabe, I., & Okada, H. (1990) *J. Ferment. Bioeng. 70*, 363–369. EcoRcsB: Stout, V., & Gottesman, S. (1990) *J. Bacteriol. 172*, 659–669. BsuComA: Weinrauch, Y., Guillen, N., & Dubnau, D. A. (1989) *J. Bacteriol. 171*, 5362–5375. EcoFimZ: Muramatsu, S., & Mizuno, T. (1990) *Mol. Gen. Genet. 220*, 325–328. EcoOmpR: Wurtzel, E. T., Chou, M.-Y., & Inouye, M. (1982) *J. Biol. Chem. 257*, 13685–13691; Comeau, D. E., Ikenaka, K., Tsung, K., & Inouye, M. (1985) *J. Bacteriol. 164*, 578–584; Nara, F., Matsuyama, S.-I., Mizuno, T., & Mizushima, S. (1986) *Mol. Gen. Genet. 202*, 194–199. PaerOmpR: Kessler, U., Maid, U., & Zetsche, K. (1992) *Plant Mol. Biol. 18*, 777–780. StyPhoP: Miller, S. I., Kukral, A. M., & Mekalanos, J. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5054–5058; Groisman, E. A., Chiao, E., Lipps, C. J., & Heffron, F. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7077–7081. BsuPhoP: Seki, T., Yoshikawa, H., Takahashi, H., & Saito, H. (1987) *J. Bacteriol. 169*, 2913–2916. EcoPhoB: Makino, K., Shinagawa, H., Amemura, M., & Nakata, A. (1986) *J. Mol. Biol. 190*, 37–44. PaePhoB: Anba, J., Bidaud, M., Vasil, M. L., & Lazdunski, A. (1990) *J. Bacteriol. 172*, 4685–4689. BsuORFX17: A. Sorokin, personal communication. RcaPetR: Gabellini, N., & Sebald, W. (1986) *Eur. J. Biochem. 154*, 569–579; Davidson, E., & Daldal, F. (1987) *J. Mol. Biol. 195*, 13–24; Davidson, E., & Daldal, F. (1987) *J. Mol. Biol. 195*, 25–29; Henikoff, S., Wallace, J. C., & Brown, J. P. (1990) *Methods Enzymol. 183*, 111–132; Tokito, M. K., & Daldal, F. (1992) *Mol. Microbiol. 6*, 1645–1654. ScoAfsQ1: Ishizuka, H., Horinouchi, S., Kieser, H. M., Hopwood, D. A., & Beppu, T. (1992) *J. Bacteriol. 174*, 7585–7594. EcoKdpE: Walderhaug, M. O., Polarek, J. W., Voelkner, P., Daniel, J. M., Hesse, J. E., Altendorf, K., & Epstein, W. (1992) *J. Bacteriol. 174*, 2152–2159. Ssp7942SphR: Aiba, H., Nagaya, M., & Mizuno, T. (1993) *Mol. Microbiol. 8*, 81–91. FdiRcaC1: G. G. Chiang, M. R. Schaefer, and A. R. Grossman, unpublished work, 1992. StyTctD: Widenhorn, K. A., Somers, J. M., & Kay, W. W. (1989) *J. Bacteriol. 171*, 4436–4441. SliCutR: Tseng, H.-C., & Chen, C. W. (1991) *Mol. Microbiol. 5*, 1187–1196. PsyCopR: Mills, S. D., Jasalavich, C. A., & Cooksey, D. A. (1993) *J. Bacteriol. 175*, 1656–1664. EcoPcoR: Brown, N. L., Rouch, D. A., & Lee, B. T. O. (1992) *Plasmid 27*, 41–51. EfaVanR: Arthur, M., Molinas, C., & Courvalin, P. (1992) *J. Bacteriol. 174*, 2582–2591; Handwerger, S., Discotto, L., Pucci, M. J., & Thanassi, J. (1992) *FEMS Microbiol. Lett. 92*, 11–14. EcoCreB: Amemura, M., Makino, K., Shinagawa, H., & Nakata, A. (1986) *J. Bacteriol. 168*, 294–302. EcoArcA: Drury, L. S., & Buxton, R. S. (1985) *J. Biol. Chem. 260*, 4236–4242. RspLcrB: Upadhyaya, M. M., Scott, K. F., Tucker, W. T., Watson, J. M., & Dart, P. J. (1992) *Mol. Plant Microb. Interact. 5*, 129–143. BfrRprY: Rasmussen, B. A., & Kovacs, E. (1993) *Mol. Microbiol. 7*, 765–776. AtuVirG: Melchers, L. S., Thompson, D. V., Idler, K. B., Schilperoort, R. A., & Hooykaas, P. J. J. (1986) *Nucleic Acids Res. 14*, 9933–9942; Winans, S. C., Ebert, P. R., Stachel, S. E., Gordon, M. P., & Nester, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8278–8282; Jin, S., Roitsch, T., Christie, P. J., & Nester, E. W. (1990) *J. Bacteriol. 172*, 531–537. EcoHnr: A. Contreras, unpublished work, 1992. PaeAlgR1: Deretic, V., Dikshit, R., Konyecsni, W. M., Chakrabarty, A. M., & Misra, T. K. (1989) *J. Bacteriol. 171*, 1278–1283. EcoBarA: Nagasawa, S., Tokishita, S., Aiba, H., & Mizuno, T. (1992) *Mol. Microbiol. 6*, 799–807. PsyLemA: Hrabak, E. M., & Willis, D. K. (1992) *J. Bacteriol. 174*, 3011–3020. EcoArcB: Iuchi, S., Matsuda, Z., Fujiwara, T., & Lin, E. C. C. (1990) *Mol. Microbiol. 4*, 715–727. MxaFrzE: McCleary, W. R., & Zusman, D. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5898–5902. EcoRcsC: Stout, V., & Gottesman, S. (1990) *J. Bacteriol. 172*, 659–669. BpeBvgS: Stibitz, S., Aaronson, W., Monack, D., & Falkow, S. (1989) *Nature 338*, 266–269; Aricó, B., Miller, J. F., Roy, C., Stibitz, S., Monack, D., Falkow, S., Gross, R., & Rappuoli, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6671–6675; Stibitz, S., & Yang, M.-S. (1991) *J. Bacteriol. 173*, 4288–4296; Aricò, B., Scarlato, V., Monack, D M., Falkow, S., & Rappuoli, R. (1991) *Mol. Microbiol. 5*, 2481–2491. BsuSpo0A: Ferrari, F. A., Trach, K., leCoq, D., Spence, J., Ferrari, E., & Hoch, J. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 2647–2651; Ikeuchi, T., Kudoh, J., & Tsunasawa, S. (1986) *Mol. Gen. Genet. 203*, 371–376. XcaRpfC: Tang, J.-L., Liu, Y.-N., Barber, C. E., Dow, J. M., Wootton, J. C., & Daniels, M. J. (1991) *Mol. Gen. Genet. 226*, 409–417. CacORF3: Hancock, K. R., Rockman, E., Young, C. A., Pearce, L., Maddox, I. S., & Scott, D. B. (1991) *J. Bacteriol. 173*, 3084–3095. KpnMrkE: Allen, B. L., Gerlach, G.-F., & Clegg, S. (1991) *J. Bacteriol. 173*, 916–920. StyCheB: Simms, S. A., Keane, M. G., & Stock, J. (1985) *J. Biol. Chem. 260*, 10161–10168. AspPatA: Liang, J., Scappino, L., & Haselkorn, R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5655–5659. BthRteA: Stevens, A. M., Sanders, J. M., Shoemaker, N. B., & Salyers,

