

Structural Basis of Ligand Discrimination by Two Related RNA Aptamers Resolved by NMR Spectroscopy

Yinshan Yang, Michel Kochoyan,* Petra Burgstaller, Eric Westhof, Michael Famulok

In a previous study, an RNA aptamer for the specific recognition of arginine was evolved from a parent sequence that bound citrulline specifically. The two RNAs differ at only 3 positions out of 44. The solution structures of the two aptamers complexed to their cognate amino acids have now been determined by two-dimensional nuclear magnetic resonance spectroscopy. Both aptamers contain two asymmetrical internal loops that are not well ordered in the free RNA but that fold into a compact structure upon ligand binding. Those nucleotides common to both RNAs include a conserved cluster of purine residues, three of which form an uneven plane containing a G:G pair, and two other residues nearly perpendicular to that surface. Two of the three variant nucleotides are stacked on the cluster of purines and form a triple contact to the amino acid side chain, whereas the edge of the third variant nucleotide is capping the binding pocket.

Since the development of *in vitro* selection in 1990 (1), more than 50 different nucleic acid aptamers that specifically recognize molecules of low or high molecular weight have been reported (2). Although detailed structural characterizations by nuclear magnetic resonance (NMR) have been described for some biologically relevant RNA fragments (3–7), little structural information is available for RNA aptamers. We now report the conformations of a citrulline-binding RNA aptamer and of its related arginine-binding triple mutant, both complexed to their cognate amino acids. These two aptamers were obtained in a previous *in vitro* selection study designed to isolate RNA aptamers for the amino acid L-citrulline and then to evolve the resulting binding motif to recognize L-arginine (8). Each aptamer binds its cognate amino acid with a dissociation constant (K_d) in the range of 10^{-5} M without detectable affinity for the noncognate amino acid. On the basis of covariations among the selected sequences, it was proposed that the secondary structure consists of two stem regions flanking two internal bulge regions (Fig. 1). The secondary structure was further refined

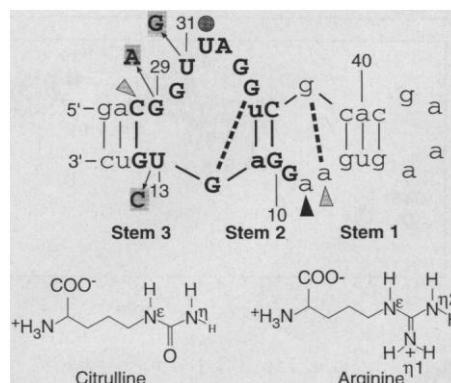
by chemical modification analysis (9). Modification of any nucleotide in the conserved region interferes with ligand binding, except for that of U32 (Fig. 1). On the basis of secondary structure refinements, shortened versions of the two aptamers were constructed (33 nucleotides) (Fig. 1). Specificity and selectivity were maintained (9). Lead cleavage experiments as well as changes in the imino proton spectra of these constructs observed by one-dimensional (1D) NMR spectroscopy indicated that substantial conformational changes take place in the RNA upon ligand binding. The lifetime of the complexes is longer than 10 ms, as shown by the occurrence of distinct spectra for the free and bound amino acids (9).

We have now determined the conformation of the two aptamer–amino acid complexes in solution by 2D NMR spectroscopy (10). Resonance assignments were obtained by standard 2D NMR techniques (11) and will be described in detail

elsewhere (12). The presence of the three stems delimiting the two bulged regions (Fig. 1) was confirmed; the stems adopt a standard A-form helix geometry. Starting from the apical GAAA tetraloop, aromatic and imino protons of the stems were sequentially and unambiguously assigned up to A7 and G38, which form an additional base pair continuing stem 1 and up to the noncanonical G12-G35 (see below) base pair closing stem 2 (Fig. 1). The nuclear Overhauser effect (NOE) pattern of aromatic and sugar protons between A11 and G14 demonstrates that nucleotides G12 and U/C13 stack continuously between stems 2 and 3. The two aptamers have similar NMR spectra, including similar NOE patterns of the three variant nucleotides (U/C13, G/A29, and U/G31) (Fig. 2). The different chemical shifts of the three variant nucleotides facilitate their assignments and, consecutively, those of the remaining residues. In particular, the G/A29 and U/G31 assignments eliminate the ambiguities in the assignments of the NOE pattern of aromatic and sugar protons in the large bulge between positions 29 and 35. The strong NOEs at short mixing time (80 ms) between H8 and H1' show that the orientation of the base is *syn* in G30 and G35.

Sequential imino proton assignments were straightforward from U36 to G12 and from G12 to U13 (Fig. 2). The amino protons of G12 are in slow rotational exchange and show a NOE connectivity to the H8 proton of G35. This observation indicates the formation of a *cis*-G12-G35 base pair involving the Watson-Crick sites of G12 and the Hoogsteen sites of G35. It also provides the assignment of the imino peak of G35, weakly NOE-connected to G12 and U36 (Fig. 2). The imino protons of U/G31, the imino proton of G29, and the H2 proton of A29 were assigned on the basis of the similarity of their arrays of

Fig. 1. Secondary structure proposed previously for the citrulline- and arginine-specific aptamers (8), based on covariations of selected sequences and on the chemical footprinting pattern obtained in the presence of the cognate amino acid as well as in damage-selection experiments (9). The sequences shown are those used in this study. The bases that were conserved among different isolates are shown in uppercase, and variant bases are in lowercase. The three nucleotides critical for arginine specificity (13, 29, and 31) are shaded. Filled triangles and circles indicate the sensitivity of A-N1 or C-N3 and U-N3 or G-N1 to chemical modification by dimethyl sulfate and 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulfonate (CMCT), respectively, in the presence of the amino acid (black triangles indicate a higher level of sensitivity than shaded ones). Thick lines indicate the noncanonical base pairs that extend stems 1 and 2. The sequences of the citrulline and arginine aptamers have been deposited with GenBank under the accession numbers U37123 (citrulline) and U37124 (arginine).



