Indian Journal of Chemistry
Vol. 64, March 2025, pp. 269-284
DOI: 10.56042/ijc.v64i3.16546

Advances in Contemporary Research

DNA binding molecules

P K Harrishnarayanan, Shreya Banerjee & Erode N Prabhakaran*
Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012, Karnataka, India
E-mail: eprabhak@iisc.ac.in

Received 20 January 2025; accepted(revised) 27 February 2025

Interactions between DNA and DNA binding molecules govern the life of cells that are the building blocks of all living organisms. Differences in gene expression form the basis for why and how different cells with diverse functions are found. Expression of undesired genes can lead to cancer, diabetes, cardiovascular diseases, immunodeficiency and a number of birth defects. This review primarily discusses about molecules like drugs and transcription factor domains that affect gene expression, different moieties (HNCCH, enediynes, strained rings, flat intercalating rings) with which such molecules bind to specific regions of DNA and synthetic analogues that have been produced from the design of parent scaffolds. Understanding different mechanisms with which molecules bind to DNA allows the design of novel molecules that can bind to any given sequence of DNA and show desired activity after binding to DNA.

Keywords: DNA binding molecules, Transcription factors, Gene expression, Intercalation, Strand cleavage, Sequence selectivity

Molecular biologists estimate that there are around 300,000 genes[1] or segments of information contained in 23 pairs of chromosomes. Each gene, typically is the blueprint for one protein[2,3] and it is the 3-dimensional shape of proteins that give our bodies form[4] and allows the proteins to carry out their functions[5] for the complex chemistry of life. Gene expression is the product of an extensive array of interactions among DNA, RNA, and proteins[6]. The subtle and complex interactions among these molecules determine the difference between health and disease[7].

Transcription of DNA leads to the synthesis of RNA. Translation of RNA leads to the production of a variety of products. The naturally found transcription factors are proteins that are classified as 3 superclasses based on the similarities found in their DNA binding domains[8,9]: helix turn helix, basic and zinc coordinated finger.

The helix turn helix superfamily is characterized by the presence of 2 bundled α-helices that are connected by another shorter α-helical strand[10,11]. One of these bundled α-helices acts as a recognition helix and inserts itself into the major groove of DNA, forming the DNA-protein interface[12]. The basic superfamily consists of the helix-loop-helix and leucine zipper that dimerize and bind to DNA with a scissor-like grip[13]. The members of this superfamily recognize palindromic DNA sequences due to their dimerization[14]. Zinc fingers consist of a coordinated zinc ion and finger-like structures that exhibit independent function in DNA recognition (discussed in section 2c). Although different types of transcription factors show different kinds of interactions with DNA, the fundamental goal of controlling RNA production and gene expression remains the same.

Transcription begins with the binding of transcription factor with DNA in the promoter region[15]. Signalling molecules, like in the nuclear receptor family may play a role in activating and inducing binding of transcription factor with DNA[16,17]. The transcription factor recruits the RNA polymerase enzyme that unwinds the segment of DNA[18]. The polymerase enzyme acts like a double zipper moving along the template strand opening closed chains in the front and closing the opened chains left behind[19]. As the polymerase reaches the terminator sequence, the transcriptional machinery and the RNA strand synthesized detaches marking termination of transcription[20]. The DNA returns to the original state.

Apart from natural transcription factors, there are various proteins, small molecules, drugs having therapeutic potential, nucleic acid strands and staining molecules that bind to DNA. Such molecules

may bear the following moieties that aid in DNA binding: HNCCH moiety, enediyne moiety, strained spirocyclic cyclopropane moiety, fused heterocyclic rings, *etc*. DNA binding molecules also show variety of binding modes such as major groove binding, minor groove binding, intercalating between nucleotide bases and alkylation with the bases. This review focuses on DNA binding molecules that bind to A, B and Z forms of DNA.

Understanding such interactions is one of the first steps towards understanding fundamental machinery of human existence and to gain a basic understanding of why certain individuals are susceptible to genetic disorders like cancer, heart disease, and mental illness. The sequence-specific recognition of DNA is an essential biological process responsible for the regulation of cellular functions including transcription[21], replication[22] and cell division[23]. A variety of small molecules and drugs having antibiotic[24], antiviral[25], or antitumor[26] properties bind to DNA to form complexes and show actions like strand cleavage. The formation of nucleic-acid drug complexes produces profound pharmacological effects by interfering with biological processes in which nucleic acids participate. Some DNA binding molecules interfere with transcription of the target gene by inhibition of key transcription factors in the promoter region or alternatively by a steric blockade in the coding region[27]. The goal is to take such small molecules that bind DNA in a modest fashion and improve them to new and novel specificities.

Different Classes of DNA binding molecules

Several DNA binding molecules have been identified and have been segregated into classes based on their mechanism of DNA binding. Inspired by the known DNA binding molecules and their mechanism of binding, many research groups have synthesized different molecules. This review is about how different classes of molecules show DNA binding by fundamentally different mechanisms and the form of DNA they bind to.

1) Chemically modifying DNA binders

DNA binding molecules can bind to DNA through the formation of reversible covalent bonds. Some molecules may bind to DNA through favorable atomic polarization interactions, hydrogen bonds[28,29], van der Waals interactions[30,31] and show their pharmacological activity. These classes of binders modify the chemical structure of DNA at the site of binding.

1a) Binding due to Alkylation by Spirocyclic Cyclopropane moiety

The spirocyclic cyclopropane moiety aids in sequence selective alkylation in 'AT' – rich minor groove sites of double stranded DNA by reversible, stereoelectronically controlled Adenine N3 addition to the least substituted cyclopropane carbon (Fig. 1). This moiety together with the indole subunit bearing it exists in a chair like conformation, where the moiety is in a plane perpendicular to that of the indole subunit. The cyclopropane ring is highly strained as the 3 sp^3 hybridised carbons are forced to form a triangular unit of each vertex angle nearly twice less than the actual tetrahedral angle. Hence such a moiety is highly susceptible to ring opening[32-37]. This alkylation is feasible due to the proximity of N-3 atom of Adenine to spirocyclic cyclopropane. This process is reversible as the methyl group bridging between Adenine and indole subunit 'kicks back' to form the spirocyclic cyclopropane. Another possibility is of a potential nucleophile that could attack from an opposite side to remove DNA alkylation. This aspect of reversibility provides a lifetime for the molecule bearing the spirocyclic cyclopropane to exist near the 'AT' rich regions of the minor groove.

The known DNA binding molecules bearing the spirocyclic cyclopropane ring are CC-1065 and family of Duocarmycins[37] (Fig. 2). CC-1065 is an antitumor agent that consists of left-hand pyrrolo[3,2-e] indole subunit containing an unprecedented spirocyclic cyclopropane and highly functionalized c-AMP phosphodiesterase inhibitors PDE-I and PDE-II as identical right-hand and central subunits[38-40]. CC-1065 binds to the B form of DNA[41] and it appears to require deformation in the DNA structure at the vicinity of adduct[42]. Studies of CC-1065 binding with Z-form of DNA revealed that in 3.5 M NaCl concentration,CC-1065 converts the Z form to the B form[43]. This molecule blocks progression of cells at G_2 and M stages of the cell cycle, thus inhibiting DNA synthesis[44]. Another exceptionally potent antitumor agent containing the spirocyclic cyclopropane moiety is the family of

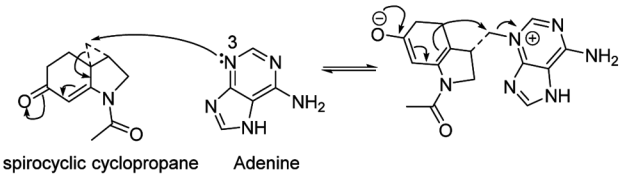

Fig. 1 — Mechanism of reversible alkylation of spirocyclic cyclopropane with N-3 atom of Adenine.

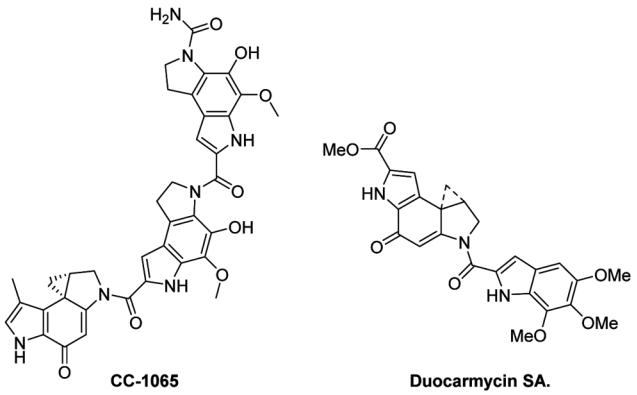

Fig. 2 — Structures of: CC-1065 and Duocarmycin SA.

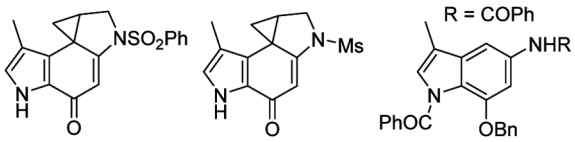

Fig. 3 — Synthetic analogues of PDE subunit of CC-1065.

