MOLECULAR AND CELLULAR BIOLOGY, Dec. 1986, p. 4425–4432
0270-7306/86/124425-08$02.00/0

Vol. 6, No. 12

Highly Mutable Sites for ICR-170-Induced Frameshift Mutations Are Associated with Potential DNA Hairpin Structures: Studies with *SUP4* and Other *Saccharomyces cerevisiae* Genes

D. MICHAEL HAMPSEY,[1]† * RAYMOND A. KOSKI,[2]‡ AND FRED SHERMAN[1,3]

Departments of Biophysics[1] and Biochemistry,[3] University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, and Department of Genetics,[2] University of Washington, Seattle, Washington 98195

Received 9 May 1986/Accepted 18 August 1986

The majority of the mutations induced by ICR-170 in both the *CYC1* gene (J. F. Ernst et al. Genetics 111:233–241, 1985) and the *HIS4* gene (L. Mathison and M. R. Culbertson, Mol. Cell. Biol. 5:2247–2256, 1985) of the yeast *Saccharomyces cerevisiae* were recently shown to be single G · C base-pair insertions at monotonous runs of two or more G · C base pairs. However, not all sites were equally mutable; in both the *CYC1* and *HIS4* genes there is a single highly mutable site where a G · C base pair is preferentially inserted at a $\frac{-G-G-G-}{-C-C-C-}$ sequence. Here we report the ICR-170 mutagen specificity at the *SUP4*-o tyrosine tRNA gene of yeast. Genetic fine structure analysis and representative DNA sequence determination of ICR-170-induced mutations revealed that there is also a single highly mutable site in *SUP4*-o and that the mutation is a G · C base-pair insertion at a monotonous run of G · C base pairs. Analysis of DNA sequences encompassing the regions of highly mutable sites for all three genes indicated that the mutable sites are at the bases of potential hairpin structures; this type of structure could not be found at any of the other, less mutable G · C runs in *SUP4*, *CYC1*, and *HIS4*. Based on these results and recent information regarding novel DNA structural conformations, we present a mechanism for ICR-170-induced mutagenesis: (i) ICR-170 preferentially binds to DNA in the β conformation; factors that increase the temporal stability of this structure, such as adjacent stem-and-loop formation, increase the frequency of ICR-170 binding; (ii) the observed mutagen specificity reflects formation of a preferred ICR-170 intercalative geometry at $\frac{-G-G-}{-C-C-}$ sites; (iii) during replication or repair, ICR-170 remains associated with the single-stranded template; (iv) stuttering or strand slippage by the polymerization complex as it encounters the mutagen results in nucleotide duplication; (v) subsequent replication or mismatch repair fixes the insertion into the genome. This mechanism accounts for both the ICR-170 mutagenic specificity and the molecular basis of the highly mutable sites in *S. cerevisiae*.

The mechanisms by which frameshift mutations occur have been the subject of intense interest since the original description of this class of mutations in 1961 (7). Streisinger et al. (48) proposed a general model to account for frameshift mutagenesis, suggesting that these mutations occur in regions of strand discontinuity, for example, as occurs during replication or repair, and that the actual frameshift is generated within sequences of base-pair redundancy. Additions or deletions therefore arise as a result of strand slippage in either the primer or the template strand, respectively. This model accounted for the increased frequency of frameshift mutations induced with intercalating agents (27, 28) by suggesting that these compounds stabilize mispaired sequences. A corollary to this model is an explanation of the molecular basis of highly mutable sites, or "hotspots," initially described by Benzer (4). If mispairing is promoted in regions of monotonous stretches of base pairs, then a higher frequency of frameshift mutations should correlate with increased lengths of base-pair redundancy (48). Indeed, this has been shown to be the case for two highly mutable sites in the lysozyme gene of bacteriophage T4 (36, 49). Despite the general acceptance as a mechanism for the generation of frameshift mutations, this model leaves a number of unanswered questions. Most notably it does not explain why certain agents known to effectively intercalate within the DNA helix are poor inducers of frameshift mutations, or why equal lengths of base-pair redundancy are often susceptible to frameshift mutations at widely varied frequencies.

The objective of the work presented here is to provide additional insight into the mechanisms involved in frameshift mutagenesis and, in particular, to explain the molecular basis of certain highly mutable sites. The inducing mutagen used in this study is from the ICR series, a class of planar, heterocyclic compounds to which an alkylating side chain is conjugated (12). The mutagenic activity of one of these compounds, ICR-170 {2-methoxy-6-chloro-9-[3-(ethyl-2-chloroethyl)aminopropylamino]acridine dihydrochloride} (Fig. 1), has been reported in *Saccharomyces cerevisiae* (8), and recent results describe its specificity. A total of 33 independent forward mutations induced by ICR-170 at the *CYC1* locus were distributed over 16 different sites, with over one-third occurring at a single site (16). Sequence analysis of representative alleles, which included the highly mutable site, indicated that ICR-170 induces primarily but not exclusively single G · C base-pair insertions at monotonous runs of three G · C base pairs. Similar specificity by ICR-170 was shown at the *HIS4* locus. Thirty-nine independent isolates with mutations mapping at 16 different sites were isolated; as for *CYC1* there was a preferred site of action, with 8 of 39 mutations occurring at a single site (13). DNA sequence analysis of 21 representatives showed that ICR-170 induced exclusively single G · C base-pair inser-

* Corresponding author.
† Present address: Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, Shreveport, LA 71130.
‡ Present address: Amgen, Thousand Oaks, CA 91320.

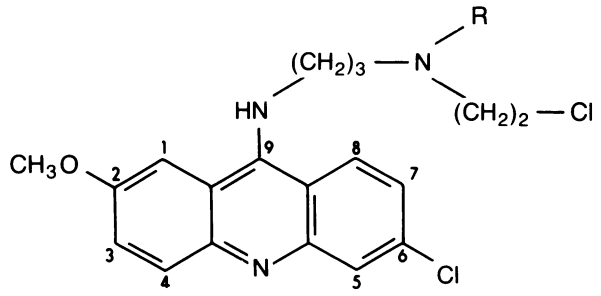

FIG. 1. Chemical structures of ICR-170 and ICR-191. For ICR-170, R is $—CH_2CH_3$; for ICR-191, R is —H.

tions at monotonous runs of two or more G · C residues (14, 32).