Footnotes to Table I

A. A. (1992) *J. Bacteriol. 174*, 2935–2942. LinORF2: M. Ding and D. B. Yelton, unpublished work, 1990. SauAgrA: Peng, H.-L., Novick, R. P., Kreiswirth, B., Kornblum, J., & Schlievert, P. (1988) *J. Bacteriol. 170*, 4365–4372. AtuVirA: Morel, P., Powell, G. K., Rogowsky, P. M., & Kado, C. I. (1989) *Mol. Microbiol. 3*, 1237–1246. AtuVirAL: Leroux, B., Yanofsky, M. F., Winans, S. C., Ward, J. E., Ziegler, S. F., & Nester, E. W. (1987) *EMBO J. 6*, 849–856. [c] Sequences with greater than 60% identity—StyCheY: Stock, A., Koshland, D. E., Jr., & Stock, J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7989–7993. AtuNtrC: Wardhan, H., McPherson, M. J., & Sastry, G. R. (1989) *Mol. Plant Microb. Interact. 2*, 241–248. BpaNtrC: Nixon, B. T., Ronson, C. W., & Ausubel, F. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7850–7854. EcoNtrC: Miranda-Ríos, J., Sánchez-Pescador, R., Urdea, M., & Covarrubias, A. A. (1987) *Nucleic Acids Res. 15*, 2757–2770. PvuNtrC: U. Steglitz-Moersdorf, G. Moersdorf, and H. Kaltwasser, unpublished work, 1992. RmeDctD: Jiang, J., Gu, B., Albright, L. M., & Nixon, B. T. (1989) *J. Bacteriol. 171*, 5244–5253. StyHydG: Chopra, A. K., Peterson, J. W., & Prasad, R. (1991) *Biochim. Biophys. Acta 1129*, 115–118. StyUhpA: Island, M. D., Wei, B.-Y., & Kadner, R. J. (1992) *J. Bacteriol. 174*, 2754–2762. StyFimZ: D. L. Swenson and S. Clegg, unpublished work, 1992. StyOmpR: Liljeström, P., Laamanen, I., & Palva, E. T. (1988) *J. Mol. Biol. 201*, 663–673. CcaOmpR: Kessler, U., Maid, U., & Zetsche, K. (1992) *Plant Mol. Biol. 18*, 777–780; Maid, U., & Zetsche, K. (1992) *Plant Mol. Biol. 19*, 1001–1010. EcoPhoP: Groisman, E. A., Heffron, F., & Solomon, F. (1992) *J. Bacteriol. 174*, 486–491; Kasahara, M., Nakata, A., & Shinagawa, H. (1992) *J. Bacteriol. 174*, 492–498. SdyPhoB: Lee, T.-Y., Makino, K., Shinagawa, H., Amemura, M., & Nakata, A. (1989) *J. Bacteriol. 171*, 6593–6599. KpnPhoB: Lee, T.-Y., Makino, K., Shinagawa, H., Amemura, M., & Nakata, A. (1989) *J. Bacteriol. 171*, 6593–6599. AtuVirG6: Melchers, L. S., Thompson, D. V., Idler, K. B., Schilperoort, R. A., & Hooykaas, P. J. J. (1986) *Nucleic Acids Res. 14*, 9933–9942; Winans, S. C., Ebert, P. R., Stachel, S. E., Gordon, M. P., & Nester, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8278–8282. AtuVirGS: Chen, C.-Y., Wang, L., & Winans, S. C. (1991) *Mol. Gen. Genet. 230*, 302–309. AtuVirG5: Powell, B. S., Powell, G. K., Morris, R. O., Rogowsky, P. M., & Kado, C. I. (1987) *Mol. Microbiol. 1*, 309–316; Rogowsky, P. M., Powell, B. S., Shirasu, K., Lin, T.-S., Morel, P., Zyprian, E. M., Steck, T. R., & Kado, C. I. (1990) *Plasmid 23*, 85–106; Pergo, M., & Hoch, J. A. (1987) *Mol. Microbiol. 1*, 125–132. ArhVirG: Aoyama, T., Hirayama, T., Tamamoto, S., & Oka, A. (1989) *Gene 78*, 173–178. EcoCheB: Mutoh, N., & Simon, M. I. (1986) *J. Bacteriol. 165*, 161–166. BparBvgS: Stibitz, S., Aaronson, W., Monack, D., & Falkow, S. (1989) *Nature 338*, 266–269; Aricó, B., Miller, J. F., Roy, C., Stibitz, S., Monack, D., Falkow, S., Gross., R., & Rappuoli, R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6671–6675; Stibitz, S., & Yang, M.-S. (1991) *J. Bacteriol. 173*, 4288–4296; Aricò, B., Scarlato, V., Monack, D. M., Falkow, S., & Rappuoli, R. (1991) *Mol. Microbiol. 5*, 2481–2491. AtuVirAW: Leroux, B., Yanofsky, M. F., Winans, S. C., Ward, J. E., Ziegler, S. F., & Nester, E. W. (1987) *EMBO J. 6*, 849–856. AtuVirAB: Melchers, L. S., Thompson, D. V., Idler, K. B., Neuteboom, S. T., de Maagd, R. A., Schilperoort, R. A., & Hooykaas, P. J. J. (1987) *Plant Mol. Biol. 9*, 635–645. [d] Similar but incomplete sequences—YenPhoP: B. Wren, unpublished work, 1992. YpsPhoP: B. Wren, unpublished work, 1992. YenPhoB: B. Wren, unpublished work, 1992. YpsPhoB: B. Wren, unpublished work, 1992. CjeCreB: B. Wren, unpublished work, 1992. YenCreB: B. Wren, unpublished work, 1992. YpsCreB: B. Wren, unpublished work, 1992. CjeRegX: B. Wren, unpublished work, 1992. EcoRegX: B. Wren, unpublished work, 1992. LmoRegX: B. Wren, unpublished work, 1992. MtuRegX: B. Wren, unpublished work, 1992. Ssp7002ORF2: Schluchter, W. M., & Bryant, D. A. (1992) *Biochemistry 31*, 3092–3102. WsuPsrX: Krafft, T., Bokranz, M., Klimmek, O., Schröeder, I., Fahrenholz, F., Kojro, E., & Kröger, A. (1992) *Eur. J. Biochem. 206*, 503–510. RspPckA: Østerås, M., Finan, T. M., & Stanley, J. (1991) *Mol. Gen. Genet. 230*, 257–269. FdiRcaC2: G. G. Chiang, M. R. Schaefer, and A. R. Grossman, unpublished work, 1992. SceORF: Özkaynak, E., Finley, D., Solomon, M. J., & Varshavsky, A. (1987) *EMBO J. 6*, 1429–1439. Ssp6803ORF1: Chitnis, P. R., Reilly, P. A., Miedel, M. C., & Nelson, N. (1989) *J. Biol. Chem. 264*, 18374–18380.

K109 pair is essentially buried, showing a negligible solvent-accessible surface area of only 7 Å^2 for the functional groups of both residues. The acidic groups of residues D12 and D13 are also quite inaccessible (6-Å^2 exposed area, total); this is consistent with the anomalous pK_a's of the Mg^{2+} binding group(s) of CheY, estimated to be ~6.0 (Lukat et al., 1990).

There are four important properties of the hydrogen-bonding network within the active site of the unphosphorylated, apo-CheY structure: (1) all potential bonds for residues D57 and K109 are satisfied; (2) the solvent structure is highly ordered; (3) one SO_4^{2-} molecule is close to the active site, with an ionic bond to the ϵ-amino group of K109; and (4) one solvent molecule (originally designated as H_2O 2) is probably an NH_4^+ ion, residing in the Mg^{2+} binding region near D12 and D13. The prsence of NH_4^+ and SO_4^{2-} ions in the structure is most likely an artifact of the crystallization conditions [~2.3 M $(NH_4)_2SO_4$]. Under normal *in vivo* circumstances, a Mg^{2+} ion would probably reside in the binding region ($K_D/[Mg^{2+}] \ll 1.0$). The D57–K109 bond strength might also be expected to increase upon loss of the interfering SO_4^{2-} anion, although the ϵ-amino group would have to compete for the attraction of D57's carboxyl group to the bound divalent cation.

This arrangement of the five essential active site residues of CheY describes the unphosphorylated state of the molecule. Phosphorylation of the central residue D57 is certain to disrupt this constellation of side chains, but the structural consequences of the activation process are not yet known.

SEQUENCE ALIGNMENT OF REGULATORY DOMAINS

Two-Component Systems and Their Response Regulators. The first indication of similarity among bacterial response regulators was reported in 1985, when the amino acid sequence of CheY was shown to be related to regulatory proteins of other cellular processes, such as membrane protein synthesis and sporulation (Stock et al., 1985). Soon after, the "two-component" nature of these systems was described (Nixon et al., 1986), and their molecular commonalities began to emerge. Excellent reviews and analyses of the genetics and molecular biology of two-component systems have recently appeared (Bourret et al., 1991; Ninfa, 1991; Stock et al., 1991; Parkinson & Kofoid, 1992). The last compilations and comparisons of primary sequences of response regulators were published over 2 years ago (Albright et al., 1989; Stock et al., 1989b; Ninfa, 1991), when only 20–30 sequences were known. New two-component systems continue to emerge; presently there are over 100 response regulator sequences with homologous regulatory domains.