Duocarmycins: Duocarmycins A, B_1 and B_2, C_1 and C_2, and D_1 and D_2. The newest and most exciting member of this family was Duocarmycin SA which is the most stable and potent member of this family of agents[45-47] with improved chemical and biological properties compared to its predecessors. Duocarmycins bind to the minor groove of B-form DNA with 'AT' rich sites[48]. They act on both dividing and non-dividing cancer cells. It is also observed that cancer cells showing hypoxic, chemoresistant, and stemness phenotypes are sensitive to Duocarmycins[49].

Boger, *et al.* and other groups had synthesized the key partial fragments of the DNA binding agents to examine the natural products and investigated the correlation between structure and properties[50-54]. A group of natural products were unusually made biosynthetically from a key common intermediate in a divergent manner. Boger termed it as 'Divergent Total Synthesis' (DTS)[55]. The key partial fragments of CC-1065 and Duocarmycin are the PDE units (Fig. 3) and the indole units bearing the spirocyclic cyclopropane.

It should be noted that Duocarmycin and CC-1065 share the same spirocyclic cyclopropane unit that shows pharmacological activity of DNA cleavage in tumorous cells. Boger, *et al.* had synthesized analogous spirocyclic cyclopropane units, that were improvised versions of the predecessors. *N*-BOC-CI was a synthetic analogue that showed more biological potency than the natural unit[56]. CBI was the improvised version of *N*-BOC-CI in terms

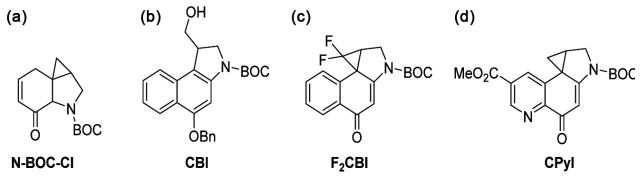

Fig. 4 — Synthetic analogues of spirocyclic cyclopropanes: a) 1,2,7,7a-Tetrahydrocyclopropa[c]indol-4-one (N-BOC-CI), b) 1,2,9,9a-Tetrahydrocyclopropa[c]benzo[e]indol-4-one (CBI), c) 9,9-Difluoro-1,2,9,9a-tetrahydrocyclopropa[c]benzo[e]indol-4-one (F2CBI), d) 1,2,9,9a-Tetrahydrocyclopropa[c]pyrido[3,2-e]indol-4-one (CPyl).

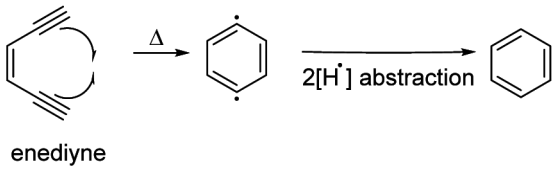

Fig. 5 — An enediyne undergoing Bergman cyclization reaction followed by hydrogen abstraction.

of biological potency and chemical stability[57]. F_2CBI was the first sythetic analogue to have functionalised spirocyclic cyclopropane moiety[58]. It was found to be 500 times more reactive than CBI, a result of the increased strain introduced by the difluorocyclopropane substitution.*N*-BOC-CPyl is a synthetic analogue with a modified heteroaromatic core. The substitution of a fused pyridine for the fused pyrrole found in Duocarmycin SA with ring expansion from a 5-membered to a 6-membered ring was expected to produce one of the more potent, stable, and selective alkylation subunits (Fig. 4).

1b) Glycon assisted delivery of Enediynes

Enediyne is a moiety that is characterized by the presence of 2 triple bonds connected by a double bond (Fig. 5). These moieties when provided temperatures around 200°C, undergo Bergman cyclization to form 1,4-benzenediyl diradical - a highly reactive species, that can abstract hydrogen atoms from donor to give the corresponding arenes[59] (Fig. 5).

K. C. Nicolaou and co-workers have worked with DNA binding molecules that bear enediyne moiety. Dynemicin A, Neocarzinostatin, Esperamicin (discussed further in section 1c) and Calicheamicin were the known DNA binding molecules bearing the enediyne moiety[60]. Calicheamicin binds to the B form of DNA[61]. These anticancer antibiotics cause double stranded DNA breaks leading to cell cycle arrest, eventually killing the cancerous cell[62]. Calicheamicin comprises of 2 distinct aryloligosaccharide and aglycon fragments[63,64]

(Fig. 6). The former unit consists of an extended sugar residue formed from four monosaccharide units and, one hexasubstituted benzene ring connected *via* a series of glycosidic, thioester, and hydroxylamine linkages. This aryloligosaccharide is responsible for: i) delivering the molecule to its target site, ii) tight binding to the minor groove of double helical DNA and iii) displaying high specificity for 5'-TCCT-3' and 5'-TTTT-3′ sequences[65,66]. Molecular modelling by Schreiber and co-workers[67] indicated that a favourable interaction between the large polarizable iodo substituent of the hexasubstituted aromatic ring and the exocyclic amino substituents of the two Guanines in the 3'-AGGA-5' significantly contribute to the sequence selectivity for 5'-TCCT-3' [68].

The latter unit of the molecule is the aglycon fragment which acts as the 'warhead' of the process. It consists of a rigid 10-membered bridged ring with enediyne functionality, awaiting activation to undergo the Bergman cyclization reaction to unleash its impact. The trisulfide group of the aglycon serves as a trigger point for activation by the following mechanism. A nucleophile (*e.g.*, glutathione) can attack the central sulphur atom of the trisulfide group to form a thiolate or a thiol which would add intramolecularly to the adjacent α, β-unsaturated ketone causing the conversion of a trigonal bridgehead position to a tetragonal centre, leading to significant change in structural geometry. This structural conversion brings immense strain on the 10 membered ring and is subsequently relieved by Bergman reaction of the enediyne to generate a highly reactive benzenoid diradical. The calicheamicin diradical generated abstracts hydrogens from duplex DNA at the C-5' in 5'-TCCT-3' segment and the C-4' of the nucleotide, three base pairs away on the 3' side of the complementary strand, causing cleavage of both strands of DNA[69].

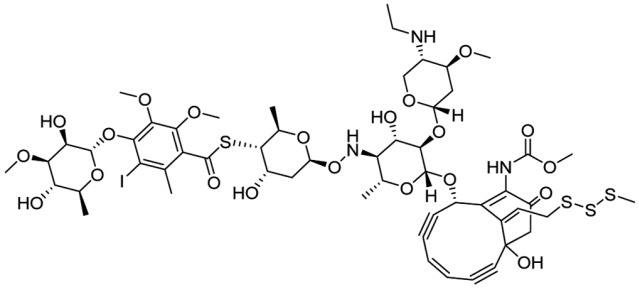

Fig. 6 — Structure of Calicheamicin.

Nicolaou and co-workers made an attempt to mimic the DNA cleaving action of Calicheamicin using simple systems[70]. However, without appropriate target delivery systems the designed enediynes show significantly lower potencies as DNA binding and cleaving agents. The strategy they adopted involved coupling of aglycon[71] (enediyne housing fragment) and oligosaccharide fragments in final stages of synthesis (Fig. 7). Another strategy adopted by Nicolaou's group was to tether the enediyne fragment with 'HNCCH' binding moiety[72] (discussed in section 2a).

1c) Chemically modifying DNA binding Intercalators

Intercalating agents are a large family of DNA binding molecules with considerable diversity in terms of structure. Their mode of DNA recognition is based on intercalation, a mechanism that neither allows DNA binding affinity nor selectivity[73]. These agents are often found sandwiched between base pairs by weak hydrophobic and π-π interactions hence they can readily be released out by random thermal motion of DNA.

A natural product that shows intercalation in the minor groove region is Bleomycin. Bleomycin is selectively toxic to cells in the M and G_2 phases of the cell cycle, and generally more effective against

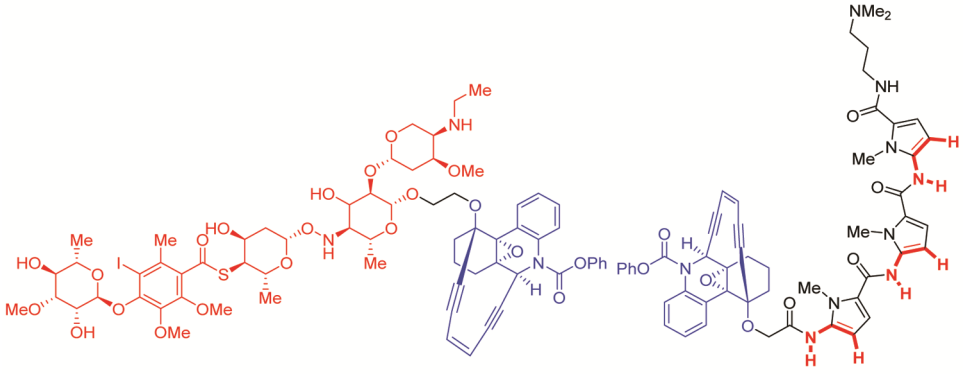

Fig. 7 — Structure of an enediyne bearing molecule (blue) tethered to a glycon (red, left) and to a 'HNCCH' containing moiety (black/red, right).