The *SUP4* tyrosine tRNA gene is another yeast gene that is amenable to genetic and biochemical analysis and that allows genetic selection of forward and reverse mutations. These methods previously have been used to define the nucleotide changes of point mutations and deletions arising spontaneously at this locus (19, 25, 26, 41). We have applied these methods to the study of ICR-170-induced mutagenesis at *SUP4*. Consistent with the ICR-170 specificity at *CYC1* and *HIS4* and in contrast to the spontaneous mutations at *SUP4* only G · C base-pair insertions occur; moreover, as in *CYC1* and *HIS4*, a single, highly mutable site is recognized. Based on flanking sequence information and potential formation of stem-and-loop structures at each of the hotspots in *SUP4*, *CYC1*, and *HIS4*, we suggest a mechanism for ICR-170-induced mutagenesis that includes a molecular basis for the presence of these highly mutable sites.

MATERIALS AND METHODS

Genetic nomenclature. Two symbols, lowercase *sup4* and uppercase *SUP4*, are used to designate various alleles of the tyrosine tRNA gene at the *SUP4* locus; the lowercase *sup4* denotes the recessive wild-type allele when used with the dominant, ochre suppressor mutant allele, designated *SUP4-o*. Second site mutations, resulting in loss of *SUP4-o* suppressor activity, are distinguished from the wild-type *sup4* allele by the designated *sup4-R*.

Strains. The *S. cerevisiae* parent strain from which the ICR-170 mutants were isolated was J15-8D (*MATa SUP4-o ade2-1 lys2-1 trp5-2 leu1-12 can1-100 met4-1 ura1-1*). This strain and *sup4-R* strains used for fine structure mapping were described by Kurjan and Hall (25). These strains were kindly provided by J. Kurjan (Columbia University).

Media. The yeast media used for isolating and screening *sup4* mutants were previously described (25). YPD medium consists of 1% yeast extract, 2% peptone, 2% glucose, 40 μg of adenine per ml, and 40 μg of uracil per ml. YT medium (33) contained ampicillin at 0.1 mg/ml.

Mutant isolation. *S. cerevisiae* J15-8D was grown to 2×10^2 to 5×10^7 cells per ml in YPD medium. Harvesting cells were washed twice by centrifugation with distilled water and suspended to 10^7 cells per ml in water. The cells were mutagenized by adding ICR-170 to 5 μg/ml and agitating for 30 min at 30°C in the dark. The cells were then diluted in YPD medium and plated on YPD under red light. After 2 days of incubation in darkness at 30°C, colonies were screened for the loss of *SUP4* suppression by the methods of Kurjan and Hall (25). Colonies were first replica plated to canavanine plates; canavanine-resistant colonies were then replica plated to media lacking adenine, methionine, lysine, or tryptophan. Putative *sup4-R* mutants were identified by the loss of suppression of the *can1-100*, *ade2-1*, *met4-1*, *lys2-1*, and *trp5-2* ochre-suppressible alleles. The *sup4-R* strains chosen for cloning and DNA sequence analysis were designated no. 895 through 900.

Fine structure mapping. A rapid plate assay for meiotic recombination at the *sup4* locus (25) was used to position ICR-170-induced *sup4-R* mutations relative to 12 *sup4-R* point mutations with known sequence alterations. ICR-170-induced mutants were mated to each of 12 characterized *sup4-R* mutant strains; diploids were selected and subsequently sporulated. The relative numbers of *ade2-1*- and *lys2-1*-suppressing *SUP4-o* recombinants derived from each cross were determined and used to establish an approximate map position of ICR-170-induced mutations within the *SUP4* gene sequence.

Isolation of *sup4-R* DNA. ICR-170-generated *sup4-R* alleles were cloned directly from yeast genomic DNA into the pBR322 vector as described by Kurjan et al. (26), with the following modifications. After double digestion of total genomic DNA with the restriction enzymes *Eco*RI and *Hin*dIII, the 4.5-kilobase *Eco*RI-*Hin*dIII restriction fragments, containing the *sup4* alleles, were enriched by electrophoresis in a 0.9% low-melting-point agarose gel (SeaKem). DNA fragments approximately 4.5 kilobases in size, as measured with *Hin*dIII-digested bacteriophage lambda DNA as the size marker, were excised from a portion of the gel in a volume of about 150 ml; DNA was recovered as described above, and the entire sample was ligated into *Eco*RI-*Hin*dIII double-digested pBR322 vector DNA (21). This solution was then used directly to transform freshly prepared transformation-competent *E. coli* HB101 cells (31). The transformed cells were plated on YT medium containing ampicillin. Ampicillin-resistant colonies were screened for the presence of *sup4-R* by using a radiolabeled oligonucleotide probe, described below, as described by Benton and Davis (3) for screening plaques, except that formamide was used at 36%. Approximately 1 to 3% of the colonies were positive.

Preparation of the *sup4*-specific probe and primer. The nucleotide sequence 5′ G-T-A-T-A-C-T-C-T-T-T-C-T-T-C-A-A-C 3′, corresponding to nucleotides −10 through −27 upstream of the *sup4* gene (24), was synthesized with the Applied Biosystems DNA Synthesizer model 380A, by using the procedures and reagents recommended by the manufacturer. This oligonucleotide, designated OL-33, was purified by polyacrylamide gel electrophoresis (29) and, when used as a probe, was radiolabeled (21) as described previously.