Search for Homologues. Response regulator homologues were identified and selected through multiple queries of the GENBANK (release 74.0) (Burks et al., 1991) and EMBL (release 30) (Hamm & Cameron, 1986) data banks. Thirty-one distantly related response regulator sequences were compared to translations of all six reading frames of the nucleotide sequences, using the program TFASTA (Pearson, 1990). The different probe sequences gave many redundant results but were successful in finding widely different sequences of low percent identity. The number of homologues retrieved by different probes was highly variable, due to the low overall degree of homology among response regulators. The ability of TFASTA to read across "termination codons" caused the largest percentage of erroneous matches. Use of a consensus sequence improved the efficiency but did not return any new sequences. The SWISPROT (Bairoch & Boekmann, 1991) and PIR (George et al., 1986) protein sequence data banks were also scanned and yielded four sequences not entered in EMBL or GENBANK. The entire search uncovered four ORFs not originally recognized as CheY homologues and six additional homologues not even recognized as ORFs. Nine sequences not in the data banks were found in the literature.

Criteria for Membership and Exceptions to the Rules. Results of the response regulator search are shown in Table I. Only full-length CheY homologues were included. There were many highly similar sequences (e.g., EcoCheY vs StyCheY, or variant forms of *Agrobacterium tumifaciens* VirG). To reduce the statistical influence of this large number of closely related sequences, only sequence pairs with less than 60% identity were retained (Table I, footnote *c*). Most true homologues showed good alignment scores in the search and were easily discriminated from spuriously matched sequences. All candidates were visually inspected for superfamily membership. The principal criterion was retention of the most highly conserved residues of the sequence collection, with special consideration given to the five functionally important residues equivalent to D12, D13, D57, T87, and K109 of CheY. The resulting consensus sequence in Table I is very similar to that originally proposed in an earlier analysis (Kofoid & Parkinson, 1988), using a smaller data base.

Six sequences—CcrFlbD, EcoFimZ, LinORF2, AspPatA, BthRteA, and AtuVirA—did not completely adhere to the functional group criteria. They did indeed fit the structural pattern and scored high in the sequence alignment calculations as well. These have been aptly termed "unorthodox" response regulators (Kofoid & Parkinson, 1988). Their deviance suggests that they may employ different types of activation mechanisms.

Seventeen incomplete entries (Table I, footnote *d*) did not span the entire length of the consensus sequence but did contain most of the essential attributes of the superfamily. They will likely show the full-length consensus once their entire sequences are known. One interesting member of this group is the SceORF, a partial sequence from a reverse open reading frame near the ubiquitin gene of *Saccharomyces cerevisiae*. This was the only convincing homologue found in sequences from eukaryotic species. This observation raises questions concerning the extent of two-component systems in the eukaryotic kingdom.[1]

Four additional response regulators have been identified (Ohta et al., 1992; Plamann, unpublished work; Shapiro, unpublished work; Zusman, unpublished work), but their sequences are not yet available. There are also cases where the existence of response regulators was inferred from identification of their cognate kinase sequences. Only two proteins reported to be response regulators (Perez-Casal et al., 1991; Taha et al., 1991) did not fit the above membership criteria. Finally, it should be mentioned that some sequences were possibly missed in this search, and perhaps others could be included by relaxing the criterion strigency (Kofoid & Parkinson, 1988).

Alignment of Primary Sequences. Alignment of the 79 sequences in Table I was done manually, using six prominent regions of CheY as registration guides: (1) β1, with D12 and D13; (2) β3, with D57; (3) α3, starting at G65; (4) β4, with T/S87; (5) α4, ending at the G102–A103 pair; and (6) the K109–P110 pair, following β5. Segments corresponding to the secondary structural elements of the CheY molecule were the same length for all sequences. Five gaps were put between the aligned sections. As can be seen in Table I, most gaps are caused by extra residues in just a few of the entries. The resulting alignments agree with automated alignment procedures [e.g., Feng and Doolittle (1987)] performed on subsets of the sequences (data not shown), with minor exceptions of placement of residues near the gaps. Some gap borders are highly variable, lacking sufficient information content to guide alignment, so residue placement within these regions was by default left or right justification. As would be expected, the two most variable regions (bordering gaps 1 and 3, which follow α1 and α3, respectively) are located on the surface of the CheY molecule most distant from the functional region.

Family Relationships. Previous compendia divided response regulator sequences into families according to homologies in their extra domains [e.g., Albright et al. (1989), Stock et al. (1989b), and Bourret et al. (1991)]. Following those examples, the entries in Table I are grouped in five families, arbitrarily named by proteins representative of each: CheY, NtrC, FixJ, OmpR, and "miscellaneous" [for an alternate, more systematic nomenclature, see Parkinson and Kofoid (1992)]. The CheY family comprises the short, "single domain" proteins. CheY and Spo0F are true single domain response regulators; the three other members contain 10–20 extra residues on their amino termini. Of them, RcaRegA has ~50 extra residues on its carboxy terminus. The functions of these regions are not known. All other response regulators are multidomain proteins. Most are positive transcriptional regulators, usually containing DNA recognition sequences in their carboxy-terminal domains. The NtrC type includes proteins with full-length homologies to NtrC, typically ~460 ± 20 residues in length, possessing two domains after the response regulator. The FixJ family contains sequences with one extra FixJ-type domain on their carboxy termini, giving a total length of ~220 residues. The OmpR family includes PhoP, PhoB, VirG, and other full-length OmpR homologues, all two domain, also ~230 amino acids long. The fifth group is not intended to be a single family but instead a miscellaneous list of smaller families and sequences lacking recognizable features common to the other groups.

The distribution of pairwise percent identities among the regulatory domains (Figure 1) generally shows that conservation is higher within families than across family lines. This is most pronounced for the OmpR family. High intrafamily conservation possibly reflects retention of features specific for interaction with their additional familial domains. However, some sequences show stronger interfamily relationships. For example, the carboxy-terminal domains of the FixJ and NodW proteins clearly define membership to the FixJ family, but their regulatory domains show greater similarity to those of the NtrC and OmpR families. The average percent identity score from all possible cross relationships in Table I is 23% (inset, Figure 1), with the lowest value of 6% between CheY and VirA. It may seem unusual that a superfamily with an average pairwise identity score of only 23% is considered highly conserved, but β/α protein families are known to contain a large proportion of structurally conserved amino acid sites which remain recognizable across wide evolutionary distances (Keim et al., 1981).

Functional Variety through Modular Design. The hypothesis that all response regulator domains share a common structure is decidedly well substantiated. It is also arguable that most response regulators utilize the same biochemical mechanism in their primary activation event. This appears plausible, considering the special conservation relationships among the sequences of the CheY superfamily. Functional divergence among these proteins is likely to arise from the interactions between the less conserved regions of the CheY-like domain and the accompanying, family-specific domains.

This modularity of domain organization illustrates the flexibility in functional selection of bacterial signal trans-

[1] One entry labeled as human partial cDNA sequence (EMBL HSA59A041) shows a high alignment score but is likely to be of prokaryotic origin. See Anderson (1993).