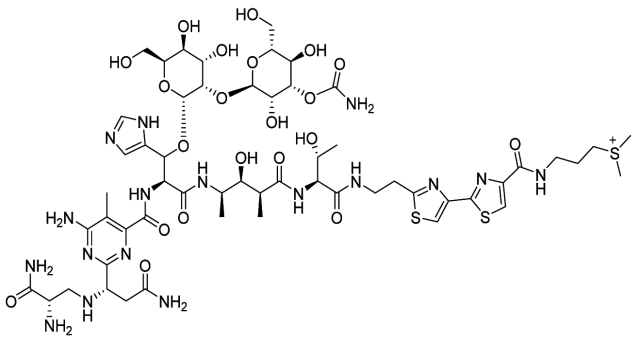

Fig. 8 — Structure of Bleomycin.

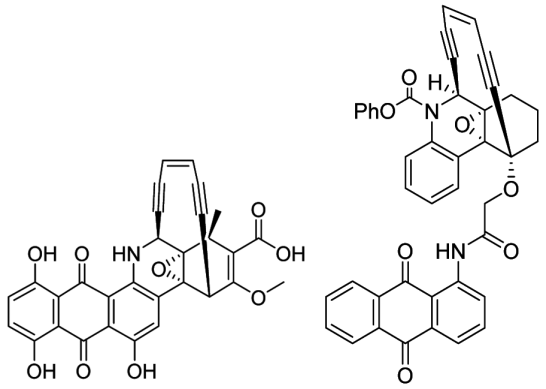

Fig. 9 — Structure of Dynemicin A (left) and Dynemicin A fragment tethered to anthraquinone (right).

actively dividing rather than resting cells. Thus, the cytokinetic state of the tumour cell population is an important determinant of drug activity[74]. Under appropriate salt concentrations in the medium, it is found that Bleomycin-metal complex reacts with the Guanine residues of B form of DNA[75-80] and at4.5 M concentration NaCl it reacts with Z-DNA[75,81]. The DNA binding affinity of this molecule arises from the C terminus sulfonium salt and the bithiazole unit is responsible for intercalation. The valerate – threonine subunits of this molecule direct binding in the minor groove[82] (Fig. 8). The bithiazole unit acts like a swivel point for 180° rotation, allowing the molecule to associate with either strand of double helical DNA from a single intercalation site, indicating the possibility of double-strand DNA cleavage. The N terminal chromophore chelates metal ions (Fe^{+3}, Cu^{+2}) which leads to the activation of O_2, and produces a powerful oxidant. The 4-aminopyrimidine anchors this metal-bound oxidant proximal to DNA for H-atom abstraction, by forming key triplex-like H bonds with G at the 5′-GC and 5′-GT cleavage sites[83-85].

Some molecules bearing the enediyne moiety bind to DNA by intercalation and show their DNA cleaving activity. Intercalating subunits of molecules provide proximity for the diradical species formed from enediyne moiety to abstract hydrogen atoms from the nucleotide bases causing disruption of hydrogen bonds and subsequent DNA cleavage. A mechanism of antitumor activity of Dynemicin A has been proposed which highlights the importance of its anthraquinone subunit (Fig. 9). In this mechanism, the anthraquinone subunit intercalates into the DNA and undergoes bioreduction, facilitating opening of the epoxide. This causes a significant conformational change in the molecule and introduces considerable strain into the enediyne system which is relieved by the new system undergoing the Bergman reaction, thus generating the DNA-damaging diradical species[86,87]. Dynemicin A binds to B form of DNA and does not react with Z form of DNA, but shows hyperreaction with B form - Z form junctions of DNA[88,89]. Nicolaou and co-workers have designed simple systems that mimic the enediyne activity of Dynemicin A and have tethered it to anthraquinone subunit for delivery[71] (Fig. 9).

Esperamicins form a family of highly potent compounds which show broad spectrum antimicrobial and antitumor activity in murine systems. This class of compounds are characterized by a central core containing several unique features (Fig. 10), including i) a bicyclo[7.3.1] ring system, ii) an allylic trisulfide attached to the bridging atom, iii) a 1,5-diyn-3-ene as part of the ring system, and iv) an α, β- unsaturated ketone in which the double bond is at the bridgehead of the bicyclic system. The ring system is attached at one end by a trisaccharide moiety consisting of a hydroxyamino sugar (ring A) connected to an isopropylamino sugar (ring C) through a glycosidic linkage and a thiomethyl sugar (ring B) through a novel NH-O linkage[90] (Fig. 10). The opposite end of the ring system connects to a 2-deoxy-L-fucose (ring D)- anthranilate (ring E) moiety. Molecular Dynamics studies of the Esperamicin A_1- d(C-G-G-A-T-C-C-G) duplex DNA Complex revealed that Esperamycin A_1 binds to the DNA at the minor groove *via* the A-B-C trisaccharide moieties with its methoxyacrylyl-anthranilate moiety intercalating into the helix. This anchors the enediyne rigidly in the minor groove such that the pro-radical centres of the enediyne are in close proximity to their potential proton abstraction sites. Esperamicin binds to the B form of DNA[91].

Chemical hydrolysis[92] has been used to synthesize various analogues of espA. These include: i) espC,

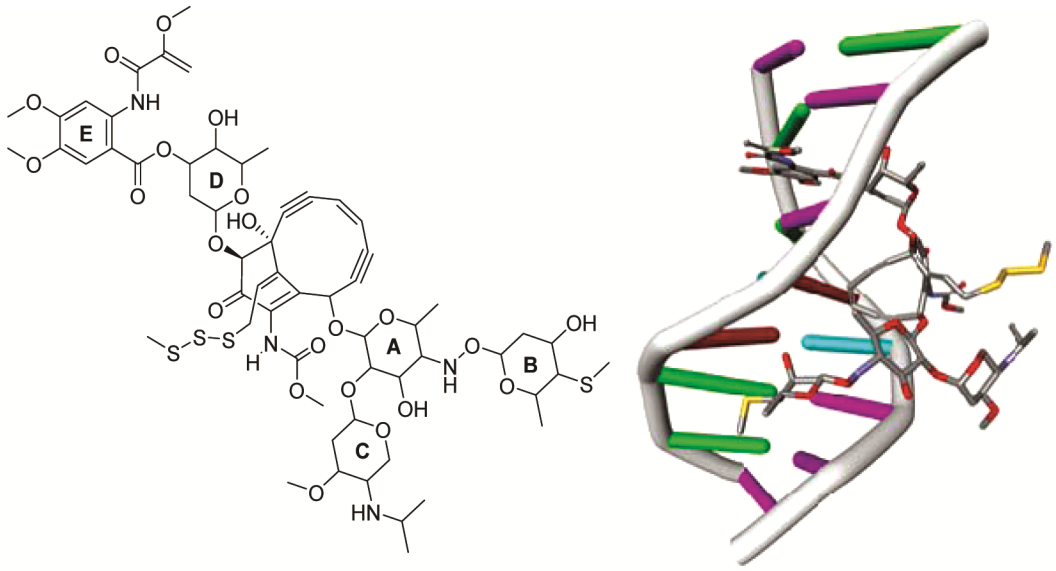

Fig. 10 — Structure of Esperamicin A_1 (left) and Esperamicin A_1/DNA complex (PDB code: 1PIK, rendered using Discovery Studio, right).

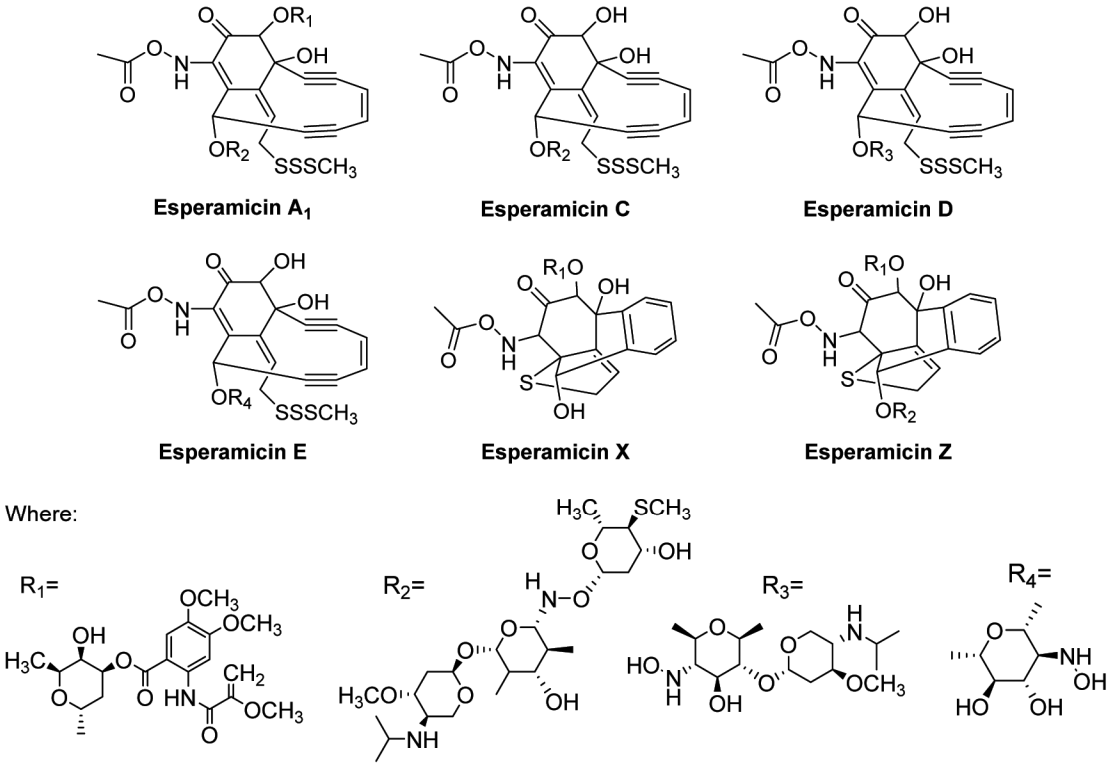

Fig. 11 — Structures of Esperamicin analogues.

which lacks both the 2-deoxy-L-fucose and the aromatic ring moieties, ii) espD, which is similar to espC but in addition, lacks the thiomethyl-hexopyranose moiety, iii) espE, which consists of the bicyclic core and the hydroxylamino sugar, iv) espZ, which is an esperamicin A_1 that has undergone reductive cyclization of the trisulfide and subsequent aromatization of the 1,5-diyn-3-ene structure, and v) espX, which is analogous to espZ but lacks the trisaccharide at the C-12 position (Fig. 11).