Plasmid purification. Plasmid DNA was prepared for sequence analysis as follows. A 10-ml culture on YT medium containing ampicillin was prepared by inoculating cells from a single bacterial colony and grown to stationary phase by agitation for 15 to 18 h at 37°C. Plasmid DNA was extracted by the alkaline lysis protocol (6). The plasmid-containing solution, in a volume of 600 μl, was extracted once with an equal volume of phenol-chloroform-isoamyl alcohol (12:12:1) and once with an equal volume of phenol alone and precipitated with 0.1 volume of 3 M sodium acetate and 2.5 volumes of ethanol. After centrifugation, the pellet was washed three times in cold 70% ethanol, dried under vacuum, and redissolved in 100 μl of TA buffer (35) containing 20 U of RNase T_1 and 2 μg of RNase A. Plasmid DNA was linearized in this solution by addition of 10 U of *Hin*dIII restriction enzyme; this solution was incubated at 37°C for 6 h. DNA was precipitated by addition of 40 μl of 30%

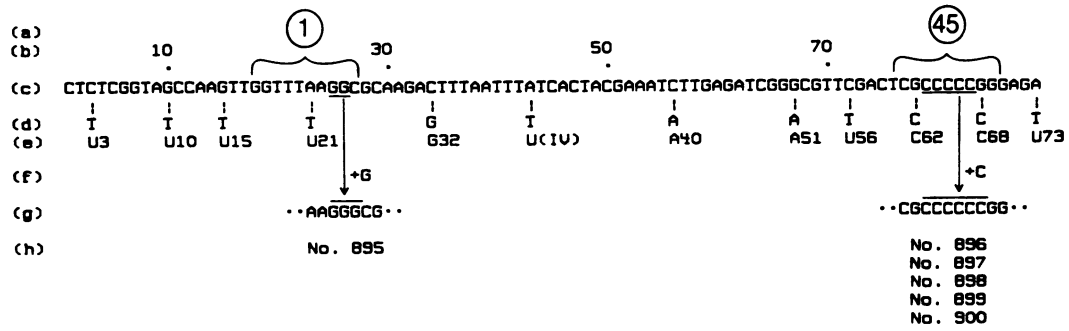

FIG. 2. Meiotic fine-structure map of ICR-170-induced *sup4-R* mutants and DNA sequences of representative mutations. Shown are the number of ICR-170-induced *sup4-R* mutants at each of the genetic fine-structure loci (a), the numbering system (b) and nucleotide sequence (c) of the *SUP4-o* gene, the mutations (d) and corresponding allelic designations (e) for the strains used in the fine-structure genetic mapping (26), and the determined ICR-170-induced mutations (f) and corresponding nucleotide sequences (g) of the *sup4-R* alleles from the indicated yeast strains (h). Nucleotide sequence 40 through 53 corresponds to the intervening sequence; the 3'-terminal CCA sequence of the tRNA is not encoded at the *sup4* locus and is therefore not indicated (19). This sequence was initially published elsewhere (19).

polyethylene glycol–1.8 M sodium chloride and stored at 4°C overnight (20). The solution was then centrifuged, and the supernatant was discarded; trace amounts of polyethylene glycol were removed with a drawn capillary pipette. The pellet was then redissolved in 40 μl of 0.3 M sodium acetate and precipitated by addition of 100 μl of ethanol and chilling to −70°C for 10 min. After centrifugation, the pellet was dried under vacuum and dissolved in 20 μl of 10 mM Tris hydrochloride (pH 7.4)–1 mM disodium EDTA. This procedure typically yielded 5 to 15 μg of purified plasmid DNA, which was then used directly for DNA sequence determination.

DNA sequence determination. The sequence of linearized plasmid DNA, purified as described above, was determined by the dideoxy terminator method (43, 45), by using the synthetic oligonucleotide OL-33, described above, as a primer. The sequencing label was α-[^{35}S]thio-dATP; sequencing reactions (5) and sequencing gels (45) were carried out as described previously.

RESULTS

Mutations of the *SUP4-o* locus. ICR-170 is an efficient mutagen for the inactivation of the *SUP4-o* ochre suppressor locus of yeast. The *sup4-R* mutants were selected by screening for concurrent loss of function of five different ochre-suppressible alleles in the mutagenized J15-8D parent strain. Such mutants were induced by ICR-170 at a frequency of approximately 10^{-3}

Meiotic fine-structure mapping demonstrated that ICR-170-induced mutations predominantly occurred at a single site within the *SUP4* gene. Indeed, 45 of the 46 mutations that mapped to the *sup4* locus are clustered in the 3′ region of the gene defined by the point mutations C62 and C68 (Fig. 2). The exceptional mutation mapped near the U21 mutation. These ICR-170-induced *sup4-R* mutations provided an opportunity to compare two sites that have distinct differences in their rate of mutation induced by ICR-170.

DNA sequences of the frameshift mutations. The most unequivocal method for establishing the nucleotide alterations induced by ICR-170 was to directly clone and sequence the *sup4-R* alleles. The appropriate DNA fragments containing *sup4-R* were directly transferred from the yeast genome into the pBR322 vector (see Materials and Methods). The DNA sequences of five representative alleles that map to the highly mutable site, along with the sequence of the parent allele, *SUP4-o*, are shown in Fig. 2. All of these *sup4-R* mutations are identical, containing a single G · C base-pair insertion at the monotonous stretch of G · C base pairs at nucleotide positions 79 through 83 (Fig. 2). Although this does not rigorously establish that all 45 of the ICR-170-induced mutations that map to this site are identical, we believe that most, if not all, of these mutations are identical, and certainly nucleotide positions 79 through 83 constitute a hotspot for ICR-170-induced G · C base-pair insertion.

The DNA sequence of the only ICR-170-induced intragenic mutant that did not map to the highly mutable site (Fig. 2) was also determined and shown to have a G · C base-pair insertion within a stretch of two G · C base pairs at nucleotide positions 25 through 26 (Fig. 2). This mutation is consistent with the mutagenic specificity of ICR-170 observed at the highly mutable site (see above) as well as with most other previously established ICR-170-induced mutations in *S. cerevisiae* (16, 32). Moreover, the DNA sequence of this mutation validates the genetic mapping, providing additional evidence that the indicated site (Fig. 2) is a hotspot for ICR-170-induced mutagenesis.