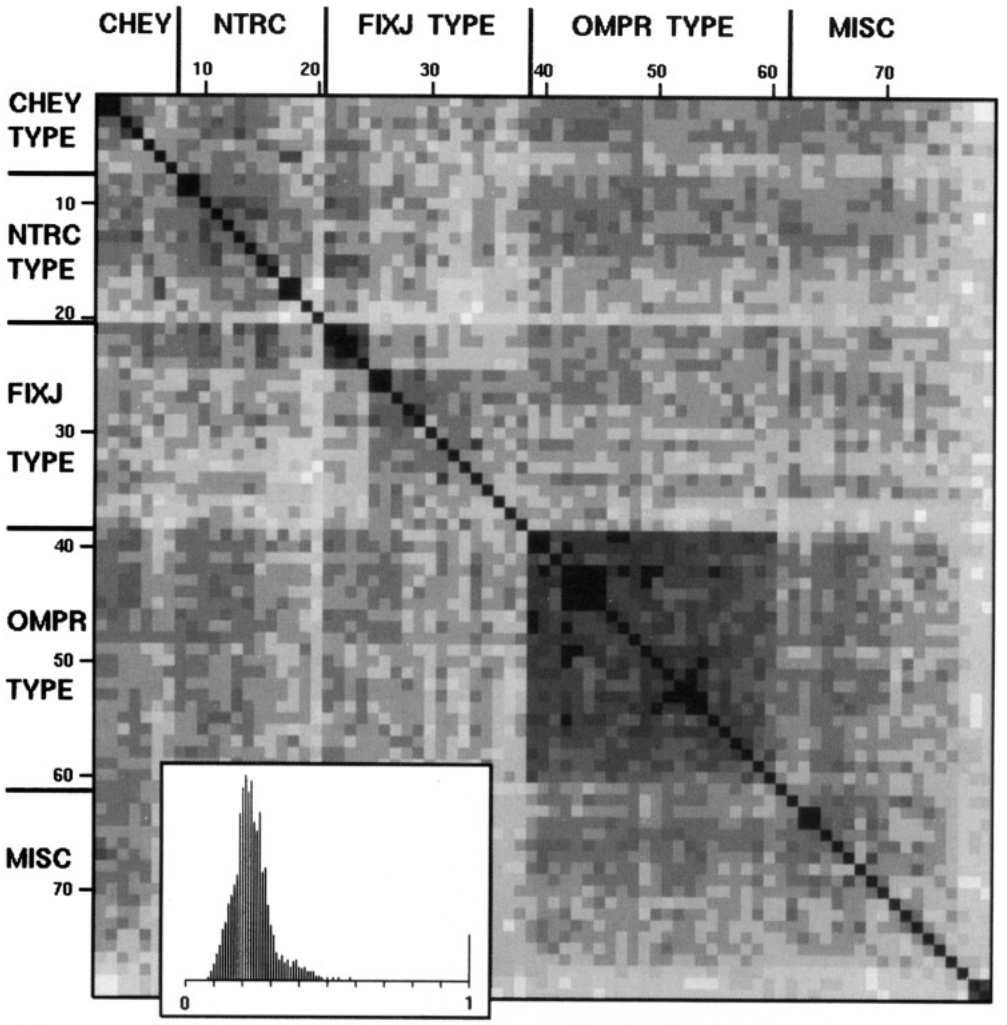

FIGURE 1: Percent identity matrix of the CheY superfamily. Sequences are aligned and sorted as in Table I. Pairwise percent identity scores are coded in nine gray scales, ranging from light to dark for values of 5–50% identity, in 5% increments. Data obscured by inset are visible on the other side of the matrix diagonal. Inset plot: distribution of pairwise percent identity scores. The average value for all cross-relationships is 0.23; values for individual families are CheY, 0.26; NtrC, 0.26; FixJ, 0.24; and OmpR, 0.37.

duction proteins, where the "mix and match" approach has been repeatedly exploited over time to achieve different biochemical activities. The evolutionary development of this extensive set of related regulatory proteins may have occurred through horizontal transmission (Parkinson & Kofoid, 1992). The functional variety found among response regulators rests in the successful coupling of different types of adjunct domains to the modified versions of the phospho-accepting regulator unit.

STRUCTURAL SIGNIFICANCE OF CONSERVED RESIDUES

Core and Tertiary Structure Conservation. A condensed but more informative representation of the multiple sequence alignment information is shown in the histogram of Figure 2, where the percent occurrence of each amino acid type is plotted as a function of residue position, CheY numbering. The most obvious structural feature of the CheY-like domain is the highly hydrophobic nature of the central β strands, β1, β3, and β4 (residues 7–11, 52–57, and 82–87). This type of core conservation has been recognized in different β/α protein families for well over a decade [e.g., Lifson and Sander (1979)]. The hydrophobic core also includes eight highly conserved sites from the surrounding α helices: sites 21 and 25 from α1; site 46 from α2; sites 68, 69, and 72 from α3; and sites 116 and 120 from α5; all presumably conserved for the sole sake of structural integrity. Twenty-six remaining contributions to the "outer shell" of hydrophobic residues come from the helices, the intervening loops, and the two outermost β strands (namely, positions 6, 17, 18, 24, 28, 30, 33, 36, 39, 42, 43, 45, 51, 60, 65, 66, 73, 81, 98, 103, 107, 111, 117, 123, 124, and 127). Three sites—58, 96, and 99—are unusual in that they are highly conserved as hydrphobic amino acids and yet, in the CheY molecule, have a high degree of solvent accessibility. These positions may be involved in interprotein contacts or interdomain contacts for the multidomain response regulators.

The five helices differ in their degree of variability for structural reasons. For instance, α3, bordered on either side by α2 and α5, is the most conserved, while the greatest variability is found in α2, the most exposed helix. The high variability in α5 is likely due to different demands placed on the structural transition to the family-specific carboxy-terminal domains. The higher intrafamily conservation of α5 supports this interpretation. In general, sections of highest variability in the primary sequence alignment coincide with regions of the CheY molecule most distant from the active site—implying that they are the regions most able to tolerate structural changes without detrimental functional consequences.

Secondary Structure Conservation. All of the other highly conserved amino acid positions coincide with important structural features of the CheY molecule (Table II). The largest example is the "γ-turn loop", a complicated and rather