The first of the bicyclic enediyne antibiotics to be discovered was Neocarzinostatin (NCS). The

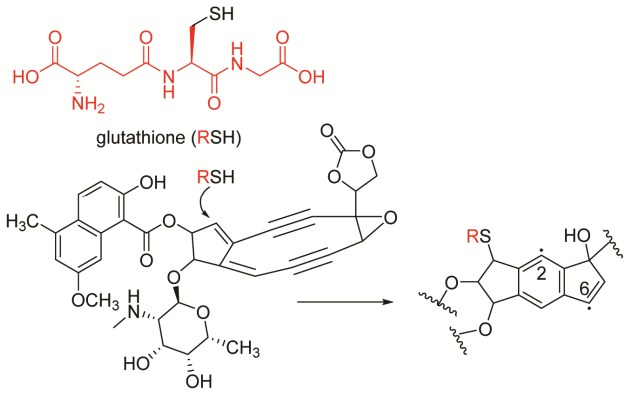

Fig. 12 — Structure of Neocarzinostatin and activation by glutathione (denoted as RSH).

moleculeconsists of 3 subunits: i) a 5- methyl-7-methoxynapthoate, ii) a 2,6-dideoxy-2- (methylamino) galactose moiety, and iii) an interconnecting C_{12} subunit, consisting of a highly strained bicyclic [7.3.0] dodecadienediyne system with a cyclic carbonate and an epoxide[93] (Fig. 12). NCS binds to the minor groove of B-DNA by intercalation[94]. The molecule binds to DNA in a 2-step process. The molecule comes in proximity to DNA by intercalation of naphthoate moiety and electrostatic interaction of the positively charged amino sugar moiety with the negatively charged sugar phosphate backbone of the DNA. Hydrodynamic, NMR, and deuterium substitution studies have demonstrated that, at AGC GCT sequences the naphthoate intercalates between the adjacent AT and GC base pairs, while the diradical is strategically positioned in the minor groove such that the radical at C-2 of the chromophore abstracts hydrogen from C-1′ of Cytosine in AGC DNA sequence, while the radical at C-6 abstracts from C-5′ of Thymine in GCT. Abstraction of hydrogen causes strand breaks[95]. The above mechanism is triggered by addition of sulfhydrils, with glutathione being the most likely *in vivo* cofactor (Fig. 12).

2) Chemically non modifying DNA binders

Some DNA binding molecules exhibit their pharmacological activity without causing any chemical change in the site where they bind[96]. These classes of molecules compete with transcription factors (TFs) whose binding with DNA can cause undesired or disease-causing gene expressions.

2a) Molecules with 'HNCCH' binding moiety

The class of molecules containing 'HNCCH' moiety as shown in Fig. 13 bind to the 'AT' rich regions of the minor groove of DNA by forming hydrogen bonds between hydrogen atom bonded to

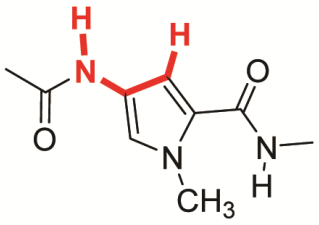

Fig. 13 — The 'HNCCH' (highlighted in red) DNA-binding moiety in N-methylpyrrolecarboxamide (N-MPC) group.

nitrogen atom of the 'HNCCH' moiety and Adenine N-3 atoms or Thymine O-2 atoms. The binding specificity to the 'AT' rich sequences arises from the close van der Waals contacts between Adenine C-2 hydrogens and CH groups of the 'HNCCH' binding moiety[97].

The known molecules that bind with DNA by this binding moiety are: Netropsin, Distamycin and Hoecht-33258 (Fig. 14). These molecules are crescent (or) concave shaped and match the concave shape of the minor groove in B-DNA[98]. Netropsin and Distamycin are natural molecules that cause protracted cell growth consisting of a prolongation of the G_1 phase of the cell cycle along with arrest in the G_2 compartment[99]. By binding to DNA both molecules block activity of certain topoisomerases that are important for transcription[100]. Netropsin binds to B form of DNA and does not bind to A or the Z form. It is also observed that binding of Netropsin to DNA favours the A to B and Z to B helix transformations[101-103]. Distamycin binds to B-DNA[104]. Netropsin and Distamycin contain 2 and 3 N-methylpyrrolecarboxamide (N-MPC) groups respectively that bear the 'HNCCH' moiety[105]. Hoecht-33258, a synthetic dye, has benzimidazole groups that bear the 'HNCCH' moiety[106,107]. This molecule binds to the 'AATT' base pairs in the minor groove of B-DNA[108]. It is shown that the inhibitory effects of Hoechst-33258 arise from the stabilization of the double-stranded organization of DNA such that the unwinding of DNA at the replication or transcription origin sites becomes greatly impaired[109].

For reading large sequences of double stranded DNA, sequence specific DNA binding molecules can be coupled with DNA binding units derived from natural products. Dervan and his co-workers have designed crescent shaped molecules with 'HNCCH' containing groups that are coupled with such units[110,111]. Perhaps, it was Dervan who pioneered synthesizing DNA binding molecules aiming to bind to any given DNA sequence. Khorlin and co-workers constructed bis (EDTA-distamycin)fumaramide (BEDF)

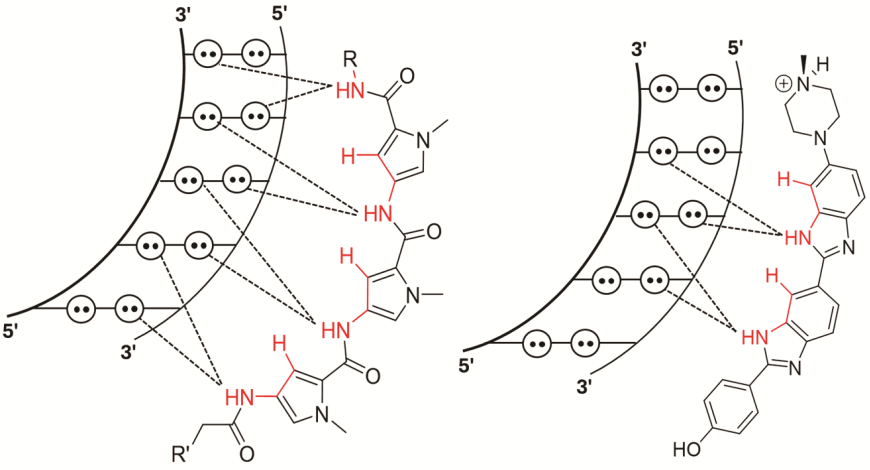

Fig. 14 — (Left) DNA-binding interactions of Netropsin and Distamycin to AT rich sequences of minor groove of DNA. (Right) DNA-binding interactions of Hoechst-33258 to AT rich sequences of minor groove of DNA. Circles with two dots represent lone pairs of electrons on N-3 of Adenine and O-2 of Thymine. The dotted lines are bridged hydrogen bonds to the NHs of the 'HNCCH' moiety (highlighted in red).

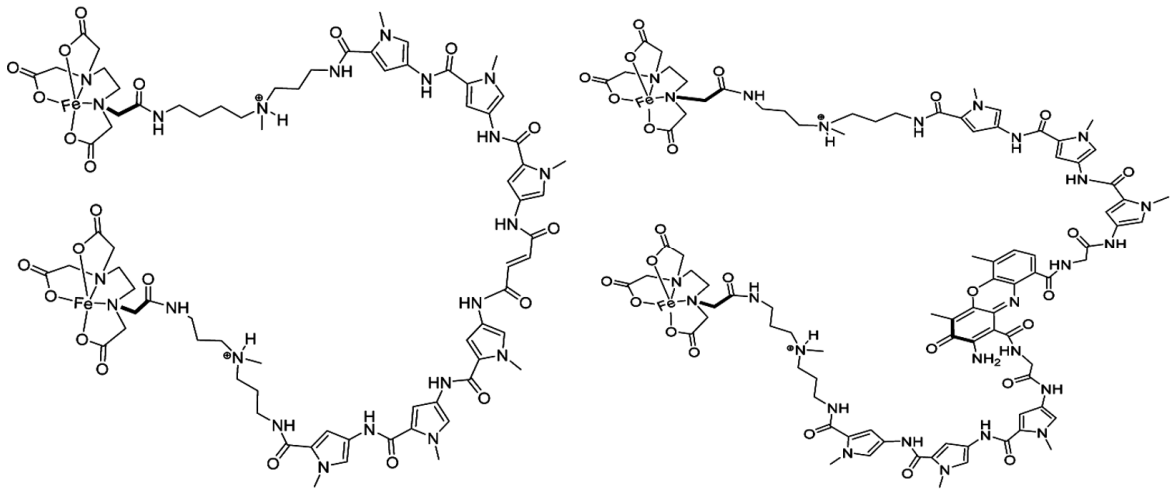

Fig. 15 — (Left) Bis [Fe (II) - EDTA-distamycin]fumaramide (BEDF). (Right) Bis [Fe (II)-EDTA – distamycin]phenoxazone (BEDP).