DISCUSSION

Genetic fine structure analysis indicates that an extensive collection of ICR-170-induced mutations in *SUP4-o* were exceptionally limited in their distribution. Of 46 mutations at the *sup4* locus, all but 1 mapped to the same site. Previous analyses showed that mutations at 30 different sites within the *SUP4-o* locus were capable of causing inactivation of the suppressor (1, 26). Since the same procedure was used to isolate the ICR-170 mutants reported here, the limited number of mutational sites cannot be attributed to the dispensibility of the other regions. It must therefore be concluded that the observed ICR-170-induced mutations reflect the mutational specificity of this mutagen.

Additional studies addressing the mutational spectrum of ICR-170 at other yeast genes have recently been reported. Characterization of ICR-170-induced mutations at the *CYC1* and *HIS4* genes showed that the majority of these mutations are identical in type to those presented here for the *SUP4* gene (16, 32). Clearly the predominant mutational specificity of ICR-170 in yeast is generation of single G · C base-pair insertions within monotonous runs of G · C base pairs. Frameshift deletions, A · T base-pair insertions, and G · C base-pair insertions at sites other than stretches of two or

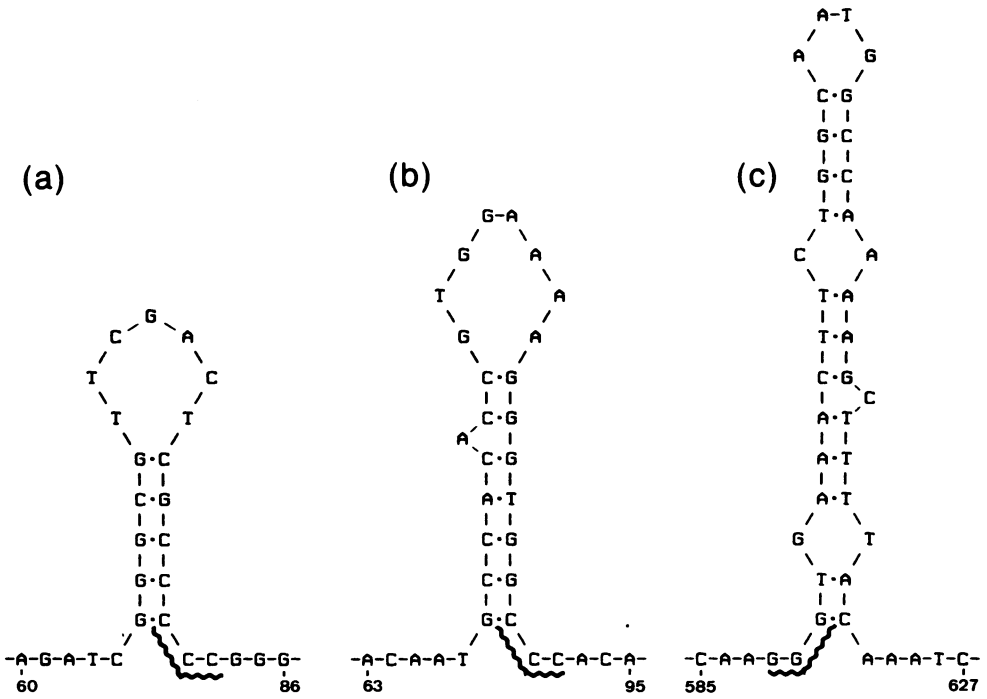

FIG. 3. Potential DNA hairpin structures of the ICR-170 preferred sites of mutagenic action in the *SUP4* (a), *CYC1* (b), and *HIS4* (c) genes. Computer-assisted analyses of potential DNA secondary structures were performed by using the program of Zuker and Stiegler (51). Both 25- and 40-nucleotide segments at 5-nucleotide intervals were analyzed by using the program of R. Moerschell (unpublished). The structures shown represent the most stable forms found over a 40-nucleotide region on either side of the mutation. The *SUP4*, *CYC1*, and *HIS4* structures shown here have, respectively, ΔG values of −12.4, −15.5, and −12.6 kcal (ca. −51.9, −64.9, and −52.7 kJ, respectively)/mol. All three structures place the site of the ICR-170-induced mutation at the base of the hairpin in a position that would preferentially leave these residues at least partially unpaired, e.g., in a β-DNA conformation. These sites are denoted by the wavy underlines. The numbers represent nucleotide positions: for *SUP4* the numbering system is consistent with Fig. 2; for *CYC1* and *HIS4* the numbering systems are such that the A of the AUG initiation codons are +1.

more G · C base pairs are never seen; a limited number of base-pair substitutions are the only exceptions to this specificity (16). A similar mutational specificity was recently reported for ICR-170 in *Neurospora crassa*: −1 frameshift in the *am* (NADP-specific glutamate dehydrogenase) gene reverted by ICR-170-induced G · C base-pair insertion at $\frac{-G-G-}{-C-C-}$ sites. These revertants arose at an approximately 100-fold higher frequency than did ICR-170-induced revertants of an *am* +1 frameshift, which is known to revert by UV-induced G · C base-pair deletion at a $\frac{-G-G-G-}{-C-C-C-}$ site (9).

Since their initial description (4), highly mutable sites have been considered to be a particularly intriguing aspect of mutagenesis. Recognition of hotspots for frameshift mutations at sites of base-pair redundancy offered substantial support for the proposal that frameshifts occur as a consequence of slipped mispairing (36, 48). However, despite characterization of numerous highly mutable sites in both procaryotic and eucaryotic organisms, explanations of their molecular basis are rare and often ambiguous. There may, in fact, be a clear explanation for the occurrence of only a single type of hotspot: 5-methylcytosine residues are highly mutable sites for G · C-to-A · T transitions (11). Both the hotspot and mutagenic specificity are readily explainable for this particular mutation: hotspots occur at the DNA methylation recognition sequence (C-C-A-G-G) and the mutation arises as a consequence of deamination of 5-methylcytosine to thymine.