Y. Yang, Centre de Biochimie Structurale (CBS), Unité Mixte de Recherche CNRS 9955, INSERM U414, Faculté de Pharmacie, 15 Avenue Ch. Flahault, 34060 Montpellier, France.

M. Kochoyan, CBS Faculté de Pharmacie, 15 Avenue Ch. Flahault, 34060 Montpellier, and Groupe de Biophysique, Unité de Recherche Associée CNRS D1254, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau, France.

P. Burgstaller and M. Famulok, Institut für Biochemie der Ludwig Maximilians Universität München–Genzentrum, Würmtalstraße 221, 81375 München, Germany.

E. Westhof, Institut de Biologie Moléculaire et Cellulaire, Unité Propre de Recherche CNRS 9002, 15 Rue R. Descartes, 67084 Strasbourg Cedex, France.

*To whom correspondence should be addressed. E-mail: michel@tome.cbs.univ-montp1.fr

NOE connectivities in the citrulline and arginine aptamer spectra. Two G imino peaks were unassigned initially with three candidates (G9, G30, and G34). The ambiguity was solved by the first stage of modeling, which excludes G34 and allowed the assignment of G9 and G30 (13). The only unassigned imino protons are those of G34 and U32, which do not give any observable NOE, probably as a result of fast exchange with the solvent (14).

The 3D structures of the citrulline and arginine aptamers complexed to their cognate amino acid were obtained by restrained molecular dynamic calculations with the X-PLOR package (15–17). As expected, the structures of both aptamers are quite similar and contain the same scaffold of noncanonical interactions (Fig. 3). At the end of the large bulge facing stem 2, the G12-G35 base pair participates in an unusual array of hydrogen bonds (18) with G9, G30, and A33. G9 comes close to the G12-G35 base pair and is tilted with respect to the base pair plane. G30 is at roughly 90° from the G12-G35 base pair plane, with hydrogen bonds between its imino H1 and the N7 nitrogen of G9 and between its amino protons and the 2' oxygen of the ribose of residue 29.

Residue 32 is extruded from this scaffold, whereas A33 stacks with G30 and closes the “quintet” structure by forming hydrogen bonds between its N3 atom, its 2' oxygen, and the amino group of G35. The “quintet” structure (Fig. 4) forms the floor of the binding cleft that carries the functional groups provided by the three variant residues (U/C13, G/A29, and U/G31) that surround the amino acid side chain. The imino proton of U13 forms a hydrogen bond with the carboxyl oxygen in the citrulline urea group. In the arginine aptamer, this contact is substituted by N3 of C13 hydrogen bonding to one amino proton of the guanido group. In both complexes, H^ε of the amino acid may form a water-mediated hydrogen bond with the carboxyl oxygen O2 of pyrimidine 13. The N7 and O6 atoms of the invariant G30 residue form hydrogen bonds to the N^η protons of citrulline; in the case of arginine, these bonds are likely directed toward the N^{η1} amino group. U/G31 caps the amino acid: The Hoogsteen edge of G31 contacts both amino groups of arginine; correspondingly, O4 of U31 binds to the single amino group of citrulline. The third variable residue (G/A29) stacks on and stabilizes the variable pyrimidine 13

that interacts with the amino acid. In the citrulline aptamer the amino group of residue G29 probably also contacts the O^η oxygen of the amino acid by way of a hydrogen bond.

The aliphatic chain of both ligands stacks over the G12-G35 base pair, thereby contributing to the overall stability of the complex by hydrophobic interactions. Similar interactions have been identified, for example, in the crystal structure of the *trp* repressor-operator complex where the methylene groups of Arg⁵⁴ stack onto the indole ring of the corepressor ligand tryptophan (19). Other hydrophobic interactions have been observed in protein-RNA complexes (20, 21); thus, BIV-TAR RNA forms a hydrophobic binding pocket for a critical isoleucine residue of BIV TAT protein (7, 20, 22). The α-carboxyl and α-amino groups of the ligand are accessible to solvent. The amino group is hydrogen bonded to the sugar O4' of residue 33, and the carboxyl moiety is partly stacked under the variable nucleotide G/U31.

At the present level of structural refinement, the main difference in the RNA scaffold of the citrulline and arginine aptamers lies in the position of bases G9-G30 relative to the G12-G35 base pair. They are further