(Fig. 15) which is a crescent shaped octamide, with two N-MPC tripeptide units coupled by the fumaramide linker[112]. This molecule was later studied by Dervan[113]. The fumaramide linker was chosen as it mimics the N-MPC binding units in shape and curvature between NHs of amide bonds. Methidiumpropyl - EDTA with 'Fe' metal centre was tethered to molecules to find the DNA sequence that they bind to[114]. BEDF showed binding to 'AT' rich sequences 5'-ATTTTTATA-3' and 5'-ATAATAAT-3'. Another molecule was synthesized by replacing the two cyclic pentapeptides of Actinomycin with tripeptides from Distamycin. Actinomycin (discussed in section 2b) is a natural molecule that consists of a phenoxazone, an aromatic chromophore and 2 identical cyclic pentapeptide lactones on either side of the phenoxazone unit. Actinomycin behaves as an intercalator using the phenoxazone unit and binds to 4 base pairs with specificity for 5'-NGCN-3' (where 'N' is any other nucleotide) sequences[115]. The Distamycin tripeptide was connected to a phenoxazone by a glycine tether formingbis(distamycin)phenoxazone (BEDP, Fig. 15) which is a potential groove binder-intercalator-groove binder that may bind to 10 bp of DNA having the sequence $(AT)_4GC_2(AT)_4$[98,116].

2b) Chemically non-modifying DNA binding Intercalators without C_2 Symmetry

Actinomycin is a well-known natural transcription inhibitor that has been widely reported to induce cell

apoptosis in several types of tumour cells by inhibiting the anti-apoptotic gene transcriptions[117]. Actinomycin shows binding to both B and Z forms of DNA depending on the concentration of the buffer system[118,119]. At 4.4 M NaCl buffer, Z-DNA is switched to B-DNA at each of the bases where Actinomycin is bound to[118]. Actinomycin has a structure constituted by phenoxazone which is coupled to two identical cyclic pentapeptide lactones (Fig. 16). The cyclic pentapeptides fit snugly above and below the phenoxazone intercalating ring in the minor groove of DNA. X-ray crystal structures of 1:2 complexes of actinomycin with d(GpC) and dG segments reveal intercalation interactions[120] and specific hydrogen bonds arising from the 2-amino group of Guanine to the carbonyl oxygen in L-threonine residue as well as two hydrogen bonds between the D-valine residues in the peptide rings[121-124]. It is evident that this molecule prefers to bind to 'GC' rich sequences in the minor groove of DNA.

There are intercalating molecules that aid in visualizing DNA and do not have any biological action of their own at optimal concentrations[125]. Such molecules can easily be replaced by other molecules with better binding affinity[126], therefore these molecules are employed as 'reporters' to find out extent of binding of other DNA binding molecules. Ethidium bromide (EtBr, Fig. 17) binds to B-DNA[127] and Z-DNA[128] formed from dinuclear bis(platinum) complex induced transition of B-DNA[129]. DNA initially saturated with EtBr can be exposed to competitive DNA binders like proteins and other ligand molecules. EtBr will then be released upon competitive binding, hence the released amount of EtBr can be correlated to the binding affinity of the DNA binding ligand. Techniques like fluoroscence emission spectroscopy employ EtBr as a reporter molecule[130]. EtBr is used as a stain for visualizing DNA through exposure in UV light[131]. Other intercalators that aid in DNA visualization are DAPI (4-6-Diamidino-2-phenylindole)[132] and Acridine Orange[133] (Fig. 17). DAPI binds to B-DNA,but does not show high binding affinity with Z-DNA[134]. Acridine Orange also binds to B form ofDNA[135].

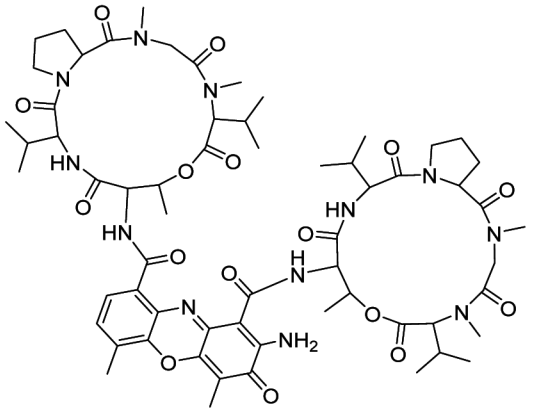

Fig. 16 — Structure of Actinomycin.

2b) Chemically non-modifying DNA binding Intercalators with C_2-Symmetry

Tröger's base (TB) is a C_2 symmetric, chiral molecule whose history began in 1887, when Carl Julius Ludwig Tröger obtained an unexpected product during the reaction of *p*-toluidine and methylal in aqueous HCl[136,137]. TB consists of a bicyclic aliphatic unit fused with two aromatic rings. The central methanodiazocine unit forces the aromatic rings to be nearly perpendicular to each other, making TB a rather rigid V-shaped molecule with a hydrophobic cavity (Fig. 18). Due to its geometry, the R and S enantiomers of TB could interact with DNA in both possibilities: by intercalation and by groove binding in different modes. First, one of the acridine rings is intercalated between adjacent base pairs, the other acridine being bound either in the major or in the minor groove. The second mode involves binding of both acridine moieties in either major or minor grooves, *i.e.*, no binding through intercalation[138]. Interaction between the S and R enantiomers of TB with different DNA sequences are as summarized: eight decamers, including two mononucleotidic sequences[139]: $d(A)_{10}.d(T)_{10}$, $d(C)_{10}.d(G)_{10}$ and six dinucleotidic sequences : $(dA\text{-}dT)_5.(dA\text{-}dT)_5$, (dT-

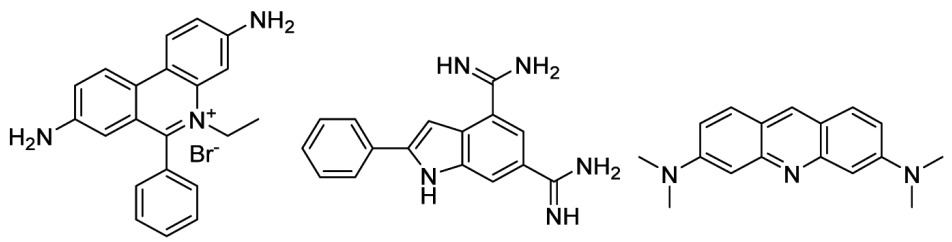

Fig. 17 — Structures of Ethidium bromide (left), 4-6-Diamidino-2-phenylindole (middle) and Acridine orange (right).

$dA)_5$.($dT\text{-}dA)_5$, $(dG\text{-}dC)_5$.$(dG\text{-}dC)_5$, $(dC\text{-}dG)_5$.$(dC\text{-}dG)_5$, $(dG\text{-}dT)_5$.$(dA\text{-}dC)_5$ and $(dT\text{-}dG)_5$.$(dC\text{-}dA)_5$. Analogues of TB that show enantioselective binding with DNA may potentially serve as molecular probes. The anticancer activity of TB is also being explored[140]. TB binds to B form of DNA[141].

Yashima and co-workers introduced the first phenanthroline analogue of TB that exhibited higher affinity towards DNA than the parent 1,10-phenanthroline[142]. Later, Demeunynck and Lhomme studied the phenanthroline analogue of TB[143]. They had synthesized the Proflavine analogue (Fig. 19) which was found to bind preferably to sequences with motifs such as 5′-GTT·AAC-3′ or 5′-ATGA·TCAT-3′[144].

2c) DNA binding zinc finger domains with constrained α-helical conformation

By the late 1950s, molecular biologists had demonstrated that the information encoded in DNA is initially copied into RNA by transcription and then

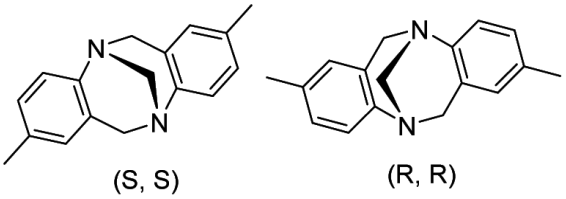

Fig. 18 — Structure of TB and its enantiomers.