In an effort to understand the mechanism of action of ICR-170 and to assess the nature of the mutagenic hotspots, we have analyzed DNA sequences flanking the highly mutable sites in *SUP4*, *CYC1*, and *HIS4*. Since the complete nucleotide sequences of the *SUP4* (19), *CYC1* (46), and *HIS4* (15, 32) structural genes are known, we also extended this analysis to a comparison of structural features at other, less mutable G · C runs in these three genes. The preferred sites of ICR-170 action in *SUP4* and *CYC1* contain contiguous or adjacent runs of two or more G · C base pairs lying in opposite orientation to those comprising the mutational site (Fig. 2; see reference 16 for the *CYC1* sequence). However, the same feature is not found at the *HIS4* hotspot (see references 15 and 32 for the *HIS4* sequence and identification of the hotspot). Moreover, there are less mutable sites of two or more G · C base pairs in these three genes that are flanked by adjacent G · C base pairs in either orientation. In fact, there is no obvious consensus in the local DNA sequences at the three mutational hotspots that is clearly distinct from sequences flanking the less mutable sites. Thus ICR-170 is *not* simply recognizing a preferred DNA sequence. Consideration of potential DNA secondary structures, however, indicates that stable hairpins could potentially form at each of the three highly mutable sites. Strikingly, each hairpin positions the site of the observed frameshift at the base of the structure (Fig. 3). Computer-assisted analyses of DNA secondary structures showed that these three structures also represent the most stable hairpins that can form in these regions. A similar analysis at the less mutable sites of ICR-170-induced frameshift mutations (1 site in *SUP4*, 3 in *CYC1*, and 15 in *HIS4*), and at all other sites of three or more monotonous G · C base-pairs (2 sites in *SUP4*, 2 in *CYC1*, and 20 in *HIS4*), showed that DNA

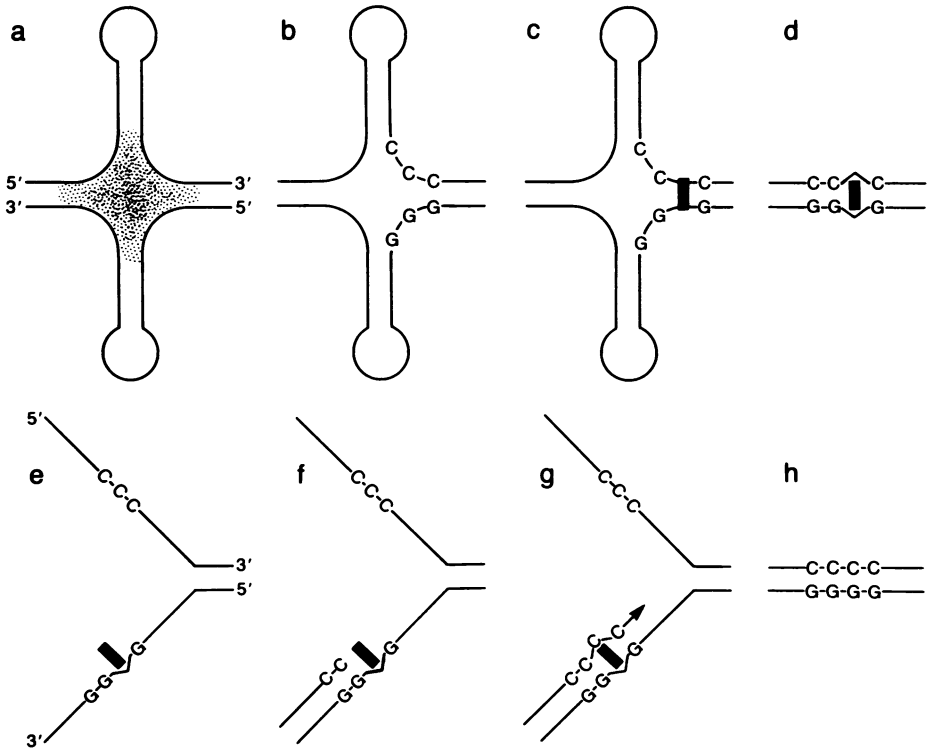

FIG. 4. Proposed mechanism of ICR-170-induced frameshift mutagenesis. (a) Factors such as adjacent inverted repeats capable of hairpin structure formation promote temporal stabilization of β-DNA (shaded area). (b) G · C base pairs in the β-DNA conformation provide an ideal intercalative geometry (c) for ICR-170, denoted by the solid rectangle. (d) ICR-170 remains associated (perhaps covalently) with the DNA in an intercalative geometry when the normal DNA conformation resumes. (e) During replication or repair ICR-170 remains associated with the single-stranded template DNA. This figure shows ICR-170 bound to the -G-G-G- sequence, although it could have been drawn as bound to the -C-C-C- sequence, unless, of course, it is covalently bound to guanine (see the text). (f) ICR-170 blocks replication, resulting in stuttering or strand slippage, such that the nucleotide preceding the mutagen is recopied (g) before proceeding past the block. (h) Replication or misrepair results in permanent G · C base-pair addition to the genome.

secondary structures simlar to those shown in Fig. 3 would not form. DNA stem-and-loop formation would invoke locally altered conformations, which have been recently characterized (see below) (47). Accordingly, the association of ICR-170 highly mutable sites with potential DNA hairpin structures suggests a mechanism of action for this compound. (Stem-and-loop DNA structures have been previously implicated at highly mutable ICR-170 sites [16]; however, the additional information from *sup4-R* and *his4* mutations has provided further insight into the specific mechanism of ICR-170-induced mutagenesis.)