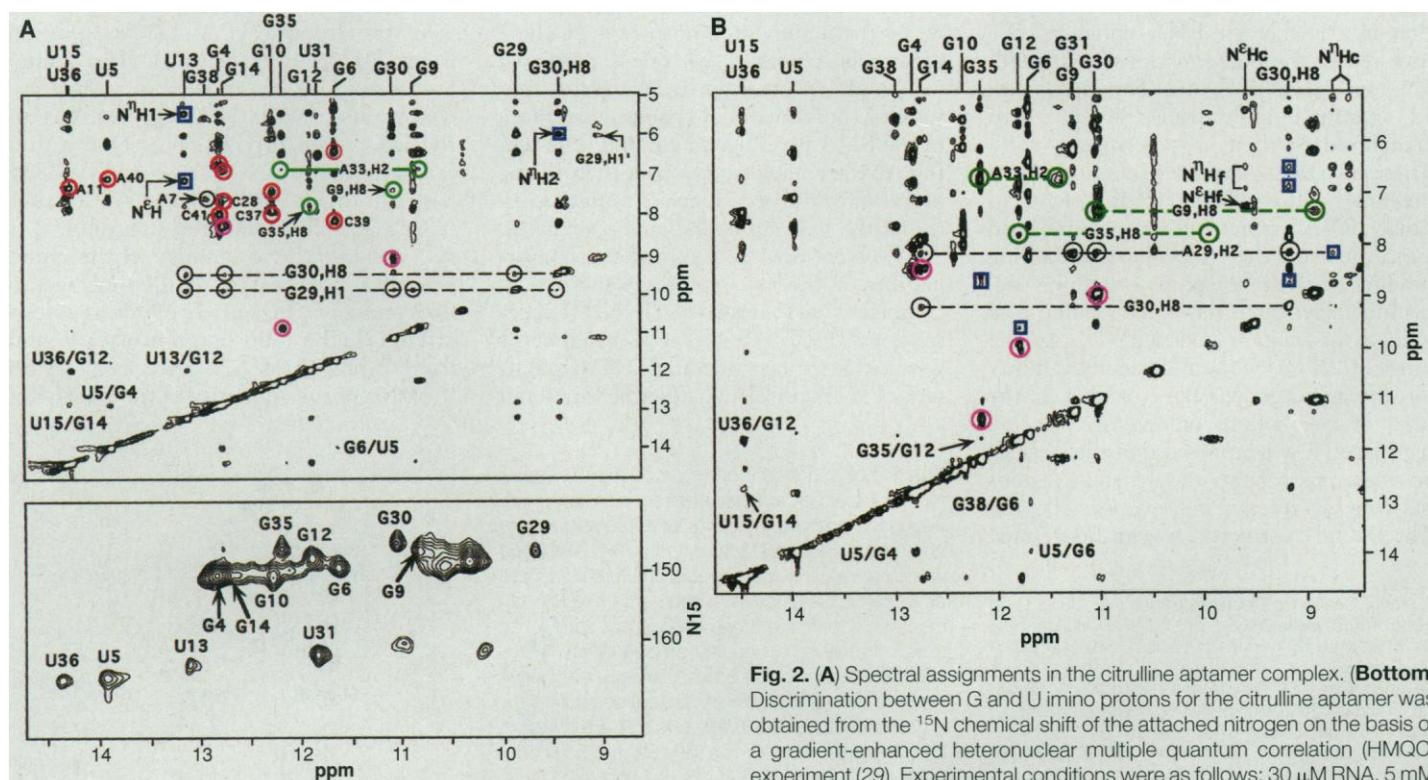


Fig. 2. (A) Spectral assignments in the citrulline aptamer complex. **(Bottom)** Discrimination between G and U imino protons for the citrulline aptamer was obtained from the ¹⁵N chemical shift of the attached nitrogen on the basis of a gradient-enhanced heteronuclear multiple quantum correlation (HMQC) experiment (29). Experimental conditions were as follows: 30 μM RNA, 5 mM citrulline, 20 mM NaCl (pH 6.7), 13°C. **(Top)** NOESY experiment (130-ms mixing time) in H₂O at 13°C. The cross-peaks (G imino to C amino protons and U imino to H2 protons) characteristic of canonical GC or AU base pairs are circled in red. The cross-peaks between G imino protons and their slowly exchanging, downfield-shifted amino protons are circled in magenta. Cross-peaks characterizing the interactions of purines 12, 30, 33, and 35 are circled in green. The cross-peaks between RNA and the amino acid are indicated by blue squares. **(B)** NOESY spectrum (130-ms mixing time) of the arginine aptamer at 5°C [colors as in (A)]. Because of exchange, some cross-peaks are observed to both the complexed (c) and free (f) forms of the amino acid. Some of the guanine amino proton resonances are broader at 13°C than at 5°C (as a result of temperature-enhanced rotational exchange) and are weaker or absent in the citrulline aptamer NOESY recorded at 13°C (but present at 5°C).

away in the arginine aptamer (average distance for the ensemble of conformers between H1 of G9 and N7 of G12 = 3.3 Å)

than in the citrulline aptamer (average distance = 2.4 Å, that is, within hydrogen bonding distance). The disposition of these

four G nucleotides results in close proximity of their four O6 oxygens (23).

A notable feature of the arginine aptamer is that it contacts the positively charged guanido group largely, if not exclusively, by hydrogen bonds without the formation of salt bridges; the RNA folds around the amino acid, which becomes an integral part of the core of the overall structure. The resulting arrangement of the bases appears quite rational considering the function for which the RNA has been selected, but differs markedly from the binding modes observed in protein–nucleic acid complexes, in which arginine is often found as a key residue (3–8, 20, 24, 25). In all known structures, the guanido group of arginine interacts by hydrogen bonding with the phosphodiester backbone of the nucleic acid, whereas sequence-specific interactions occur at the Hoogsteen edge of G nucleotides. In protein–DNA complexes, these interactions do not alter the base-pairing scheme, whereas in the case of the arginine amide HIV-TAR (5) and the BIV-TAT RNA binding peptide (6, 22), NMR experiments demonstrated modifications of the base-pairing scheme of RNA upon ligand binding. Even in those cases in which arginine is the most critical determinant for specific interactions (5, 26), the binding site is located at the edge of the RNA deep