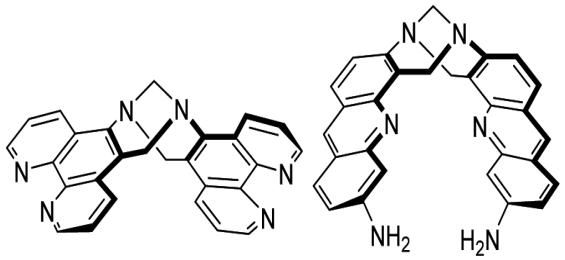

Fig. 19 — Structures of Phenanthroline (left) and Proflavine (right) analogues of Tröger's base (TB).

into proteins followed by translation[145]. There are 4 nucleotide bases and 20 amino acids, hence translation cannot be comprehended for a direct one to one correspondence between a nucleotide of RNA and an amino acid of protein. The sequence of nucleotides of an mRNA molecule is read serially in a non-overlapping manner in groups of three and is termed as a triplet 'codon'. The 64 combinations of triplet codons code for 20 amino acids with each amino acid being coded by more than 1 triplet codon (degenerate)[146] (Fig. 20).

Transcription factors (TFs) contain DNA binding domains (DBD) through which they bind to DNA to carry out gene expression that determines the fate of the cell[147]. Zinc finger proteins (ZFP)[148], leucine zipper motif[149], helix loop helix motif[150], helix turn helix motif[151] are some of the DBD. The zinc-finger domain is one of the most abundantly found DNA-binding motifs found in eukaryotic TFs[152]. Zinc finger proteins consist of an α helix packed against a β hairpin (Fig. 21). 2 cysteine residues in the β sheet and 2 histidine residues in the α helix form coordinate bonds with a zinc ion (Cys2His2 ZFP), giving a stable globular structure[153] (Fig. 21). DNA binding occurs through amino acid residues in the α-helix at positions -1,2,3,6 (where position 1 means 1st amino acid residue of α-helix). Hence the first 7 residues of the α-helix are vital for DNA binding. –1st position amino acid plays a role in anchoring to DNA, to facilitate DNA binding by 2nd, 3rd and 6th position amino acids[154-157]. From Fig. 21, it can be observed that the DBD of ZFP does not coordinate with the zinc ion. However, without the zinc ion, the ZFP loses stability and structure.

Patterns of gene expression rely on triggers such as intracellular and extracellular signals, hence there is both up and down regulation of transcription of a gene[158-160]. Malfunctioning of transcriptional regulators alters the gene expression and results in various diseases like

codons	GCA	AGA AGG CGA						GGA			UUA UUG CUA				AGC AGU UCA	ACA			GUA	
	GCC	CGC						GGC		AUA	CUC			CCA	UCC	ACC			GUC	UAA
	GCG	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUC	CUG	AAA		CCC	CCG	ACG		UAC	GUG	UAG
	GCU	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUU	CUU	AAG	AUG	UUU	CCU	UCU	ACU	UGG	UAU	GUU	UGA
amino acids	Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
	A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

Fig. 20 — Table listing all the three-nucleotide codons in mRNAs that specify an amino acid (mentioned with 1 letter and 3 letters abbreviation). Reproduced from *B. Alberts, Molecular Biology of the Cell, 4th Ed.*

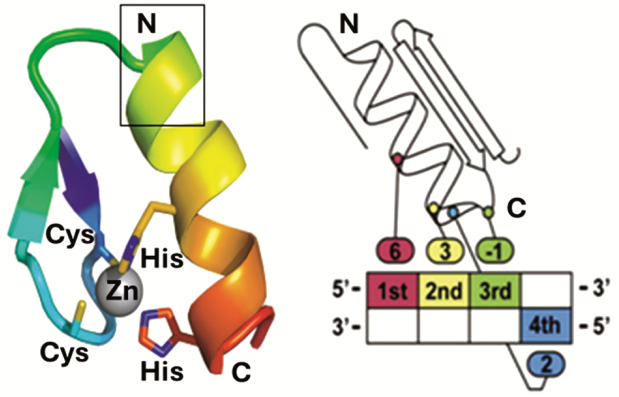

Fig. 21 — (Left) 3-Dimensional structure of the protein Xfin showing zinc finger 31. The grey sphere represents the zinc ion. The boxed segment in the α-helix contains the DBD. Reproduced from *Neuhaus, D., Zinc finger structure determination by NMR: Why zinc fingers can be a handful. Prog. Nucl. Magn. Reson. Spectrosc. 2022, 130, 62-105.*[158] (Right) Cartoon diagram showing the binding of a finger to a dsDNA sequence. Reproduced from *Sera, T., Zinc-finger-based artificial transcription factors and their applications. Adv. Drug Delivery Rev. 2009, 61 (7-8), 513-526.*[157]

diabetes[161], immunodeficiency[162] and cancer[163,164]. Controlling the gene expression through artificial transcription factors (ATFs) has been the goal for several research efforts over the past couple of decades. The general construction of an ATF, involves a nuclear localization signal (NLS: aids in the entry of ATF inside nucleus of cell), DBD (motif that binds to DNA) and an effector domain (ED: controls up or downregulation of the gene in DNA)[157].

Zinc fingers represent the most promising candidates for DBD as they bind to DNA with independently functioning monomeric fingers (modular) and show triplet specificity. With more number of fingers, there is an increase in sequence selectivity[165]. Several groups have engineered ZFPs to modulate endogenous gene expression in cells[166-192].

Single α-helical turns (4 to 6 residues) in shorter peptide sequences (around 15 residues) influence several biological activities[193-195]. Although being the best fit for DBD, constraining the first 7 residues of the zinc fingers in the α-helical conformation poses a challenge[196-198].There are various approaches to constrain the helical turns. The non-covalent techniques include the metal clamp approach[199-201], side chain-side chain electrostatic salt bridge interactions[202], cation-π interactions[203] and π-π stacking[204]. Introduction of unnatural amino acids like α-aminoisobutyryl (Aib) residue promote helicity in α-helices[205-207] and 3_{10} helices[208]. The covalent approaches consist of promoting helicity through covalent cross linking in peptides[209], side chain olefin derivatization[210] using ring closing metathesis and Hydrogen Bond Surrogates[211] (HBS).

HBS technique involves replacement of the labile hydrogen bond by a rigid linker and has been explored by various groups. In 1999, Satterthwait and co-workers pioneered the work on HBS with their hydrazone linker (13_82 structure)[212,213]. Arora, *et al.* had proposed 3 HBS models that were synthesized by ring closing metathesis (RCM): i) olefin model (13_82 structure) with a C-C double bond linker[214,215], ii) disulfide model (13_62 structure) with disulfide linkage[216], and iii) covalent alkane linker model[217]. Alewood, *et al.* had made an attempt to replace the (i, i+4) hydrogen bond with an ethylene bridge to synthesize small truncated peptides (13_63 structure)[218]. Broussy and co-workers developed solid phase synthesis of short peptides by incorporating a diaminopropane linker to replace the (i, i+4) hydrgen bond (13_63 structure)[219-220].

Prabhakaran, *et al.* had proposed their HBS model (13_73 structure) with a propyl linker[221,222]. Highest helicity for short α-helical turns can be derived in the propyl linker model as it retains the planar nitrogen atom in N termini of the i+1st residue[223-225] and the side chain of i+1st residue which forms a stereogenic centre and a crucial recognition element in various interactions[226-235]. CD studies have shown that mimics retaining the planar sp^2 hybridized nitrogen atom of i+1st residue resemble α-helix the closest[221,236]. The propyl linker HBS model could mimic α-helices[222, 237], single α-helical turns[221,238], 3_{10} helices[239] and β-sheets[240] with superior stability. Their model is easy to design as it is very similar to that of the native peptide and it also bears sufficient non-peptidic character such that it is not easily biodegraded. Since each zinc finger recognizes a triplet codon, a library of amino acids sequences of such fingers can be made. Such a library[241,242] can be employed in tailoring constrained α-helical zinc fingers that can bind to any of the 64 triplet codons. Optimization of DNA binding affinity can be achieved by mutations of amino acid sequences in DBD and obtaining them through protein expression in *E. coli* by plasmids[243-248]. Thus, this model can be expected to produce promising molecules in future for DNA binding in a triplet recognition fashion.

Conclusions

Interactions between proteins, nucleic acids and corresponding gene expression determine the fate of the cell. Numerous natural molecules showing antibiotic, antiviral, antitumor activity and transcriptional control are being studied to produce novel entities with higher efficacy. The work summarized above describes various DNA binding molecules and their modes of binding. HBS models have been constantly evolving as different groups have made their attempts to mimic the natural helix the closest. Prabhakaran's HBS model is the closest mimic to the natural helix, as it retains the N termini planarization of $i+1^{st}$ residue and the side chain of α-carbon which is crucial for molecular recognition. Since the model has sufficient non peptidic character, it is resistant to biodegradation. Recently, various drug molecules that have been shown to inhibit the processes in which DNA and RNA participate, have been synthesized and are being approved by the FDA. Such molecules have immense scope for therapeutic potential. Tovorafenib was approved by the FDA in April, 2024 and is prescribed for the treatment of paediatric low-grade glioma. It is a type II pan-RAF inhibitor[249,250]. Similarly, Imetelstat which was approved in June, 2024, is an oligonucleotide telomerase inhibitor[251] and is indicated for the treatment of Myelodysplastic Syndrome. Doxorubicin is one of the prominent drugs for the treatment of breast cancer[252] and it shows DNA binding through intercalation[253]. It is anticipated that chemists will continue to bring fresh perspectives and a zest for understanding biology at the molecular level in this highly fertile ground to combat genetic, oncogenic and viral diseases.