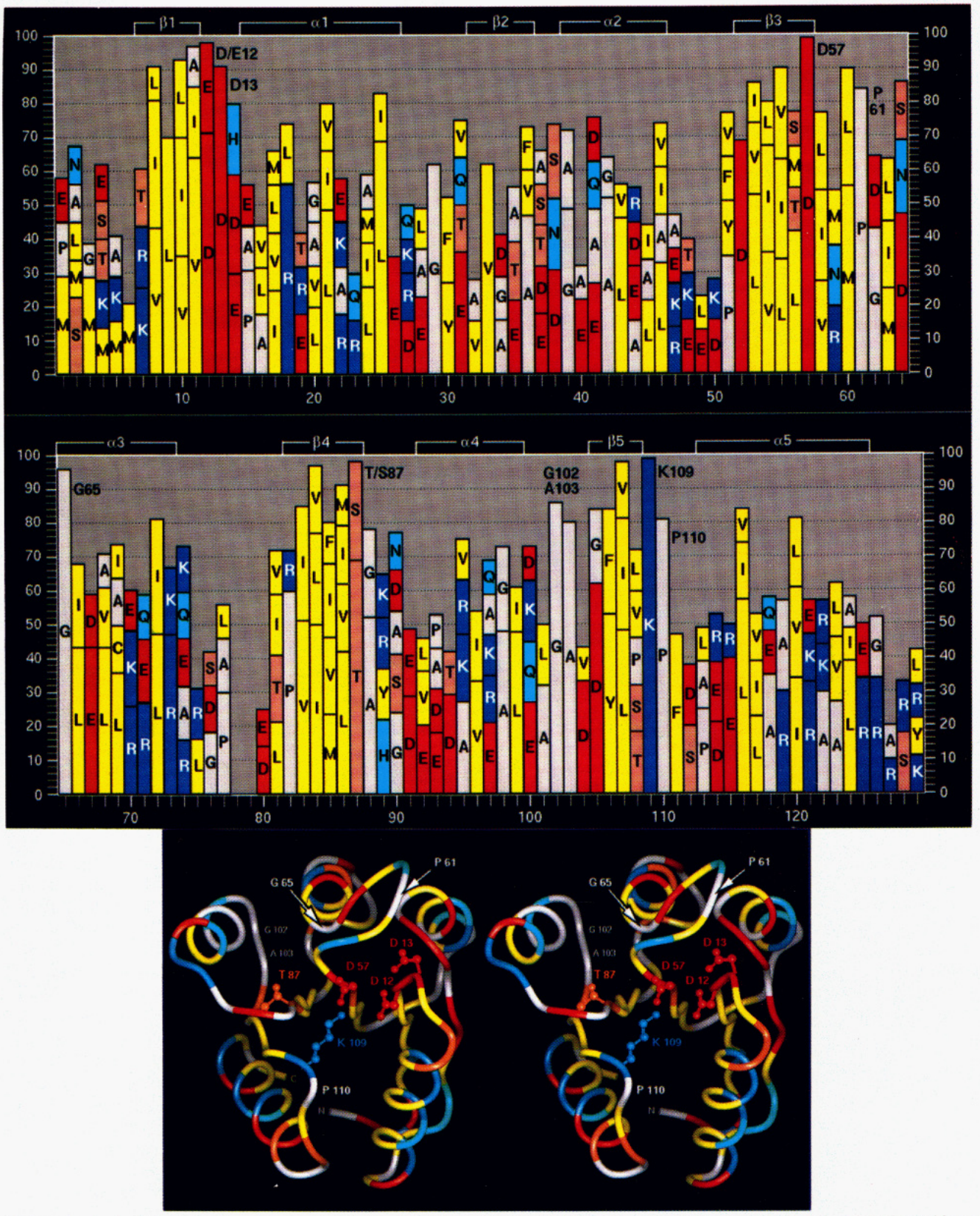

FIGURE 2: (a, top) Amino acid preference plot for the CheY superfamily. A histogram of the percent occurrence of each amino acid type is plotted against CheY numbering. Each bar represents the occurrence of each amino acid type divided by the total number of amino acids per site in the multiple alignment of Table I. Gaps are not counted. Values less than 10% are omitted for clarity. The color code is acidic (D, E), red; hydroxyl (S, T), rose; hydrophobic (C, F, I, L, M, V, W, Y), yellow; polar (H, N, Q), light blue; basic (K, R), dark blue; and structural (A, G, P), light gray. The secondary structural elements of the CheY molecule are shown at the top. (b, bottom) Stereoview of the overall structure of CheY. Amino acid residue types are color coded as in (a). Highly conserved positions are labeled. The side chains of D12, D13, D57, T87, and K109 are also shown.

rigid solvent-exposed structure, prominently located adjacent to the phosphorylation region of the molecule (Figure 3). In CheY, this loop includes the residues MPNMDG. The relative "immutability" of this surface loop is the consequence of a combination of five structural elements: (1) the exclusive occurrence of a buried hydrophobic residue at position 60 (M

Table II: Structural Roles of "Conserved" Residues in CheY

residue	role
F8	hydrophobic core sequence of internal β1 strand
L9	hydrophobic core sequence of internal β1 strand
V10	hydrophobic core sequence of internal β1 strand
V11	hydrophobic core sequence of internal β1 strand
D12	Mg^{2+} binding
D13	Mg^{2+} binding and catalysis
M17	hydrophobic cover of α1–α5 interface
R18	hydrophobic cover of α1–β1 interface
V21	hydrophobic core
L25	hydrophobic core
G29	α_L conformation; C-cap of α1
F30	hydrophobic cluster with F8, F53, and F124
V33	sole hydrophobic residue of exposed β2 strand; buried
L46	hydrophobic core
G52	aspartate by consensus; start of β3
F53	hydrophobic cluster with F8, F30, and F124
V54	hydrophobic core sequence of internal β3 strand
I55	hydrophobic core sequence of internal β3 strand
S56	hydrophobic core sequence of internal β3 strand
D57	invariant; required for phosphorylation-induced activation
W58	hydrophobic closure to α3–β4–α4 interface
M60	hydrophobic plug for crevice between β1, α2, β3, and α3
P61	rigid group in γ-turn loop
N62	positive ϕ conformation; required for γ-turn
D64	N-cap of α3
G65	internal start of α3; severe steric restrictions on β substituent
L66	internal residue of α3
L68	internal residue of α3; hydrophobic core
L69	internal residue of α3; hydrophobic core
L72	internal residue of α3; hydrophobic core
P82	start of β4
V83	hydrophobic core sequence of internal β4 strand
L84	hydrophobic core sequence of internal β4 strand
M85	hydrophobic core sequence of internal β4 strand
V86	hydrophobic core sequence of internal β4 strand
T87	invariant hydroxyl; proton donor to active site
A88	reduced bulk; required for access to D57 active site
I96	conserved solvent-accessible hydrophobic residue
G102	α_L conformation; C-cap of α4
A103	reciprocally covariant with position 106
Y106	conserved aromatic; reciprocally covariant with 103
V107	sole hydrophobic residue of exposed β5 strand; buried
K109	invariant; required for activation
P110	*cis* peptide following K109; rigidity to β5–α5 loop
F111	outer shell of hydrophobic core
L116	internal residue of α5; hydrophobic core
L120	internal residue of α5; hydrophobic core
F124	hydrophobic cluster with F8, F30, and F53

+ L + I, 98%); (2) a proline (84%) at position 61; (3) the γ turn centered at position 62 (G + D + E + N + K + R, 87%), imposing a positive ϕ value on that residue; (4) the N-capping residue of α3 at position 64 (D + N + S + T, 91%); and (5) the internal start of α3 at position 65 (G, 96%). Why is such an unusual, intricate structure so highly conserved? Its large degree of solvent accessibility and proximity to the phosphorylation region suggests that it plays a role in intermolecular contacts. The γ-turn loop might serve as a "trademark", retained by all response regulators for recognition by the kinases. Furthermore, residues in the nearby variable sites may contribute to different systems' specificities. One might also speculate that two-component kinases possess their own complementary recognition surface.

A second uncommon structural feature located in a highly conserved region is the type VIb turn involving the residues KPF, numbered 109–111 (Figure 3). This turn features a *cis* peptide bond between K109 (99%) and P110 (81%). The FixJ family exhibits the highest variability at position 110; when excluded, the percent identity for proline at 110 is 95%. Conservation of this turn may be indirectly related to function, perhaps to confer rigidity to the loop connecting β5 to α5. Hypothetically, if repositioning of the K109 backbone atoms occurs during the activation process, the effect would propagate more directly to surrounding parts of the molecule through a rigid link.

Another region in the active site periphery contains the two highly conserved positions, 87 and 88 (Figure 3). Except for the CcrFlbD and AtuVirA proteins, residue 87 is exclusively a hydroxyamino acid (T/S, 71/29%), likely for functional purposes. The side chain of position 88 appears to be restricted in bulk (A + G + S, 84%), probably a requirement for access to the phosphorylation region. The FixJ family is most variable at this position.

There are also many localized sites of high conservation that serve specialized secondary structural roles in the CheY molecule. For instance, conservation of arginine at position 18 (56%) is not for functional reasons but is instead due to the hydrophobic role of the arginine's aliphatic methylene groups in sealing the α1–β2 interface. This interpretation is based on the fact that arginine exchanges primarily with hydrophobic amino acids (L + A + V + I + C, 35%) at that site. The other structurally important arginine in CheY is at position 73. The entire guanidinium group of R73 is packed

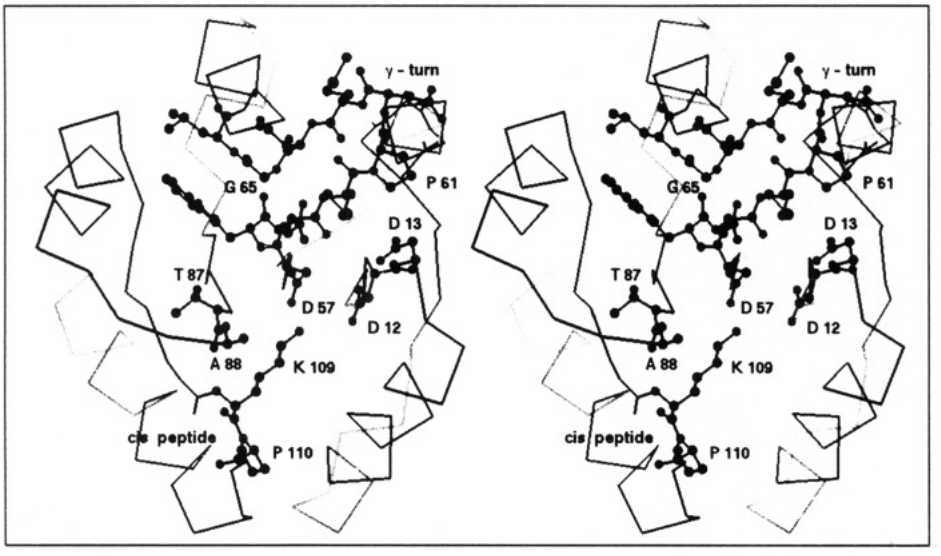

FIGURE 3: Stereo diagram of the CheY region containing the most highly conserved residues of the superfamily. The orientation is approximately the same as in Figure 2b. See discussion in text.