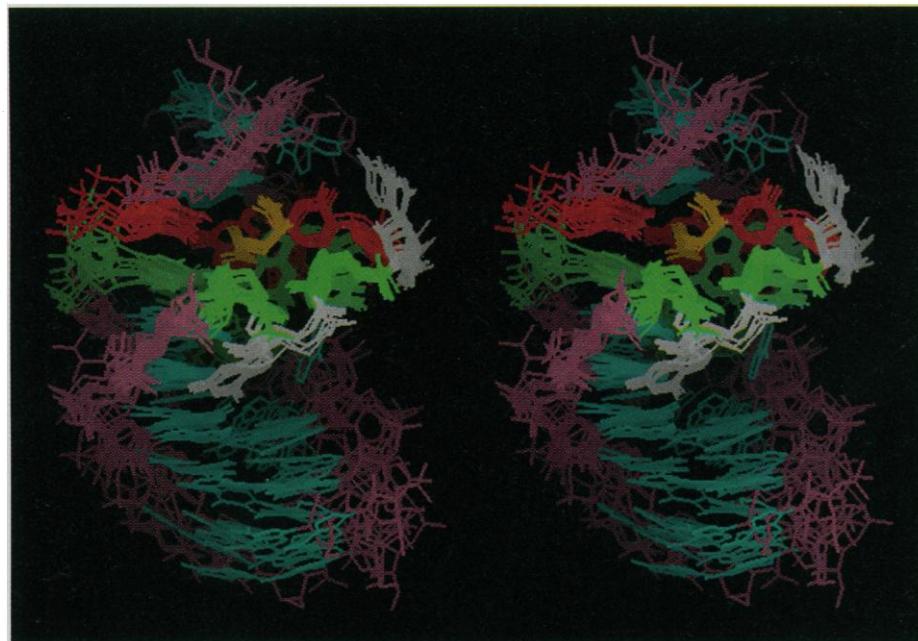


Fig. 3. Ensemble of 10 structures generated for the citrulline aptamer (16, 17). The amino acid is shown in yellow; residues 9, 12, 30, 33, and 35 are in green; variant residues 13, 29, and 31 are in red; residues 32 and 34 are in white; and others are in blue (bases) and pink (backbone). The average rms deviation between pairs of conformers for nucleotides G9 to G14 and C29 to C39 is 0.63 Å for the bases and 1.2 Å for the backbone. A comparable precision was obtained for the arginine aptamer.