The DNA sequence at the base of a hairpin structure would be partially single stranded and occupy a unique structural geometry. It has been proposed that locally denatured DNA, such as the DNA at the replication fork, exists as a structural intermediate denoted β-DNA (2, 42). β-DNA is a structural variant of Watson-Crick B-DNA, differing in its sugar puckering pattern and backbone conformation angle and exhibiting varied degrees of base unstacking. Crystallographic studies of various dye- and drug-nucleic acid complexes provide direct evidence for the existence of β-DNA (23, 37, 42). Recently, Sobell (47) postulated that β-DNA is an obligatory structural intermediate in DNA melting; β-DNA necessarily exists at the boundary between single-stranded and double-stranded entities within the DNA molecule. Accordingly, the structures presented in Fig. 4, or even partially formed stem-and-loop structures, would generate β-DNA structural intermediates. Furthermore, Sobell and his colleagues demonstrated that the β-DNA structure provides a geometry that promotes intercalative binding of 9-aminoacridine to DNA (42). It is therefore likely that β-DNA also provides a tight binding site for intercalating agents such as ICR-170, which is a derivative of 9-aminoacridine (Fig. 1).

The observation that stable DNA hairpin structures could form at the ICR-170 highly mutable sites in the *SUP4*, *CYC1*, and *HIS4* genes, combined with the characterization of β-DNA structural intermediates, suggests that following mechanism for ICR-170-induced mutagenesis. (i) ICR-170 preferentially binds to DNA in the β conformation (Fig. 4a); factors that increase the temporal stability of this structure, such as transient stem-and-loop formation, would increase the frequency of ICR-170 association. (ii) The observed mutagen specificity reflects formation of a preferred ICR-170 intercalative geometry at $\frac{-G-G-}{-C-C-}$ sites (Fig. 4b and c). (iii) As the DNA unwinds during replication or repair, ICR-170 remains associated with the single-stranded template (Fig. 4d and e). (iv) The polymerization complex encounters the mutagen, resulting in either stuttering or strand slippage so that the preceding nucleotide is recopied, resulting in a nucleotide duplication on the growing primer or daughter strand (Fig. 4f and g) (42, 49). (v) Subsequent replication or mismatch repair then "fixes" the mutation into the genome (Fig. 4h).

This model accounts for the nucleotide specificity of the mutation (+G · C), the requirement for stretches of monot-

onous G · C base pairs at the mutational site, and the observation that the frameshift is always an insertion and never a deletion. Studies of ICR-170-induced mutagenesis in *Neurospora* species show a requirement for the polyamine side chain (30) (Fig. 1), indicating that ICR-170 may covalently bind the DNA, probably at the N-7 position of guanine (40); this has also been suggested for the binding of a related compound, ICR-191 (see below), to DNA in *Escherichia coli* (44). Although the mechanism of ICR-170-induced mutagenesis in yeast might also involve covalent attachment to the DNA, the actual specificity must be determined by the intercalative geometry at $\frac{-G-G-}{-C-C-}$ sites to account for frameshift mutations occurring only at runs of two or more G · C base pairs. This mechanism is necessarily different from the mechanism(s) of spontaneous frameshift mutation in yeast, since the only spontaneous frameshift identified within the ICR-170 hotspot of *SUP4* is a G · C base-pair deletion (26). The recent characterization of ICR-170-induced frameshift revertants in the *am* gene of *N. crassa* (9), showing preferential G · C base-pair insertions at stretches of G · C base pairs, is consistent with our proposed model.

We wish to emphasize that the hairpin structures depicted in Fig. 3 and 4a through c would not be expected to exist as permanent or even stable structural entities within the genome. Their formation is expected to be transient, and they might occur only during metabolic processes that involve local DNA strand separation, such as during transcription or replication. Furthermore, their formation need occur only infrequently, since ICR-170-induced mutations are also rare events. There is also increasing evidence that DNA hairpin structures do indeed form in vivo. For example, cruciform structures were recently observed in electron microscopic studies of psoralen cross-linked simian virus 40 DNA (22).

Although the mechanism proposed here associates ICR-170 highly mutable sites with formation of DNA hairpin structures, it is fundamentally different from other proposals that invoke hairpin structures, either as part of their mutagenic mechanisms or to account for differential rates of mutation. For example, Ripley and Glickman (38) suggested that the stem of imperfect hairpin structures might serve as a template during repair so that one strand of the stem is "repaired" to generate a closer complement of the other. This mechanism has been invoked to account for a variety of complex multiple-base-pair changes (38; M. Hampsey, J. Ernst, J. Stewart, and F. Sherman, submitted) and more recently was suggested as a mechanism to account, in part, for nucleotide variability in human alpha interferon gene phylogeny (18). The formation of DNA hairpin structures has also been suggested to account for varied rates of UV-induced G · C-to-A · T transitions (50). According to this proposal, the potential to position the mutagenic target site in a hairpin loop correlates with increased frequency of UV-induced mutation. There are data from studies in the *lacI* gene of *E. coli* both supportive (50) and contrary (34) to this proposal. By contrast, the mechanism presented here proposes that hairpin structure formation correlates to increased frequency of mutagenesis simply by generating, immediately adjacent to the hairpin (Fig. 3), the preferred structural geometry for ICR-170 intercalation. This represents a novel explanation for the molecular basis of certain highly mutable sites. Recent evidence also describes spontaneous deletions and duplications of two base-pair sequences that are consistently found adjacent to DNA hairpin structures in bacteriophage T4, supporting a role for DNA secondary structure in promoting adjacent frameshift mutations (L. S. Ripley, A. Clark, and J. G. deBoer, J. Mol. Biol., in press); it was previously suggested that such hairpin structures might act to slow or block processive DNA elongation and thus promote frameshift mutations (39).