Table III: Spontaneous Point Mutations in CheY and CheY Homologues

organism–protein	mutation site	phenotypic/biochemical consequences[a]	CheY position	references[b]
RmeFixJ	V6M	decreased activity	8	Weinstein et al., 1992
AtuVirG	V7I	(double mutant with I106T) superactivator	11	Chen et al., 1991
EcoCheB	D10N	inactive; phosphorylation stable	12	Stewart et al., 1990; Stewart, 1993
EcoPhoB	E9K	loss of function	12	Yamada et al., 1989
EcoCheY	D13N	loss of function	13	Matsumura et al., unpublished work
RmeFixJ	E12G	decreased activity; phosphorylation defect	13	Weinstein et al., 1992
RmeFixJ	E13K	decreased activity	14	Weinstein et al., 1992
BsuSpo0A	N12K	pleotropic; change of function	14	Hoch et al., 1985
BsuSpo0A	N12K	pleotropic; change of function	14	Spiegelman et al., 1990
BsuDegU	H12L	pleotropic; constitutive activity	14	Henner et al., 1988
BsuDegU	H12L	phosphorylation stable; constitutive activity	14	Dahl et al., 1991, 1992
BsuSpo0A	E14V,A	sporulation-proficient suppressor (*sof*)	16	Spiegelman et al., 1990
EcoOmpR	R15C	altered osmoregulation	16	Nara et al., 1986
EcoOmpR	R15C	$OmpF^-$ $OmpC^+$; dephosphorylation defective	16	Aiba et al., 1989
EcoOmpR	L16Q	allele-specific phosphorylation defective	17	Matsuyama et al., 1986
EcoOmpR	L16Q	*envZ11* suppressor; phosphorylation defective	17	Aiba et al., 1989
BsuSpo0A	L15P	suppressor of Spo0E deletion (*sed*)	17	Perego & Hoch, 1991
AtuVirG	M13T	sensitized; increased activity	17	Han et al., 1992
AtuVirG	H15R	sensitized; increased activity	19	Han et al., 1992
EcoNarL	T20I	low-level constitutive	19	Egan & Stewart, 1991
EcoCheY	G39D	loss of function	39	Matsumura et al., unpublished work
EcoCheB	R42H	phosphotransfer defective; lower activity	43	Stewart, 1993
EcoOmpR	S48F	$OmpF^-$ $OmpC^-$; loss of function; intramolecular suppressor of D11N mutant	50	Brissette et al., 1991a
BsuDegU	D56N	nonphosphorylatable; loss of function	57	Dahl et al., 1991
BsuDegU	D56N	(double mutant with H12L) loss of function	57	Msadek et al., 1990
AtuVirA	D766N	decreased activity	57	Pazour et al., 1991
EcoNarL	D59N	loss of function	57	Egan & Stewart, 1991
AtuVirG	D52E	loss of function	57	Han et al., 1992
AtuVirG	L53D	loss of function	58	Han et al., 1992
AtuVirG	N54D	increased activity	59	Pazour et al., 1992
AtuVirG	N54D	lowered transcriptional activity	59	Han et al., 1992
AtuVirG	N54D	VirA-independent vir expression	59	Jin et al., 1993
EcoCheB	E58K	phosphorylation independent; increased activity	59	Stewart, 1993
BsuSpo0A	P60S	pleotropic; change of function	61	Olmedo et al., 1990
BsuSpo0A	P60S	pleotropic; change of function	61	Spiegelman et al., 1990
BsuSpo0A	P60S	intragenic suppressor of D10Q mutant	61	Green et al., 1991
EcoOmpR	R71T	constitutive	73	Nara et al., 1986
EcoCheB	R73H	decreased methylesterase activity	74	Stewart, 1993
BsuSpo0A	D75	deletion mutant(s); constitutively active	76	Ireton et al., 1993
EcoNarL	S80P,L	low-level constitutive	80	Egan & Stewart, 1991
RmeFixJ	P77S	decreased activity	82	Weinstein et al., 1992
RmeFixJ	T82I	decreased activity; unphosphorylatable	87	Weinstein et al., 1992
EcoCheY	T87I	loss of function; phosphorylatable	87	Matsumura et al., unpublished work
EcoOmpR	T83A	$OmpF^-$ $OmpC^+$; suppressor of D55Q mutant	87	Brissette et al., 1991b
EcoCheY	A88T	loss of function; phosphorylatable	88	Matsumura et al., unpublished work
BsuSpo0A	A87V	change of function	88	Olmedo et al., 1990
EcoNarL	V88A	constitutive	88	Egan & Stewart, 1991
BsuSpo0A	Q90K	change of function	91	Olmedo et al., 1990
BsuSpo0A	D92Y	sporulation-proficient suppressor (*sof*)	93	Spiegelman et al., 1990
EcoCheB	E91K	phosphorylation independent; increased activity	93	Stewart, 1993
EcoOmpR	G94S	$OmpF^+$ $OmpC^+$; suppressor of D55Q	98	Brissette et al., 1991b
EcoOmpR	G94D	$OmpF^-$ $OmpC^-$ loss of function	98	Brissette et al., 1992
EcoOmpR	E96A	$OmpF^-$ $OmpC^-$; phosphorylation normal; loss of function	100	Nakashima et al., 1991
BsuDegU	T98I	increased phosphorylation; hyperactivity	101	Dahl et al., 1991
BsuDegU	T98I	pleotropic; hyperactivity	101	Msadek et al., 1990
EcoCheB	L99P	inactive; nonphosphorylatable	101	Stewart et al., 1990
BsuSpo0A	A102P	suppressor of Spo0E deletion (*sed*)	103	Perego & Hoch, 1991
BsuSpo0A	F105S	sporulation-proficient suppressor (*sof*)	106	Spiegelman et al., 1990
EcoOmpR	Y102C	constitutively active	106	Kanamaru & Mizuno, 1992
BsuDegU	E107K	pleotropic; constitutive activity	110	Henner et al., 1988
EcoNarL	E112G	low-level constitutive	112	Egan & Stewart, 1991
AtuVirG	I106T	(double mutant with V7I) superactivator	113	Chen et al., 1991
EcoOmpR	E111K	$OmpF^-$ $OmpC^-$; loss of function	115	Brissette et al., 1992
EcoOmpR	R115S	$OmpF^-$ $OmpC^-$; phosphorylation normal; loss of function	119	Nakashima et al., 1991
BsuSpo0A	Q121R	sporulation-proficient suppressor (*sof*)	122	Spiegelman et al., 1990
CheY Suppressor Mutations				
StyCheY		L9W, L24X, K45N, L66Q, A99T, S104R, P110T, A113V, T115A		Sockett et al., 1992
EcoCheY		V11M, E27K, S56F, A90T, A90V, V108M, F111V, T112I, E117K		Roman et al., 1992

[a] Mutational results are only briefly described. See original references for full details. [b] Aiba, H., Nakasai, F., Mizushima, S., & Mizuno, T. (1989) *J. Biol. Chem. 264*, 14090–14094. Brissette, R. E., Tsung, K., & Inouye, M. (1991a) *J. Bacteriol. 173*, 601–608. Brissette, R. E., Tsung, K., & Inouye, M. (1991b) *J. Bacteriol. 173*, 3749–3755. Brissette, R. E., Tsung, K., & Inouye, M. (1992) *J. Bacteriol. 174*, 4907–4912. Chen, C.-Y., Wang, L., & Winans, S. C. (1991) *Mol. Gen. Genet. 230*, 302–309. Dahl, M. K., Msadek, T., Kunst, F., & Rapaport, G. (1991) *J. Bacteriol. 173*, 2539–2547. Dahl, M. K., Msadek, T., Kunst, F., & Rapaport, G. (1992) *J. Biol. Chem. 267*, 14509–14514. Egan, S. M., & Stewart, V. (1991) *J. Bacteriol. 173*, 4424–4432. Green, B. D., Olmedo, G., & Youngman, P. (1991) *Res. Microbiol. 142*, 825–830. Han, D. C., Chen, C.-Y., Chen, Y-F., & Winans, S. C. (1992) *J. Bacteriol. 174*, 7040–7043. Henner, D. J., Yang, M., & Ferrari, E. (1988) *J. Bacteriol. 170*, 5102–5109. Hoch, J. A., Trach, K.,