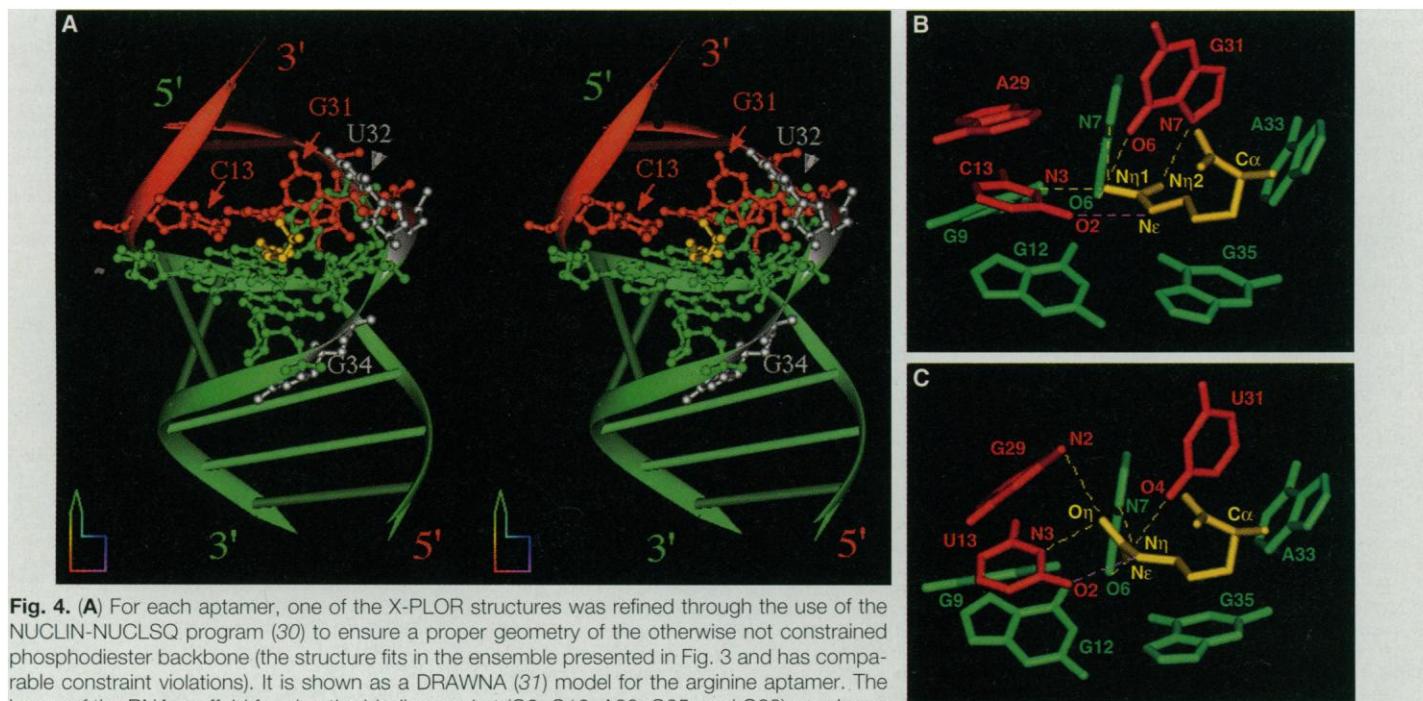


Fig. 4. (A) For each aptamer, one of the X-PLOR structures was refined through the use of the NUCLIN-NUCLSQ program (30) to ensure a proper geometry of the otherwise not constrained phosphodiester backbone (the structure fits in the ensemble presented in Fig. 3 and has comparable constraint violations). It is shown as a DRAWNA (31) model for the arginine aptamer. The bases of the RNA scaffold forming the binding pocket (G9, G12, A33, G35, and G30) are shown in green, the three variant nucleotides (C13, A29, and G31) in red, the amino acid in yellow, and other nucleotides of the consensus sequence in gray. Residue 34 points toward base pair G10–C37 with its O6 oxygen hydrogen bonded to the external amino proton of C37 (see the downfield shift of this proton in Fig. 2), which probably explains why its imino is protected from chemical modification. The phosphodiester backbone along the large bulge spanning from residues 29 to 35 is wide open but maintains a right-handed helical conformation. The nucleotides along this bulge are arranged so as to surround the amino acid. (B) Enlargement of the binding pocket. The nucleotides are colored as in (A). Hydrogen bonds are indicated by yellow dashed lines. For the ensemble of conformers, the average distance between H^ε of citrulline and O2 of U13 is 2.8 Å. This distance suggests the existence of a water-mediated rather than a direct hydrogen bond, and a different color is therefore used. (C) Same as in (B), but for the citrulline aptamer. The average distance between H^ε and O2 is now 2.5 Å.

groove and the amino acid contacts both the phosphate backbone and the Hoogsteen site of a G residue.

Our findings demonstrate that small RNAs are able to fold into compact structures, thereby bringing into close spatial proximity various functional groups belonging to residues scattered along the primary structure of the macromolecule. They reveal the capacity of RNA for building well-defined binding pockets by organizing complex arrays of nucleotides and exploiting the diversity of their polar atoms. Variability in specificity is achieved by mutations of a subset of residues leaving invariant those that mould the recognition scaffold as in protein enzymes. Whether these principles of molecular recognition are or have been relevant in biologic or prebiotic processes remains to be investigated.

