European Journal of Cancer Vol. 30A, No. 4, pp. 567–568, 1994.

0959–8049/94 $7.00 + 0.00

0959-8049(93)E0091-4

The Relevance of Drug DNA Sequence Specificity to Anti-tumour Activity

S. Neidle, M.S. Puvvada and D.E. Thurston

IT IS well-established that many clinically useful anti-tumour agents act at the DNA level [1]. The intercalators form "cleavable complexes" involving DNA topisomerase II, and alkylators, such as cis-platinum or the mustards, produce covalent cross-links in DNA. Few of these classic drugs have significant sequence specificity that extends beyond two to three base pairs (although several mustards react preferentially with runs of GC base pairs which, it is suggested [2], may relate to their anti-tumour effectiveness in tumours such as Burkitt's lymphoma with long GC stretches in the associated Epstein–Barr viral sequences). The drug-induced cleavage of DNA by topoisomerase II does show site selectivity that depends on the nature of the drug, but this only involves bases in the immediacy of a cleavage site [3]. Many DNA binding agents tend to have general cytotoxic properties, and selectivity is generally in favour of rapidly growing tumour cells, rather than discriminating against any fundamental differences between normal and tumour cells. The identification of human genes that are responsible for transforming activity (oncogenes) [4] has prompted studies in many laboratories aimed at discovering potential new drugs that are capable of specifically inhibiting the expression of particular oncogene proteins, thereby being specific for tumour cells that depend on such oncoproteins for their viability, and therefore lacking general cytotoxic properties. Much current activity is focused on antisense and antigene oligonucleotides targeted against mRNA and genomic DNA, respectively, which, in *in vitro* experiments, have been shown to act as specific inhibitors of translation and transcription for genes such as c-*myc*, *bcl*-2 and c-Ha-*ras* [5, 6]. However, oligonucleotides as drugs have numerous scientific, technological and economic problems associated with them. Concomitant with this has been the development in a number of laboratories, including our own, of natural and synthetic lower molecular weight molecules that are also capable of recognising specific double-stranded DNA sequences, yet have considerable potential stability, synthetic, pharmacokinetic and economic advantages compared to natural and modified oligonucleotides (see, for example [7–9]). As yet, such compounds are only capable of recognising short (six to eight base pairs maximum) lengths of DNA; a length of approximately 15–16 consecutive base pairs has been considered to constitute a unique site on the human genome, on a statistically random bias [10].

Correspondence to S. Neidle at the CRC Biomolecular Structure Unit, The Institute of Cancer Research, Cotswold Road, Sutton, Surrey SM2 5NG, UK.
M.S. Puvvada and D.E. Thurston are at the Division of Medicinal Chemistry, School of Pharmacy and Biomedical Sciences, University of Portsmouth, King Henry I Street, Portsmouth, Hants PO1 2DZ, U.K.
Revised 15 Oct. 1993; accepted 21 Oct. 1993.

We have recently designed a novel series of *bis*-DC-81 (an anthramycin analogue) interstrand DNA cross-linking agents, some of which possess extraordinarily high cytotoxicity in L1210 and other cell lines [11, 12]. The cytotoxic activity of the most active compound, "DSB-120", is also highly selective within a panel of ovarian cell lines (Kelland L, personal communication). We have been concerned to know whether this activity is related to the DNA sequence specificity of DSB-120, which is selective for sequences of the type 5′-(Pu/Py)GATC(Py/Pu). We are also interested in the wider question of whether and how the anti-tumour activity of sequence-selective compounds in general is related to the actual genomic DNA sequences involved. It has been firmly established that the DNA cross-linking ability of these agents correlates with cytotoxicity and anti-tumour activity [12]. The repairability of cross-linked adducts from groove-binding agents may also play a role.

A priori, the four sequences 5′-(Pu/Py)GATC(Py/Pu) being short, should occur with a high frequency, of about 1 in 500 in total [9], assuming a random distribution of sequences. We have analysed a representative number of oncogene sequences, from the EMBL databank, in order to examine the *actual* frequencies of the occurrence of these DSB-120 sequences. The total number of DNA bases in the sample is 80 138.