Table III (Continued)

Kawamura, F., & Saito, H. (1985) *J. Bacteriol. 161*, 552–555. Ireton, K., Rudner, D. Z., Siranosian, K. J., & Grossman, A. D. (1993) *Genes Dev. 7*, 283–294. Jin, S., Song, Y.-N., Pan, S. Q., & Nester, E. W. (1993) *Mol. Microbiol. 7*, 555–562. Kanamaru, K., & Mizuno, T. (1992) *J. Biochem. (Tokyo) 111*, 425–430. Matsuyama, S.-I., Mizuno, T., & Mizushima, S. (1986) *J. Bacteriol. 168*, 1309–1314. Msadek, T., Kunst, F., Henner, D., Klier, A., Rapaport, G., & Dedonder, R. (1990) *J. Bacteriol. 172*, 824–834. Nakashima, K., Kanamaru, K., Aiba, H., & Mizuno, T. (1991) *J. Biol. Chem. 266*, 10775–10780. Nara, F., Matsuyama, S.-I., Mizuno, T., & Mizushima, S. (1986) *Mol. Gen. Genet. 202*, 194–199. Olmedo, G., Ninfa, E. G., Stock, J., & Youngman, P. (1990) *J. Mol. Biol. 215*, 359–372. Pazour, G. J., Ta, C. N., & Das, A. (1991) *Proc. Natl. Acad. Sci, U.S.A. 88*, 6941–6945. Perego, M., & Hoch, J. A. (1991) *J. Bacteriol. 173*, 2514–2520. Roman, S. J., Meyers, M., Volz, K., & Matsumura, P. (1992) *J. Bacteriol. 174*, 6247–6255. Sockett, H., Yamaguchi, S., Kihara, M., Irikura, V. M., & Macnab, R. M. (1992) *J. Bacteriol. 174*, 793–806. Spiegelman, G., van Hoy, B., Perego, M., Day, J., Trach, K., & Hoch, J. A. (1990) *J. Bacteriol. 172*, 5011–5019. Stewart, R. C., Roth, A. F., & Dahlquist, F. W. (1990) *J. Bacteriol. 172*, 3388–3399. Stewart, R. C. (1993) *J. Biol. Chem. 268*, 1921–1930. Weinstein, M., Lois, A. F., Monson, E. K., Ditta, G. S., & Helinski, D. (1992) *Mol. Microbiol. 6*, 2041–2049. Yamada, M., Makino, K., Amemura, M., Shinagawa, H., & Nakata, A. (1989) *J. Bacteriol. 171*, 5601–5606.

internally, forming hydrogen bonds with the backbone carbonyl oxygens of S79, L81, and G102. Conservation of a basic amino acid in this position (R + K, 66%) is probably for retention of these strongly stabilizing interactions. Two more obvious examples of localized, high conservation are the helix-terminating glycines with α_L configurations: G29 past the end of α1, 62% identity, and G102 near the end of α4, 86% identity (the third α_L glycine—G128—is unique to CheY). The only puzzling conserved site is the alanine at position 103 (80%). This site's conservation might be explained by the substitutional covariance it shares with position 106. Residue 106 is predominantly an aromatic (Y/F, 53/31%), and the side chains for residues 103 and 106 are in van der Waals contact. Substitution by a smaller nonaromatic residue at 106 correlates positively with increased bulk at position 103. Possible functional significance of this covariance is currently being investigated (Zhu, Volz, and Matsumura, unpublished work).

An interesting conservative trend seen for the interior β strands is the presence of either a charged group or proline residue preceding the hydrophobic tetrad (see positions 7, 52, and 82 preceding β1, β3, and β4, respectively; Figure 2). For example, position 52, although a glycine in CheY, is by consensus an aspartate (69%). This type of conservation is apparent in the multiple sequence alignment of other β/α protein families, as was discussed in a survey of β breakers in parallel β sheets (Colloc'h & Cohen, 1991).

One last pattern worth noting is the commonly observed polarized distribution of charged amino acids along the course of the α helices (Ptitsyn, 1969). The best examples are α3 and α5 (most pronounced in the OmpR family), where acidic residues cluster at the amino termini and basic residues dominate the carboxy termini.

Mutant Sites in Regulatory Domains. Many mutant forms of two-component response regulators have been characterized phenotypically and, in many cases, by *in vitro* biochemical assay. It is instructive to view the results of these experiments collectively, in terms of the distribution of the point mutations on a generalized response regulator structure. Table III briefly summarizes the positions and consequences of all spontaneous response regulator mutations known to date (site-specific mutagenesis experiments were excluded). With few exceptions, the mutant sites reside on the surface of the molecule. There is an obvious clustering of mutations about the functionally important region of the response regulator domain; most are located near the phosphorylation site of the CheY molecule. Other mutations cluster about the two loops following the β4 and β5 strands, forming a spatially contiguous surface adjacent to the active site. In two separate mutational suppression experiments, this same region has been tentatively mapped out as the effector surface of the CheY molecule (Sockett et al., 1992; Roman et al., 1992). This proposed signaling surface does not completely overlap with the phosphorylation region of the molecule.

CONCLUSIONS AND PERSPECTIVES

The Conserved Template. The scaffold for the CheY superfamily is built upon a singular, highly conserved domain. This detailed interpretation of the molecular architecture of CheY provides well-founded explanations for the conservation within the aligned sequences. It is inevitable that the overall secondary structure and even many of the specialized structural features of CheY will be present in all homologues. These conclusions strongly argue for a common molecular mechanism of phosphorylation-dependent activation for all *bona fide* members of the superfamily.

Mg^{2+} Binding Site and Possible Acyl Phosphate Locations. The phosphoryl group transfer reactions of CheY require Mg^{2+} ion (Lukat et al., 1990). The Mg^{2+} binding site on the CheY molecule is close to the top of β1 and β3, in the same location as the putative NH_4^+ ion in the wild-type *E. coli* structure. The coordination sphere of the Mg^{2+} ion includes the side chains of the highly conserved residues D12 and D13, the carbonyl oxygen of residue 59, solvent molecules, and apparently the carboxyl group of D57 itself (Stock et al., 1992).

Because of the extremely short half-life of the acyl phosphate moiety on CheY ($t_{1/2} \sim 10$ s; Hess et al., 1988), a structural determination of phosphorylated CheY is not possible with conventional X-ray diffraction methods. However, model building experiments with the wild-type *E. coli* CheY structure show two stereochemically reasonable, but slightly different positions for the acyl phosphate group in the active site. One hypothetical position (Volz & Matsumura, 1991) brings the oxygen atoms of the phosphoryl group into reasonable bonding contact with the Mg^{2+} ion site. This position gives a low-energy conformation for the acyl phosphate group on D57 but requires a minor reorientation of the D12 carboxyl side chain. The second and possibly more favorable arrangement would place the phosphoryl group in an extended conformation, fully *trans* to the C^β atom of D57. In this orientation, the phosphoryl group is directed toward β4 and brings the acyl phosphate closer to residue 87—the site of high conservation for a hydroxyamino acid. Each of the above modeling approaches has its disadvantages; it is not possible to model the acyl phosphate ligand so it simultaneously interacts with both the Mg^{2+} site and the T87 side chain. Perhaps the two opposing sites function not only separately but also sequentially, playing different roles in the phosphorylation/dephosphorylation reactions. Evidence from phosphorylation-defective mutants (see below) appears to support this dual role possibility.

Possible Activation Mechanism for CheY. Phosphorylation is the required, primary event in the activation of normally functioning response regulators, but phosphorylation and activation can be unlinked. The two most interesting types of phosphorylation mutants are (1) phosphorylatable, but not active, and (2) nonphosphorylatable, but constitutively active.

Recent biochemical and structural results have been obtained for both types of mutants in CheY.