REFERENCES AND NOTES

1. D. L. Robertson and G. F. Joyce, *Nature* **344**, 467 (1990); C. Tuerk and L. Gold, *Science* **249**, 505 (1990); A. D. Ellington and J. W. Szostak, *Nature* **346**, 818 (1990).
2. J. W. Szostak, *Trends Biochem. Sci.* **17**, 89 (1992); G. F. Joyce, *Curr. Opin. Struct. Biol.* **4**, 331 (1994); S. J. Klug and M. Famulok, *Mol. Biol. Rep.* **20**, 97 (1994); L. Gold, *J. Biol. Chem.* **270**, 13581 (1995); B. Polisky, O. Uhlenbeck, M. Yarus, *Annu. Rev. Biochem.* **64**, 763 (1995); A. D. Ellington, *Curr. Biol.* **4**, 427 (1994); I. Hirao and A. D. Ellington, *ibid.* **5**, 1017 (1995).
3. F. H. Allain and G. Varani, *J. Mol. Biol.* **250**, 333 (1995); C. Cheong, G. Varani G, I. Tinoco Jr., *Nature* **346**, 680 (1990); G. Varani, C. Cheong, I. Tinoco Jr., *Biochemistry* **30**, 3280 (1991); J. Santa Lucia Jr. and D. H. Turner, *Biochemistry* **32**, 12612 (1993); H. A. Heus and A. Pardi, *Science* **253**, 191 (1991); E. V. Puglisi, J. D. Puglisi, J. R. Williamson, U. L. RajBhandary, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11467 (1994).
4. B. Wimberly, G. Varani, I. Tinoco Jr., *Biochemistry* **32**, 1078 (1993); G. Varani, B. Wimberly, I. Tinoco Jr., *ibid.* **28**, 7760 (1989); J. D. Puglisi, J. R. Wyatt, I. Tinoco Jr., *J. Mol. Biol.* **214**, 437 (1990); *Biochemistry* **29**, 4215 (1990).
5. J. D. Puglisi, R. Tan, B. J. Calnan, A. D. Frankel, J. R. Williamson, *Science* **257**, 76 (1992).
6. J. D. Puglisi, L. Chen, S. Blanchard, A. D. Frankel, *ibid.* **270**, 1200 (1995).
7. R. D. Peterson, D. P. Bartel, J. W. Szostak, S. J. Horvath, J. Feigon, *Biochemistry* **33**, 5357 (1994).
8. M. Famulok, *J. Am. Chem. Soc.* **116**, 1698 (1994).
9. P. Burgstaller, M. Kochoyan, M. Famulok, *Nucleic Acids Res.* **23**, 4769 (1995).
10. Both aptamers were obtained by T7-RNA polymerase transcription of synthetic DNA templates in milligram quantities. The aptamers were purified as described [J. R. Wyatt, M. Chastain, J. D. Puglisi, *Bio-techniques* **11**, 764 (1991)]. ¹⁵N labeling of the RNA aptamers was carried out as described [E. P. Nikonowicz *et al.*, *Nucleic Acids Res.* **20**, 4508 (1992)]. NMR samples consisted of 1 to 2 mM RNA, twice as much of amino acid, and 60 mM NaCl (pH 6.7), unless otherwise specified. Spectra were acquired on the 600-MHz Bruker spectrometer at the Centre de Biochimie Structurale. Two-dimensional spectra in D₂O (NOESY, nuclear Overhauser exchange spectroscopy; DQ-COSY, double quantum filtered spectroscopy; and TOCSY, total correlation spectroscopy) were acquired at 5°, 13°, and 20°C in order to remove ambiguities resulting from overlapping cross-peaks. NOESY spectra in H₂O were performed with a jump and return (27) reading pulse at 5° and 13°C. Assignments were obtained up to the 3' sugar protons for most residues. Resonance assignments of the bound amino acid were obtained from the exchange cross-peaks with the free amino acid in the NOESY spectra and confirmed by TOCSY experiments (12).
11. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
12. Y. Yang and M. Kochoyan, unpublished results.
13. The two unassigned G imino resonances show a NOE to the independently assigned G29-imino or A29-H2 proton. In the preliminary structures generated without these assignments, the G34 imino proton was always located at >8 Å from the H2 or imino proton of nucleotide 29. Therefore, the two unassigned resonances belong to G9 and G30. One of them shows a NOE to the H8 proton of G9 and was consequently assigned to G30, because such a NOE cannot be observed within a nucleotide. The other resonance was assigned by default to G9.
14. For U32 this observation is further supported by the chemical probing data because it can be modified by CMCT in the presence of the bound amino acid, indicating that the imino proton is solvent accessible (Fig. 1A). For G34, see Fig. 4 legend.
15. A. T. Brünger, *X-PLOR User Manual, Version 3.1* (Yale University, New Haven, CT, 1992).
16. Only the portion going from base pair U5-A40 to base pair U15-A27 was modeled. The starting structures consisted of two independent strands in random conformation plus the amino acid. Distance restraints were obtained from NOESY spectra recorded with mixing times ranging from 80 to 150 ms. They were classified as strong (2.7 Å), medium (3.5 Å), weak (4.5 Å), and absent (4.5 Å, for a few sugar to amino acid protons) and calibrated with respect to the H5-H6 cross-peak of pyrimidines for nonexchangeable protons or to the imino to H2 cross-peaks of A-U base pairs for exchangeable protons (a cross-peak intensity correction was then performed to take into account the jump- and-return pulse (JR) frequency response). A few sugars were constrained as C2'-endo or C3'-endo on the basis of coupling data. Modeling was performed by means of a high-temperature simulated-annealing protocol (15-ps dynamics at 1200 K with reduced van der Waals radii) followed by 8-ps cooling to 300 K (15) without electrostatic, dihedral, or hydrogen bonding potential and with a repulsive van der Waals potential. These dynamics were followed by 300 steps of conjugate gradient minimization with attractive van der Waals potential. Twenty or more structures were generated, and only those with a constraint energy within a limit of 30% of the lowest constraint energy obtained for the best structure (usually more than half and at least 10) were retained. A first step of modeling was used to generate an ensemble of structures compatible with the restraints obtained by NMR. At this stage, no hydrogen bonding constraints were used except those defining the base pairs of the three duplex stems 1 to 3 (except pair G12-G35 of stem 2). New hydrogen bonds were added to the constraints list as short distance restraints (2.2 ± 0.3 Å) between a hydrogen and an acceptor atom after the following conditions were satisfied: (i) the acceptor and donor heavy atoms were within a distance of <4 Å in all the structures generated; (ii) the bonded protons were in slow exchange with the solvent (and in slow rotational exchange for the amino groups); and (iii) the structures generated in the presence of the new constraints did not result in higher violation of the other experimental constraints than those generated in its absence. As a result, six hydrogen bonding constraints were added for the citrulline aptamer, and only four of them were used for the arginine aptamer. The ones common to both aptamers are those defining the G12-G35 base pair (amino group of G12 to N7 of G35 and imino proton of G12 to O6 of G35, supported by the downfield shift of G12 imino proton, the slow rotational exchange of the amino group, and the NOEs to the H8 proton of G35), between the amino group of G35 and the N3 nitrogen of A33 [both amino protons are in slow rotational exchange, showing a strong NOE to H2 of A33 with one of them shifted downfield (Fig. 2)] and between the imino proton of G30 and the N7 nitrogen of G9 (supported by the imino to H8 NOE). For the citrulline aptamer, two extra bonding constraints were used between the imino proton of G29 and the O2 proton of U13 as well as between the imino proton of U13 and the carboxyl oxygen of the urea moiety of the amino acid. This latter bond is supported by the downfield shift of the U13 imino proton and by the strong NOEs observed to H^ε and H^η of citrulline.
17. Apart from the base-pairing constraints used along stems 1, 2, and 3 (see above), 174 NOEs and 8 dihedral angles for the citrulline aptamer and 161 NOEs and 8 dihedral angles for the arginine aptamer were used to generate the structures (containing residues G9 to G14 and C28 to C37). They consist of 25 intranucleotide, 78 sequential, and 55 long-range NOEs (26, 74, and 46, respectively, for the arginine aptamer) together with 16 intermolecular NOEs between the amino acid and the RNA. Apart from those indicated by blue squares (Fig. 2) and involving the imino protons of G12 and G35, the H8 proton of G30, the imino proton of U13 or the H2 proton of G29 and the guanido or urea group of the amino acid, other contacts are observed between the sugar protons of residues 31 and 33 and the aliphatic protons of the amino acid and between the H8 proton of residue 35 and the Hβ protons of the amino acid. Four repulsive constraints were used to take into account the absence of NOE cross-peaks between the aromatic H6 or H8 proton of variable residue 31 and the Hγ and δ protons of the ligand. The maximum (0.35 Å) and average (0.05 Å) distance violations are comparable for both aptamers. The average deviation from ideality for bond length, bond angle, and improper for the ensemble of structures is 0.007 Å (no violation >0.05 Å), 1.4°, and 0.7°, respectively (similar for both aptamers). The average root mean square (rms) deviations between the starting and the converged structures are >15 Å. The largest chemical-shift anomalies are observed for the H8 proton of G30 [9.65 and 9.4 parts per million (ppm), for both aptamers] and for the H5 proton of U31 of the citrulline aptamer (3.95 ppm). They can be explained qualitatively by the locations of these protons in the 3D structure. The H8 proton of G30 is in an unusual environment, surrounded by several aromatic rings and in close proximity to the N^η protons of the amino acid. The H5 proton of U32 is in axial contact with the six-membered ring of A33 in a position where ring current effects are expected to be at a maximum (28). The involvement of G9 in the overall stability of the structure is supported by the observation that a mutant RNA with a G9-C42 base pair is unable to bind its cognate amino acid (12). Furthermore, the localization of G9 in the 3D structure of the complexes implies partial unstacking and solvent exposure of the functional groups of adenine 8, and indeed that residue is more sensitive to chemical modification in the presence of the amino acid (9).
18. A hydrogen bond is assumed when the distance between the acceptor and donor heavy atoms is <3.5 Å in all X-PLOR conformers. The term probable hydrogen bond is used when only 8 out of 10 conformers satisfy this condition.
19. Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
20. L. Chen and A. D. Frankel, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5077 (1995).
21. H. Siomi, M. Choi, M. C. Siomi, R. L. Nussbaum, G. Dreyfuss, *Cell* **77**, 33 (1994).
22. X. Ye, R. A. Kumar, D. J. Patel, *Chem. Biol.* **2**, 827 (1995).
23. In the aptamer of citrulline, whose urea group is uncharged, the oxygen array might be stabilized by a sodium ion [as observed for G quartet in tetraplex structures (32)]. In the aptamer of arginine, the positively charged guanido group might provide some stabilization to the array, while hindering cation binding.
24. C. G. Burd and G. Dreyfuss, *Science* **265**, 615 (1994); K. Nagai and I. W. Mattaj, Eds., *RNA-Protein Interactions* (Oxford Univ. Press, Oxford, 1994).
25. C. Pabo and R. Sauer, *Annu. Rev. Biochem.* **61**, 1053 (1992).
26. J. Tao and A. D. Frankel, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2723 (1992).
27. P. Plateau and M. Guéron, *J. Am. Chem. Soc.* **104**, 7310 (1982).
28. C. Giessner Prettre, B. Pullman, P. Borer, L.-S. Kan, P. Ts'o, *Biopolymers* **15**, 2277 (1976).
29. V. Sklenar and A. Bax, *J. Magn. Reson.* **74**, 469 (1987); G. Kellogg and P. Moore, *FEBS Lett.* **327**, 261 (1993).
30. E. Westhof, P. Dumas, D. Moras, *J. Mol. Biol.*