The specificity of action of a related compound, ICR-191 {2-methoxy-6-chloro-9-[3-(2-chloroethyl)aminopropylamino]acridine dihydrochloride} (Fig. 1) has been investigated in *E. coli* at the *lacI* gene (10) and in lambda prophage (44). The extensive analysis by Calos and Miller (10) showed the following. (i) ICR-191 acts preferably at monotonous stretches of G · C base pairs, where either frameshift additions or deletions are predominantly observed. (ii) There are relative hotspots for this mutagen; a greater than 30-fold difference in the frequency of mutation at sites with three or more G · C base pairs is seen. (iii) There is at best only a weak correlation between the frequency of mutagenesis and the DNA context flanking the G · C runs. Since the entire sequence of the *lacI* gene is known (17), we were able to assess whether hotspots for ICR-191-induced mutations correlate with formation of hairpin structures of the type shown in Fig. 4. Despite an extensive analysis at each of the 24 $\frac{-G-G-G-}{-C-C-C-}$ or $\frac{-G-G-G-G-}{-C-C-C-C-}$ sites at which frameshifts occur, no strong correlation between frequency of mutagenesis and potential formation of hairpin structures at the target site was found. Even though ICR-170 and ICR-191 have a common preference for monotonous G · C runs as their targets, the actual mechanisms of promoting frameshift mutations might be fundamentally different. This is consistent with other differences between these two mutagens, including the apparent generation of only frameshift insertions and not deletions by ICR-170 and the relative inactivity of ICR-191 in *S. cerevisiae*.

The mechanism proposed here for ICR-170-induced frameshift mutagenesis at monotonous runs of G · C base pairs makes a number of testable predictions. ICR-170 should bind preferentially to poly(dG-dC). Crystal structure studies of the type showing binding of 9-aminoacridine to DNA (42) should show that ICR-170 bound to $\frac{-G-G-}{-C-C-}$ is intercalated in an asymmetric binding mode consistent with the β-DNA conformation. Moreover, factors that would change the temporal stability of β-DNA should correlate to altered frequencies of ICR-170-induced mutations. Such experiments are directly approachable by in vitro construction of phenotypically neutral base-pair changes, as can readily be made in the *S. cerevisiae CYC1* gene (21), that would increase or decrease the stability of hairpin structures at the highly mutable sites.

ACKNOWLEDGMENTS

We thank Benjamin Hall (University of Washington) for guidance during the genetic aspects of this work, Henry M. Sobell (University of Rochester) for valuable discussions, Richard Moerschell for assistance with the computer analyses of secondary structures, and Sarah Fingar for synthesizing the oligonucleotide OL-33.

This investigation was supported by Public Health Service training grant T32 GM07098 from the National Institutes of Health, by the National Institutes of Health research grant GM11895, and by the United States Department of Energy grant DE-FG02-85ER60281. D.M.H. was an American Cancer Society postdoctoral fellow; R.A.K. was a National Institutes of Health postdoctoral fellow.

LITERATURE CITED

1. **Allison, D. S., S. H. Goh, and B. D. Hall.** 1983. The promoter sequence of a yeast $tRNA^{tyr}$ gene. Cell **34**:655–664.