References

1. Fields C, Adams M D, White O & Venter J C, *Nat Genet,* 7 (1994) 345.
2. Henry C, Overbeek R & Stevens R L, *Biotechnol J,* 5 (2010) 695.
3. Szostak J W, Bartel D P & Luisi P L, *Nature,* 409 (2001) 387.
4. Johnson M S, Sutcliffe M J & Blundell T L, *J Mol Evol,* 30 (1990) 43.
5. Orengo C A, Todd A E & Thornton J M, *Curr Opin Struct Biol,* 9 (1999) 374.
6. Lockhart D J & Winzeler E A, *Nature*, 405 (2000) 827.
7. Robinette S L, Holmes E, Nicholson J K & Dumas M E, *Genome Med*, 4 (2012) 1.
8. Luscombe N M, Austin S E, Berman H M & Thornton J M, *Genome Biol,* 1 (2000) 1-37.
9. Hughes T R, *A handbook of transcription factors,* (Springer Science & Business Media) 2011.
10. Menon S & Lawrence C, *Brenner's Encyclopedia of Genetics,* 3 (2013).
11. Brennan R G & Matthews B W, *J Biol Chem,* 264 (1989) 1903.
12. Aravind L, Anantharaman V, Balaji S, Babu M M & Iyer L M, *FEMS Microbiol Rev,* 29 (2005) 231.
13. Vinson C R, Sigler P B & McKnight S L, *Science,* 246 (1989) 911.
14. Amoutzias G D, Robertson D L, Peer V D Y & Oliver S G, *Trends Biochem Sci,* 33 (2008) 220.
15. Buratowski S, *Transcription Sci*, 322 (2008) 1804.
16. Huang P, Chandra V & Rastinejad F, *Chem Rev,* 114 (2014) 233.
17. Carlberg C, *Biofactors,* 10 (1999) 91.
18. Cole A S & Eastoe J E, *Biochemistry and oral biology* Butterworth-Heinemann: 2014.
19. Fong P, *Proc Nati Acad Sci*, 58 (1967) 501.
20. Richardson J P & Roberts J W, *Crit Rev Biochem Mol Biol* 1993 28 (1) 1.
21. Pabo C O & Sauer R T, *Annu Rev Biochem,* 61 (1992) 1053.
22. Bramhill D & Kornberg A, *Cell,* 54 (1988) 915.
23. Stillman B, *FEBS Lett,* 579 (2005) 877.
24. Polyanovsky O, Nosikov V, Zhuze A, Braga E & Karlyshev A, *Adv Enzyme Regul,* 17 (1979) 307.
25. De Clercq E, *J Clin Virol,* 22 (2001) 73.
26. Auclair C, *Arch Biochem Biophys,* 259 (1987) 1.
27. Amine B M, Lambert M & David-Cordonnier M.-H, *Curr Top Med Chem,* 15 (2015) 1323.
28. Coulocheri S A, Pigis D G, Papavassiliou K A & Papavassiliou A G, *Biochimie,* 89 (2007) 1291.
29. Fersht A R, *Trends Biochem Sci,* 12 (1987) 301.
30. Czarny A, Boykin D W, Wood A A, Nunn C M, Neidle S, Zhao M & Wilson W D, *J Am Chem Soc,* 117 (1995) 4716.
31. Chang D-K & Cheng S-F, *Int J Biol Macromol,* 19 (1996) 279.
32. Boger D L & Johnson D S, *Angew Chem Int Ed Engl,* 35 (1996) 1438.
33. Boger D L, Johnson D S & Yun W, *J Am Chem Soc,* 116 (1994) 1635.
34. Boger D L & Yun W, *J Am Chem Soc,* 115 (1993) 9872.
35. Sugiyama H, Ohmori K, Chan K L, Hosoda M, Asai A, Saito H & Saito I, *Tetrahedron Lett,* 34 (1993) 2179.
36. Warpehoski M A, Harper D E, Mitchell M A & Monroe T J, *Biochem*, 31 (1992) 2502.
37. Boger D L, Boyce C W, Garbaccio R M & Goldberg J A, *Chem Rev,* 97 (1997) 787.
38. Chidester C, Krueger W, Mizsak S, Duchamp D & Martin D, *J Am Chem Soc,* 103 (1981) 7629.
39. Enomoto Y, Furutani Y, Naganawa H, Hamada M, Takeuchi T & Umezawa H, *Agric Biol Chem,* 42 (1978) 1331.
40. Nakamura H, Enomoto Y, Takeuchi T, Umezawa H & Iitaka Y, *Agric Biol Chem,* 42 (1978) 1337.
41. Boger D L, Invergo B J, Coleman R S, Zarrinmayeh H, Kitos P A, Thompson S C, Leong T & McLaughlin L W, *Chem Biol Interact,* 73 (1990) 29.
42. Hurley L H, Needham-VanDevanter D R & Lee C.-S, *Proc Nati Acad Sci,* 84 (1987) 6412.
43. Krueger W C & Prairie M D, *Chem -Biol Interact,* 62 (1987) 281.
44. Bhuyan B K, Newell K A, Crampton S L & Von H D D, *Cancer Res,* 42 (1982) 3532.