One constitutively active mutant of CheY bears an aspartate to lysine substitution at position 13 (Bourret et al., 1990). Discovery of this mutant raised the exciting prospect that its structure might represent the activated conformation of the CheY molecule. Suprisingly, the refined 2.3-Å resolution structure of the D13K mutant reveals no significant differences in backbone conformation when compared to wild-type CheY (Jiang, Bourret, Simon, and Volz, unpublished work). Assuming that the wild-type CheY structure is the inactive form, these results might suggest that no conformational change accompanies the activation process of CheY. But a closer inspection shows that the active site bonding pattern in D13K is extensively rearranged, resulting in a significant overall weakening of the hydrogen bond interactions. Thus it is plausible that the D13K crystal structure represents an inactive but metastable state of the molecule, where its transformation to the activated form would proceed dynamically, in solution. This hypothesis should be verifiable with NMR spectroscopy (Drake et al., 1993; Bourret et al., 1993).

The other window into the activation mechanism of CheY is through structural analysis of phosphorylatable but inactive mutants. The two known CheY mutants of that type are K109R (Lukat et al., 1991) and T87I (Matsumura et al., unpublished work). The T87I mutant structure has recently been solved and refined at 2.1-Å resolution (Ganguli, Matsumura, and Volz, unpublished work). The active site region of the T87I structure is very similar to that of wild type, as would be expected, but the molecule does show significant conformational differences in the substituted region, distal from the active site. This location partially overlaps the putative signaling surface of the CheY molecule inferred by conformational suppression analysis (Sockett et al., 1992; Roman et al., 1992).

The activation mechanism of CheY is still poorly understood. Many more structural and functional issues must be addressed. For instance, what are the structural consequences of Mg^{2+} binding? What are the precise roles of K109 and T87 in the phosphorylation, activation, and dephosphorylation reactions of CheY? How will the transiently activated form of CheY be structurally determined? Answers to most of these questions hopefully will emerge from the combined results of ongoing genetic, biochemical, and biophysical experiments.

Two-Component Systems as Antimicrobial Targets. The fact that two-component regulatory systems are essential for bacterial survival and virulence makes them attractive targets for antimicrobial drug design (DiRita & Mekalanos, 1989; Miller et al., 1989; Deretic et al., 1991). Indeed, two inhibitors of one signal transduction system have already been isolated (Roychoudhury et al., 1993). These studies demonstrate the feasibility of selective inhibition of a bacterial signal transduction pathway and introduce two-component systems as a new class of biochemical targets for eventual development of therapeutically effective antimicrobial compounds. By consideration of the widespread occurrence of bacterial two-component systems, an extension of this approach—based on the molecular commonalities of the systems—may offer distinct advantages. This would be a multitarget approach, aimed at the dozens of homologous two-component systems found in any one bacterial species. If a single compound could be designed to simultaneously inhibit as many kinase/response regulator pairs as possible, the effects might be cumulative. In other words, maximum inhibition of one critical pathway would not be the objective; even low-level inhibition of many mechanistically similar pathways could cause systemic confusion within the organism. Microbial specificity seems assured, since no homologues of two-component systems have yet been found in vertebrate DNA sequences. Such a multitarget approach would also not be as susceptible to the common chromosomal bacterial resistance mechanisms arising from single point mutations.

ACKNOWLEDGMENTS

I thank J. S. Parkinson and E. Kofoid for advice and critical reading of the manuscript; R. Bourret for comments, corrections, and constructive criticism; R. Barnett for discussions concerning the active site solvent structure; A. Sorokin and colleagues for use of the BsuORFX17 sequence before publication; and the Scientific Computer Workstation and Biostatistics Facility of the Research Resources Center, University of Illinois at Chicago, for computer time and access to data banks and programs.

REFERENCES

Albright, L. M., Huala, E., & Ausubel, F. M. (1989) *Annu. Rev. Genet. 23*, 311–336.
Anderson, C. (1993) *Science 259*, 1685.
Bairoch, A., & Boekmann, B. (1991) *Nucleic Acids Res. 19*, 2247–2249.
Bourret, R. B., Hess, J. F., & Simon, M. I. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 41–45.
Bourret, R. B., Borkovich, K. A., & Simon, M. I. (1991) *Annu. Rev. Biochem. 60*, 401–441.
Bourret, R. B., Drake, S. K., Chervitz, S. A., Simon, M. I., & Falke, J. J. (1993) *J. Biol. Chem. 268*, 13081–13088.
Burks, C., Cassidy, M., Cinkosky, M. J., Cumella, K. E., Gilna, P., Hayden, J. E.-D., Keen, G. M., Kelley, T. A., Kelly, M., Kristofferson, D., & Ryals, J. (1991) *Nucleic Acids Res. 19*, 2221–2225.
Colloc'h, N., & Cohen, F. E. (1991) *J. Mol. Biol. 221*, 603–613.
Deretic, V., Mohr, C. D., & Martin, D. W. (1991) *Mol. Microbiol. 5*, 1577–1583.
DiRita, V. J., & Mekalanos, J. J. (1989) *Annu. Rev. Genet. 23*, 455–482.
Drake, S. K., Bourret, R. B., Luck, L. A., Simon, M. I., & Falke, J. J. (1993) *J. Biol. Chem. 268*, 13089–13096.
Feng, D.-F., & Doolittle, R. F. (1987) *J. Mol. Evol. 25*, 351–360.
George, D. G., Barker, W. C., & Hunt, L. T. (1986) *Nucleic Acids Res. 14*, 11–15.
Hamm, G. H., & Cameron, G. N. (1986) *Nucleic Acids Res. 14*, 5–9.
Hess, J. F., Oosawa, K., Kaplan, N., & Simon, M. I. (1988) *Cell 53*, 79–87.
Keim, P., Heinrikson, R. L., & Fitch, W. M. (1981) *J. Mol. Biol. 151*, 179–197.
Kofoid, E. C., & Parkinson, J. S. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4981–4985.
Lifson, S., & Sander, C. (1979) *Nature 282*, 109–111.
Lukat, G. S., Stock, A. M., & Stock, J. B. (1990) *Biochemistry 29*, 5436–5442.
Miller, J. F., Mekalanos, J. J., & Falkow, S. (1989) *Science 243*, 916–922.
Ninfa, A. J. (1991) *Genet. Eng. 13*, 39–72.
Nixon, B. T., Ronson, C., & Ausubel, F. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7850–7854.
Ohta, N., Lane, T., Ninfa, E. G., Sommer, J., & Newton, A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 10297–10301.
Parkinson, J. S., & Kofoid, E. C. (1992) *Annu. Rev. Genet. 26*, 71–112.
Pearson, W. R. (1990) *Methods Enzymol. 183*, 63–98.
Perez-Casal, J., Caparon, M. G., & Scott, J. R. (1991) *J. Bacteriol. 173*, 2617–2624.
Ptitsyn, O. B. (1969) *J. Mol. Biol. 42*, 501–510.

Roman, S. J., Meyers, M., Volz, K., & Matsumura, P. (1992) *J. Bacteriol. 174*, 6247–6255.
Roychoudhury, S., Zeilinski, N. A., Ninfa, A. J., Allen, N. E., Jungheim, L. N., Nicas, T. I., & Chakrabarty, A. M. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 965–969.
Sanders, D. A., Gillece-Castro, B. L., Stock, A. M., Burlingame, A. L., & Koshland, D. E., Jr. (1980) *J. Biol. Chem. 264*, 21770–21778.
Sockett, H., Yamaguchi, S., Kihara, M., Irikura, V. M., & Macnab, R. M. (1992) *J. Bacteriol. 174*, 793–806.
Stock, A., Koshland, D. E., Jr., & Stock, J. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7989–7993.
Stock, A. M., Mottonen, J. M., Stock, J. B., & Schutt, C. E. (1989a) *Nature 337*, 745–749.
Stock, J. B., Ninfa, A. J., & Stock, A. M. (1989b) *Microbiol. Rev. 53*, 450–490.
Stock, J. B., Lukat, G. S., & Stock, A. M. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 109–136.
Stock, J. B., Surette, M. G., McLeary, W. R., & Stock, A. M. (1992) *J. Biol. Chem. 267*, 19753–19756.
Taha, M. K., Dupuy, B. D., Saurin, W., So, M., & Marchal, C. (1991) *Mol. Microbiol. 5*, 137–148.
Volz, K., & Matsumura, P. (1991) *J. Biol. Chem. 266*, 15511–15519.