2. **Banerjee, A., and H. M. Sobell.** 1983. Presence of nonlinear excitations in DNA structure and their relationship to DNA premelting and to drug intercalation. J. Biomol. Struct. Dyn. **1**:253–262.
3. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
4. **Benzer, S.** 1961. On the topography of the genetic fine structure. Proc. Natl. Acad. Sci. USA **47**:403–416.
5. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
6. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
7. **Brenner, S., L. Barnett, F. H. C. Crick, and A. Orgel.** 1961. The theory of mutagenesis. J. Mol. Biol. **3**:121–124.
8. **Brusick, D. J.** 1970. The mutagenic specificity of ICR-170 in *Saccharomyces cerevisiae*. Mutat. Res. **10**:11–19.
9. **Burns, P. A., J. H. Kinnaird, B. J. Kilbey, and J. R. S. Fincham.** 1986. Sequencing studies of ICR-170 mutagenic specificity in the *am* (NADP-specific glutamate dehydrogenase) gene of *Neurospora crassa*. Genetics **113**:45–51.
10. **Calos, M. P., and J. H. Miller.** 1981. Genetic and sequence analysis of frameshift mutations induced by ICR-191. J. Mol. Biol. **153**:39–66.
11. **Coulondre, C., J. H. Miller, P. J. Farabaugh, and W. Gilbert.** 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. Nature (London) **274**:775–780.
12. **Creech, H. J., E. Breuninger, R. F. Hankwitz, Jr., G. Polsky, and M. L. Wilson.** 1960. Quantitative studies of the effects of nitrogen mustard analogs and other alkylating agents on ascites tumors in mice. Cancer Res. **20**(Suppl.):471–494.
13. **Culbertson, M. R., L. Charnas, M. T. Johnson, and G. R. Fink.** 1977. Frameshifts and frameshift suppressors in *Saccharomyces cerevisiae*. Genetics **86**:745–764.
14. **Donahue, T. F., P. J. Farabaugh, and G. R. Fink.** 1981. Suppressible glycine and proline four base codons. Science **212**:455–457.
15. **Donahue, T. F., P. J. Farabaugh, and G. R. Fink.** 1982. The nucleotide sequence of the *HIS4* region of yeast. Gene **18**:47–59.
16. **Ernst, J. F., D. M. Hampsey, and F. Sherman.** 1985. DNA sequence of frameshift and other mutations induced by ICR-170 in yeast. Genetics **111**:233–241.
17. **Farabaugh, P. J.** 1978. Sequence of the *lacI* gene. Nature (London) **274**:765–769.
18. **Golding, G. B., and B. W. Glickman.** 1985. Sequence-directed mutagenesis: evidence from a phylogenetic history of human α-interferon genes. Proc. Natl. Acad. Sci. USA **82**:8577–8581.
19. **Goodman, H. M., M. V. Olson, and B. D. Hall.** 1977. Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor *SUP4-o*. Proc. Natl. Acad. Sci. USA **74**:5453–5457.
20. **Guo, L.-H., and R. Wu.** 1983. Exonuclease III: use for DNA sequence analysis and in specific deletions of nucleotides. Methods Enzymol. **100**:60–96.
21. **Hampsey, D. M., G. Das, and F. Sherman.** 1986. Amino acid replacements in yeast iso-1-cytochrome *c*. Comparison with the phylogenetic series and the tertiary structure of related cytochromes *c*. J. Biol. Chem. **261**:3259–3271.
22. **Hsu, M.-T.** 1985. Electron microscopic evidence for the cruciform structure in intracellular SV40 DNA. Virology **143**:617–621.
23. **Jain, S. C., K. K. Bhandary, and H. M. Sobell.** 1979. Visualization of drug-nucleic acid interaction at atomic resolution. VI. Structures of two drug-dinucleotide monophosphate crystalline complexes: ellipiticine-5-iodocytidylyl (3′-5′) guanosine and 3,5,6,8-tetramethyl-N-methyl phenanthrolinium-5-iodocytidylyl (3′-5′) guanosine. J. Mol. Biol. **135**:813–840.
24. **Koski, R. A., D. S. Allison, M. Worthington, and B. D. Hall.** 1982. An *in vitro* RNA polymerase III system from *S. cerevisiae*: effects of deletions and point mutations upon *SUP4* gene transcription. Nucleic Acids Res. **10**:8127–8143.
25. **Kurjan, J., and B. D. Hall.** 1982. Mutations at the *Saccharomyces cerevisiae SUP4* $tRNA^{Tyr}$ locus: isolation, genetic fine-structure mapping and correlation with physical structure. Mol. Cell. Biol. **2**:1501–1513.
26. **Kurjan, J., B. D. Hall, S. Gillam, and M. Smith.** 1980. Mutations at the yeast *SUP4* $tRNA^{Tyr}$ locus: DNA sequence changes in mutants lacking suppressor activity. Cell **20**:701–709.
27. **Lerman, L. S.** 1961. Structural considerations in the interaction of DNA and acridines. J. Mol. Biol. **3**:18–30.
28. **Lerman, L. S.** 1963. The structure of the DNA-acridine complex. Proc. Natl. Acad. Sci. USA **49**:94–102.
29. **Lo, K.-M., S. S. Jones, N. R. Hackett, and H. G. Khorana.** 1984. Specific amino acid substitutions in bacterioopsin: replacement of a restriction fragment in the structural gene by synthetic DNA fragments containing altered codons. Proc. Natl. Acad. Sci. USA **81**:2285–2289.
30. **Malling, H. V.** 1967. The mutagenicity of the acridine mustard (ICR-170) and the structurally related compounds in *Neurospora*. Mutat. Res. **4**:265–274.
31. **Mandel, M, and A. Higa.** 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159–162.
32. **Mathison, L., and M. R. Culbertson.** 1985. Suppressible and nonsuppressible +1 G · C base-pair insertions induced by ICR-170 at the *his4* locus in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **5**:2247–2256.
33. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Miller, J. H.** 1985. Mutagenic specificity of ultraviolet light. J. Mol. Biol. **182**:45–68.
35. **O'Farrell, P. H., E. Kutter, and M. Nalcanishi.** 1980. A restriction map of the bacteriophage T4 genome. Mol. Gen. Genet. **179**:421–435.
36. **Okada, Y., G. Streisinger, J. E. Owen, J. Newton, A. Tsugita, and M. Inouye.** 1972. Molecular basis of a mutational hot spot in the lysozyme gene of bacteriophage T4. Nature (London) **236**:238–241.
37. **Reddy, B. S., T. P. Seshadri, T. D. Sakore, and H. M. Sobell.** 1979. Visualization of drug-nucleic acid interaction at atomic resolution. V. Structures of two aminoacridine-dinucleoside monophosphate crystalline complexes, proflavin 5′-iodocytidylyl (3′-5′) guanosine and acridine orange-5-iodocytidylyl (3′-5′) guanosine. J. Mol. Biol. **135**:787–812.
38. **Ripley, L. S., and B. W. Glickman.** 1983. Unique self-complementarity of palindromic sequences provides DNA structural intermediates for mutation. Cold Spring Harbor Symp. Quant. Biol. **47**:851–861.
39. **Ripley, L. S., B. W. Glickman, and N. B. Shoemaker.** 1983. Mutator versus antimutator activity of a T4 DNA polymerase mutant distinguishes two different frameshifting mechanisms. Mol. Gen. Genet. **189**:113–117.
40. **Roth, J. R.** 1974. Frameshift mutations. Annu. Rev. Genet. **8**:319–346.
41. **Rothstein, R.** 1979. Deletions of a tyrosine tRNA gene in *S. cerevisiae*. Cell **17**:185–190.
42. **Sakore, T. D., B. S. Reddy, and H. M. Sobell.** 1979. Visualization of drug-nucleic acid interactions at atomic resolution. IV. Structure of an aminoacridine-dinucleotide monophosphate crystalline complex, 9-aminoacridine-5-iodocytidylyl (3′-5′) guanosine. J. Mol. Biol. **135**:763–785.
43. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
44. **Skopek, T. R., and F. Hutchinson.** 1984. Frameshift mutagenesis of lambda prophage by 9-aminoacridine, proflavin and ICR-191. Mol. Gen. Genet. **195**:418–423.
45. **Smith, A. J. H.** 1980. DNA sequence analysis by primed synthesis. Methods Enzymol. **65**:560–580.
46. **Smith, M., D. W. Leung, S. Gillam, C. Astell, D. L. Montgomery, and B. D. Hall.** 1979. Sequence of the gene for iso-1-cytochrome *c* in *Saccharomyces cerevisiae*. Cell **16**:753–761.
47. **Sobell, H. M.** 1985. Actinomycin and DNA transcription. Proc. Natl. Acad. Sci. USA **82**:5328–5331.
48. **Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E.**

Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31**:77–84.
49. **Streisinger, G., and J. E. Owen.** 1985. Mechanisms of spontaneous and induced frameshift mutations in bacteriophage T4. Genetics **109**:633–659.
50. **Todd, P. A., and B. W. Glickman.** 1982. Mutational specificity of UV light in *Escherichia coli*: indications for a role of DNA secondary structures. Proc. Natl. Acad. Sci. USA **79**:4123–4127.
51. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.