- 184, 119 (1985).
 31. C. Massire, C. Gaspin, E. Westhof, *J. Mol. Graphics* **12**, 201 (1994).
 32. G. Laughlan *et al.*, *Science* **265**, 520 (1994).
 33. Supported by grants from the Deutsche For-

schungsgemeinschaft to M.F. and the European Union (grant Biot-CT93-0345) to M.K., E.W., and M.F. M.K. thanks M. Guéron and J. L. Leroy for critical reading of the manuscript. Coordinates have been deposited at the Protein Data Bank (accession

number 1KOC and 1KOD) and will be directly accessible on the Web server of the Centre de Biochimie Structurale at <http://tome.cbs.univ-montp1.fr>.

28 December 1995; accepted 25 March 1996

Stress-Induced Phosphorylation and Activation of the Transcription Factor CHOP (GADD153) by p38 MAP Kinase

XiaoZhong Wang and David Ron*

CHOP, a member of the C/EBP family of transcription factors, mediates effects of cellular stress on growth and differentiation. It accumulates under conditions of stress and undergoes inducible phosphorylation on two adjacent serine residues (78 and 81). In vitro, CHOP is phosphorylated on these residues by p38 mitogen-activated protein kinase (MAP kinase). A specific inhibitor of p38 MAP kinase, SB203580, abolished the stress-inducible in vivo phosphorylation of CHOP. Phosphorylation of CHOP on these residues enhanced its ability to function as a transcriptional activator and was also required for the full inhibitory effect of CHOP on adipose cell differentiation. CHOP thus serves as a link between a specific stress-activated protein kinase, p38, and cellular growth and differentiation.

CHOP, also known as growth arrest and DNA damage-inducible gene 153 (GADD153), is expressed in response to various metabolic stresses in all cells tested (1). By forming heterodimers with members of the C/EBP family of transcription factors, CHOP influences gene expression as both a dominant-negative regulator of C/EBP binding to one class of DNA targets (2) and by directing CHOP-C/EBP heterodimers to other sequences (3). Both modes of action are implicated in the effects of CHOP on cellular growth (4, 5) and differentiation (6). In addition to inducing CHOP expression, stress increases the ability of CHOP to activate gene expression (3). This latter observation prompted us to study the possible stress-induced posttranslational modification of CHOP.