Table 1 shows distributions of occurrence of these four sequences, selected from a set of 19 randomly chosen oncogenes in the DNAstar database [13]. Several oncogenes (for example, v-*src* and *bcl*-3) have higher actual frequencies of occurrence of DSB-120 binding sites than predicted on the basis of randomly expected sites. Others such as *abl*, v-*abl*, *bcl*-2 and the two *ras* genes are significantly underrepresented, and indeed several of them do not have any binding sites at all. We have also examined the actual frequency of occurrence of one of these four sequences (5′-GCATGC) in the total primate database, of 12 497 613 bases in the EMBL databank. There are 1645 occurrences in total, i.e. a 1 in 7597 frequency. This is to be compared with the statistically random expected frequency of 1 in 2080, indicating that this particular sequence is underrepresented in a much larger sample.

We have repeated our searches, now with the 12 base pair sequence 5′-(Pu)GATC(Py)(Pu)GATC(Py) (one of the four possibilities) as a representative example of a longer sequence based on this motif, for which it may be possible to synthesise longer molecules in the future. Over the 19 oncogene sequences, we find only two highly significant non-zero occurrences, one on each of two genes. We would expect a random frequency of occurrence of approximately 1 in 525 000. Instead, we find a site on each of the two genes *raf* and *int*-2 (which are 4249 and 11608 bases long, respectively).

These preliminary results suggest that DSB-120 will interact to a greater extent with some DNA sequences compared to others. Since DNA binding of cross-linkers is generally directly related to their cytotoxic effects, we therefore suggest that DSB-120 will have a proportionally greater effect in those cells that are dependent on and strongly expressing particular oncogenes rather than others. This in turn may suggest that these agents have the potential to show differential effects on tumours that have elevated levels of particular oncogenes. Conversely, one may not expect compound DSB-120 to have a significant effect in, for example, B-cell lymphomas/leukaemias, for which the *bcl*-2 translocation has been implicated, or in chronic myelogenous leukaemia where the *bcr-abl* translocation is strongly implicated [4]. We might expect other sequence-specific agents such as CC-1065 and its recently developed analogue carzelesin [14] (currently in clinical trial) to show equivalent differential oncogene selectivity. As the length of DNA sequence being recognised

Table 1. Binding site distribution for DSB-120 in representative oncogene sequences

DNAstar reference	Gene sequence	No. of bases in gene fragment	No. of expected sites for DSB-120 based on a 1:500 probability*	Search for 5′-PuGATCPy†	Search for 5′-PyGATCPu†	Ratio of sites found to those expected based on probability§
				No. of sites found‡	No. of sites found‡	
M30444	c-*myc*	369	0.74	1	0	1.4
X02305	v-*src*	399	0.79	2	1	3.8
X01228	Hs-242	458	0.92	0	0	0
M20408	D-*ets*-2-proto	631	1.26	1	1	1.6
M16598	D*raf*-1	938	1.88	2	0	1.1
X06691	abl	986	1.97	0	0	0
K01884	Blym-1	1004	2.0	0	0	0
M20013	Hum-Burkitts	1222	2.44	2	0	0.8
K01043	dsrc	1343	2.69	2	3	1.9
X51898	*bcl-2*	1394	2.79	0	0	0
M36181	D-*jun*	1440	2.88	4	0	1.4
X17363	Ha-*ras*1	1464	2.93	0	0	0
M31731	*bcl-3*	2282	4.56	6	1	1.5
M132354	v-*abl*	2465	4.93	1	1	0.4
X07181	*raf* proto	4249	8.5	8	4	1.4
M54968	K-*ras*	5775	11.55	2	1	0.3
M19720	L-*myc*	7011	14.0	6	1	0.5
X14445	*int*-2-proto	11608	23.2	9	1	0.4
X14720	c-*fms*	35100	70.2	46	16	0.9

*Number of predicted binding sites for DSB-120 on the gene fragments shown based on a probability of 1:500 due to the length of the ligand and variable base requirements at either end [9]. †Search for 5′-PuGATCPy and 5′-PuGTACPy, the known DSB-120 binding sites based on NMR and modelling studies [10, 11]. ‡Number of binding sites found on the gene fragment by the DNAstar programme [12]. §Ratio of the total number of sites found on the gene fragment by DNAstar to the number predicted by probability (e.g. "over-representation" or "under-representation" of binding sites on the particular gene). Pu = purine, Py = pyrimidine.

becomes greater with the development of more rationally designed compounds, we can look forward to a considerable enhancement of such effects, with corresponding possible increases in clinically useful therapeutic indices for cancers where a particular oncogene is found to be strongly associated with a disease.

This study shows that analysis of DNA sequence-specific recognition of real gene sequence targets should not be based solely on considerations of the presumed statistically random distribution of binding sites. In general, non-random sequence distribution is not unexpected—the dinucleotide sequence CpG has a well-documented underrepresentation in the human genome. Therefore, the *actual* frequency of occurrence of a particular sequence needs to be examined, as we report here. We suggest then, that elevated occurrences of sequences selective for particular DNA recognising drugs, may confer therapeutic advantage. Conversely, reduced or zero occurrence within a particular oncogene suggests that such a drug may not be effective in tumours where this oncogene is important for tumour growth and progression.

1. Hurley LH. DNA and associated targets for drug design. *J Med Chem* 1989, **32**, 2027–2033.
2. Hartley JA, Lown JW, Mattes WB, Kohn KW. DNA sequence specificity of antitumour agents. Oncogenes as possible targets for cancer therapy. *Acta Oncol* 1988, **27**, 503–510.
3. Ralph RK, Judd W, Pommier Y, Kohn KW. DNA topoisomerases. In Neidle S, Waring MJ, eds. *Molecular Aspects of Anticancer Drug-DNA Interactions*, Vol. 2. London, Macmillan Press, 1994, 1–95.
4. Vile RG (ed). *Introduction to the Molecular Genetics of Cancer*. Chichester, John Wiley, 1992.
5. Simons M, Edelman ER, DeKeyser J-L, Langer R, Rosenberg RD. Antisense c-*myb* oligonucleotides inhibit intimal arterial smooth muscle cell accumulation *in vitro*. *Nature* 1992, **359**, 67–70.
6. Cooney M, Czernuszewicz G, Postel EH, Flint SJ, Hogan ME. Site-specific oligonucleotide binding represses transcription of the human c-*myc* gene *in vitro*. *Science* 1988, **241**, 456–459.
7. Dwyer TJ, Geierstanger BH, Bathini Y, Lown JW, Wemmer DE. Design and binding of a distamycin A analog to d(CGCAAGTTGGC)·d(GCCAACTTGCG): synthesis, NMR studies, and implications for the design of sequence-specific minor groove binding oligopeptides. *J Am Chem Soc* 1992, **114**, 5911–5919.
8. Mrksich M, Wade WS, Dwyer TJ, Geierstanger BH, Wemmer DE, Dervan PB. Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc Natl Acad Sci USA* 1992, **89**, 7586–7590.
9. Laughton CA, Jenkins TC, Fox KR, Neidle S. Interaction of berenil with the tyrT DNA sequence studied by footprinting and molecular modelling. Implications for the design of sequence-specific DNA recognition agents. *Nucleic Acids Res* 1990, **18**, 4479–4488.
10. Dervan PB. Design of sequence-specific DNA-binding molecules. *Science* 1986, **232**, 464–471; Hélène C, Toulmé J-J. In *Oligodeoxynucleotides, Antisense Inhibitors of Gene Expression* (Cohen JS, ed.). London, Macmillan Press, 1989.
11. Bose DS, Thompson AS, Ching J, *et al.* Rational design of a highly efficient non-reversible DNA interstrand cross-linking agent based on the pyrrolobenzodiazepine ring system. *J Am Chem Soc* 1992, **114**, 4939–4941.
12. Bose DS, Thompson AS, Smellie M, *et al.* Effect of linker length on DNA-binding affinity, cross-linking efficiency and cytotoxicity of C8-linked pyrrolobenzodiazepine dimers. *J Chem Soc Chem Commun* 1992, 1518–1592.
13. DNAStar software. DNASTAR, Inc. Madison, WI, U.S.A.
14. Li LH, DeKoning TF, Kelly RC, *et al.* Cytotoxicity and antitumour activity of carzelesin, a prodrug cyclopropylpyrroloindole analogue. *Cancer Res* 1992, **52**, 4904–4913.

Acknowledgements—We are grateful to the Cancer Research Campaign and SERC for support (grants to DET and SN).