In stressed cells, CHOP is present in two forms, distinguishable by their migration in SDS-polyacrylamide gel electrophoresis (Fig. 1A) (7). Isoelectric focusing revealed the form of CHOP with decreased mobility to have a more acidic isoelectric point, suggestive of phosphorylation. To study the possible effects of stress on CHOP phosphorylation independently of the stress-induced increase in protein, we used cells constitutively expressing an epitope-tagged form of CHOP that is distinguishable in size from the endogenous protein. In vivo labeling with [³²P]orthophosphate followed by immunoprecipitation revealed two to four

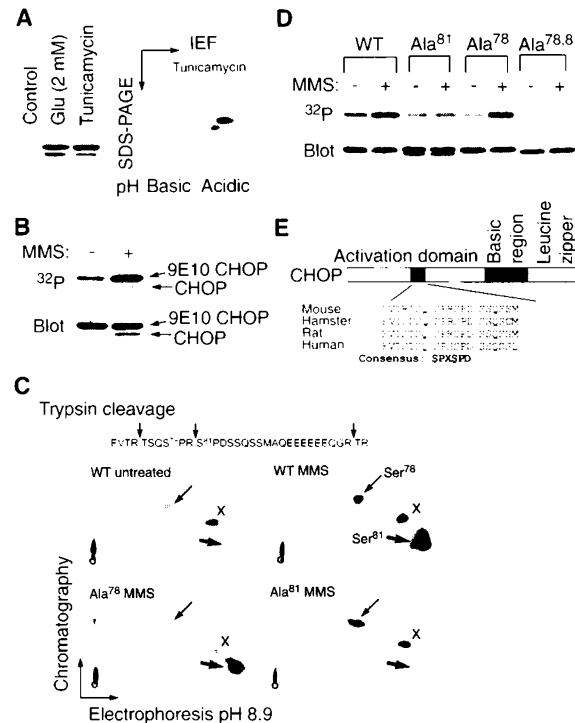
times as much phosphorylation of tagged CHOP in response to stress with no change in the amount of protein (Fig. 1B). The

stress-induced endogenous CHOP was also phosphorylated.

Comparison of the tryptic phosphopeptide maps of the tagged CHOP from stressed and unstressed cells showed most of the inducible phosphorylation to occur on two distinct peptides (Fig. 1C). All known CHOP proteins contain two adjacent serine residues [amino acids 78 and 81 in the mouse sequence (2)] in a context that may serve as a substrate for members of the MAP kinase family (Fig. 1E). Members of this family are activated by many of the same insults that induce CHOP expression (8) and may therefore participate in the stress-induced phosphorylation of CHOP. Conversion of Ser⁷⁸ or Ser⁸¹ to Ala led to the selective loss of inducible phosphorylation of one tryptic peptide (Fig. 1C). The minimal residual phosphorylation of the Ala-substituted peptides may have resulted from the presence of other phosphorylated residues. The mutant CHOP proteins exhibited

Fig. 1. Stress-induced phosphorylation of CHOP on Ser⁷⁸ and Ser⁸¹.

(A) Two isoforms of CHOP are present in stressed cells. NIH 3T3 cells were cultured in a medium with a low concentration of glucose [(Glu), 2 mM, 16 hours] or treated with tunicamycin (25 μg/ml, 4 hours). CHOP, detected by protein immunoblotting with the 9C8 monoclonal antibody (6), migrates as a doublet with 11% SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The form with less mobility is more acidic on isoelectric focusing (IEF). (B) Endogenous CHOP and Myc-epitope-tagged CHOP [9E10 CHOP (5)] from [³²P]orthophosphate-labeled NIH 3T3 cells (500 μCi/ml, 5 hours) that were treated with the stress-inducing alkylating agent MMS (100 μg/ml, 3 hours) were immunoprecipitated with 9C8. Autoradiography (top) and protein immunoblotting with rabbit antiserum to CHOP (bottom) are shown. (C) Wild-type (WT) and mutant CHOP, immunoprecipitated from treated or untreated NIH 3T3 cells with the antibody to 9E10 (17), were digested with trypsin and the phosphopeptides separated by electrophoresis and thin-layer chromatography and visualized by autoradiography (18). The circles mark the origin of the run. The arrows mark the position of the two peptides that underwent inducible phosphorylation upon MMS treatment. The peptide marked "X" was constitutively phosphorylated. The predicted sequence of the tryptic phosphopeptides, with small arrows denoting the tryptic cleavage sites, is shown above the autoradiograms (19). (D) Comparison of the in vivo phosphorylation of wild-type and Ala substitution mutants of CHOP from untreated cells and cells treated with MMS. Autoradiography (top) and CHOP immunoblot (bottom) are shown. (E) Schematic diagram of the CHOP protein. The region containing the stress-inducible phosphorylation sites is stippled, and the peptide sequence of this area, from four mammalian species, shows the conservation of context of Ser⁷⁸ and Ser⁸¹ (19).



Departments of Medicine and Cell Biology, Skirball Institute of Biomolecular Medicine, and the Kaplan Cancer Center, New York University Medical Center, New York, NY 10016, USA.

*To whom correspondence should be addressed. E-mail: rond01@saturn.med.nyu.edu