45. Ichimura M, Ogawa T, Takahashi K.-I, Kobayashi E, Kawamoto I, Yasuzawa T, Takahashi I, Nakano H & Duocarmycin S A, *J Antibiot,* 43 (1990) 1037.
46. Yasuzawa T, Saitoh Y, Ichimura M, Takahashi I & Sano H, *J Antibiot,* 44 (1991) 445.
47. Ichimura M, Ogawa T, Katsumata S, Takahashi K.-i, Takahashi I & Nakano H, *J Antibiot* 44 (1991) 1045.
48. Alder A, Struck N S, Xu M, Johnson J W, Wang W, Pallant D, Cook M A, Rambow J, Lemcke S, Gilberger T W *Cell Chem Biol* 29 (2022) 840.
49. Yao H.-P, Zhao H, Hudson R, Tong X.-M & Wang M-H, *Drug Disc Today,* 26 (2021) 1857.
50. Wierenga W, *J Am Chem Soc,* 103 (1981) 5621.
51. Kraus G A & Yue S, *J Chem Soc Chem Commun,* (1983) 1198.
52. Kraus G A, Yue S & Sy J, *J Org Chem,* 50 (1985) 283.
53. Boger D L & Zarrinmayeh H *J Org Chem,* 55 (1990) 1379.
54. Sundberg R J, Baxter E W, Pitts W J, Ahmed-Schofield R & Nishiguchi T, *J Org Chem,* 53 (1988) 5097.
55. Li L, Chen Z, Zhang X, Jia Y *Chem Rev,* 118 (2018) 3752.
56. Boger D L, Mesini P, Tarby C M *J Am Chem Soc,* 116 (1994) 6461.
57. Boger D L & Mesini P, *J Am Chem Soc,* 116 (1994) 11335.
58. Boger D L & Jenkins T J, *J Am Chem Soc,* 118 (1996) 8860.
59. Bergman R G, *Acc Chem Res,* 6 (1973) 25.
60. Nicolaou K C, Smith A & Yue E, *Proc Natl Acad Sci,* 90 (1993) 5881.
61. Ikemoto N, Kumar R A, Ling T.-T, Ellestad G A, Danishefsky S J & Patel D J, *Proc Nati Acad Sci,* 92 (1995) 10506.
62. Carr M I, Zimmermann A, Chiu L.-Y, Zenke F T, Blaukat A & Vassilev L T, *Front oncol,* 10 (2020)127.
63. Lee M D, Dunne T S, Siegel M M, Chang C C, Morton G O & Borders D B, *J Am Chem Soc,* 109 (1987) 3464.
64. Lee M D, Ellestad G A & Borders D B, *Acc Chem Res,* 24 (1991) 235.
65. Zein N, Poncin M, Nilakantan R, Ellestad G A *Science* 244 (1989) 697.
66. Walker S, Landovitz R, Ding W, Ellestad G A & Kahne D, *Proc Nati Acad Sci U S A,* 89 (1992) 4608.
67. Hawley R C, Kiessling L L & Schreiber S L, *Proc Nati Acad Sci U S A,* 86 (1989) 1105.
68. Nicolaou K, Tsay S, Suzuki T & Joyce G, *J Am Chem Soc* 114 (1992) 7555.
69. Hangeland J J, De Voss J J, Heath J A, Townsend C A, Ding W D, Ashcroft J S & Ellestad G A, *J Am Chem Soc,* 114 (1992) 9200.
70. Nicolaou K, Zuccarello G, Riemer C, Estevez V & Dai W M, *J Am Chem Soc,* 114 (1992) 7360.
71. Nicolaou K, Schreiner E P, Iwabuchi Y & Suzuki T, *Angew Chem Int Ed Engl,* 31 (1992) 340.
72. Nicolaou K & Dai W M, *Angew Chem Int Ed Engl* 30 (1991) 1387.
73. Vázquez M E, Caamaño A M & Mascarenas J, *Chem Soc Rev* 32 (2003) 338.
74. Sikic B, *Cancer Surv* 5 (1986) 81.
75. Morii T, Otsuka M & Sugiura Y, *Biochim Biophys Acta,* 1308 (1996) 169.
76. Ehrenfeld G M, Murugesan N & Hecht S M, *Inorg Chem,* 23 (1984) 1496.
77. Chang C & Meares C, *Biochem,* 23 (1984) 2268.
78. Kuwahara J, Suzuki T & Sugiura Y, *Biochem Biophys Res Comm,* 129 (1985) 368.
79. Guan L, Totsuka R, Kuwahara J, Otsuka M & Sugiura Y *Biochem Biophys Res Comm,* 191 (1993) 1338.
80. Guan L L, Kuwahara J & Sugiura Y, *Biochem,* 32 (1993) 6141.
81. Barton J K, *Science,* 233 (1986) 727.
82. Winston C T & Boger D L, *Chem Biol,* 11 (2004) 1607.
83. Boger D L & Cai H, *Angew Chem Int Ed Engl,* 38 (1999) 448.
84. Stubbe J, Kozarich J W, Wu W & Vanderwall D E, *Acc Chem Res,* 29 (1996) 322.
85. Hecht S M, *J Nat Prod,* 63 (2000) 158.
86. KONISHI M, OHKUMA H, MATSUMOTO K, TSUNO T, KAMEI H, MIYAKI T, OKI T, KAWAGUCHI H, VANDUYNE G D & Clardy J, *J Antibiot,* 42 (1989) 1449.
87. Konishi M, Ohkuma H, Tsuno T, Oki T, VanDuyne G D & Clardy J, *J Am Chem Soc,* 112 (1990) 3715.
88. Guo Q, Lu M, Shahrestanifar M, Sheardy R D & Kallenbach N R, *Biochem,* 30 (1991) 11735.
89. Rodolfo C, Lanza A, Tornaletti S, Fronza G & Pedrini A, *Nuc Acids Res,* 22 (1994) 314.
90. Golik J, Dubay G, Groenewold G, Kawaguchi H, Konishi M, Krishnan B, Ohkuma H, Saitoh K & Doyle T W, *J Am Chem Soc,* 109 (1987) 3462.
91. Kumar R A, Ikemoto N & Patel D J, *J Mol Biol,* 265 (1997) 173.
92. Long B H, Golik J, Forenza S, Ward B, Rehfuss R, Dabrowiak J C, Catino J J, Musial S T, Brookshire K W & Doyle T W, *Proc Nati Acad Sci U S A,* 86 (1989) 2.
93. Goldberg I H, *Acc Chem Res,* 24 (1991) 191.
94. Dasgupta D & Goldberg I H, *Nuc Acids Res,* 14 (1986) 1089.
95. Povirk L F, *Mutat Res,* 355 (1996) 71.
96. Lemon B & Tjian R, *Genes Dev,* 14 (2000) 2551.
97. Kopka M L, Yoon C, Goodsell D, Pjura P & Dickerson R E, *Proc Nati Acad Sci U S A,* 82 (1985) 1376.
98. Dervan P B, *Science* 232 (1986) 464.
99. Poot M, Kausch K, Köhler J, Haaf T & Hoehn H, *Cell structure and function,* 15 (1990) 151.
100. Nelson S M, Ferguson L R & Denny W A, *Mutat Res,* 623 (2007) 24.
101. Kopka M L, Yoon C, Goodsell D, Pjura P & Dickerson R E, *Proc Nati Acad Sci,* 82 (1985) 1376.
102. Ivanov V I, Minchenkova L E, Minyat E E, Frank-Kamenetskii M D & Schyolkina A K, *J Mol Biol,* 87 (1974) 817.
103. Zimmer C, Marck C & Guschlbauer W, *FEBS Lett* 154 (1983) 156.
104. Majumder P & Dasgupta D, *PLoS One* 6 (2011) e26486.
105. Zimmer C, *Prog Nucleic Acid Res Mol Biol,* 15 (1975) 285.
106. Harshman K D & Dervan P B, *Nucleic Acids Res,* 13 (1985) 4825.
107. Martin R F & Holmes N, *Nature,* 302 (1983) 452.
108. Teng M.-K, Usman N, Frederick C A & Wang A H.-J, *Nuc Acids Res,* 16 (1988) 2671.
109. Dubey D & Raman R, *Exp Cell Res,* 149 (1983) 419.
110. Dervan P B & Bürli R W, *Curr Opin Chem Biol,* 3 (1999) 688.
111. Bremer R E, Baird E E & Dervan P B, *Chem Biol,* 5 (1998) 119.
112. Khorlin A, Krylov A, Grokhovsky S, Zhuze A, Zasedatelev A, Gursky G & Gottikh B, *FEBS Lett,* 118 (1980) 311.

113. Schultz P & Dervan P, *J Am Chem Soc,* 105 (1983) 7748.
114. Hertzberg R P & Dervan P B, *J Am Chem Soc,* 104 (1982) 313.
115. Takusagawa F, Dabrow M, Neidle S & Berman H M, *Nature,* 296 (1982) 466.
116. Krivtsova M, Moroshkina E, Glibin E & Frisman E, *Molekuliarnaia Biologiia,* 18 (1984) 950.
117. Lu D.-F, Wang Y.-S, Li C, Wei G.-J, Chen R, Dong D.-M & Yao M, *Int J Clin Exp Med,* 8 (2015) 1904.
118. Walker G T, Stone M P, Krugh T R *Biochemistry* 24 (1985) 7471.
119. van de Sande J H, Jovin T M Z *EMBO J,* 1 (1982) 115.
120. Kamitori S, Takusagawa F *J Mol Biol,* 225 (1992) 445.
121. Scamrov A, Beabealashvilli R S *FEBS lett,* 164 (1983) 97.
122. Van Dyke M W, Hertzberg R P, Dervan P B, *Proc Nati Acad Sci U S A,* 79 (1982) 5470.
123. Lane M J, Dabrowiak J C, Vournakis J N, *Proc Nati Acad Sci,* 80 (1983) 3260.
124. Wartell R M, Larson J E, Wells R D, *J Biol Chem,* 250 (1975) 2698.
125. Crissman H A, Hirons G T, *Method Cell Biol,* 41 (1994) 195.
126. Haq I, *Arch Biochem Biophys,* 403 (2002) 1.
127. Vardevanyan P, Antonyan A P, Parsadanyan M, Davtyan H & Karapetyan A, *Exp Mol Med,* 35 (2003) 527.
128. Wu P K, Kharatishvili M, Qu Y & Farrell N, *J Inorg Biochem,* 63 (1996) 9.
129. Johnson A, Qu Y, Van Houten B & Farrell N B, *Nuc Acids Res,* 20 (1992) 1697.
130. Heller D P & Greenstock C L, *Biophys Chem,* 50 (1994) 305.
131. Jin L T & Choi J K, *Electrophoresis,* 25 (2004) 2429.
132. Kapuscinski J, *Biotech Histochem,* 70 (1995) 220.
133. Arndt-Jovin D J & Jovin T M, *Cell Biol,* 30 (1989) 417.
134. Nayak A K, Mishra A, Jena B S, Mishra B K & Subudhi U, *Sci Rep,* 6 (2016) 26855.
135. Sayed M, Krishnamurthy B & Pal H, *Phys Chem Chem Phys,* 18 (2016) 24642.
136. Spielman M, *J Am Chem Soc,* 57 (1935) 583.
137. Wagner E, *J American Chemical Society,* 57 (1935) 1296.
138. Rúnarsson Ö V, Artacho J, Wärnmark K *Eur J Org Chem* 2012 (2012) 7015.
139. Coppel Y, Coulombeau C, Coulombeau C, Lhomme J, Dheu-Andries M, Vatton P *J Biomol Struct Dyn* 12 (1994) 637.
140. Manda B R, Alla M, Ganji R J, Addlagatta A *Eur J Med Chem* 86 (2014) 39.
141. Veale E B & Gunnlaugsson T, *J Org Chem,* 75 (2010) 5513.
142. Yashima E, Akashi M & Miyauchi N, *Chem Lett,* 6 (1991) 1017.
143. Salez H, Wardani A, Demeunynck M, Tatibouet A & Lhomme J, *Tetrahedron Lett,* 36 (1995) 1271.
144. Bailly C, Laine W, Demeunynck M & Lhomme J, *Biophys Res Comm,* 273 (2000) 681.
145. Crick F, *Nature,* 227 (1970) 561.
146. SONNEBORN T M, *Degeneracy of the genetic code: extent nature and genetic implications In Evolving genes and proteins* (Elsevier) 1965, p. 377.
147. Stadhouders R, Filion G J & Graf T, *Nature,* 569 (2019) 345.
148. Klug A & Schwabe J W, *FASEB J,* 9 (1995) 597.
149. Alber T, *Curr Opin Genet Dev,* 2 (1992) 205.
150. Atchley W R & Fitch W M, *Proc Nati Acad Sci,* 94 (1997) 5172.
151. Harrison S C & Aggarwal A K, *Annu Rev Biochem,* 59 (1990) 933.
152. Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, Melino G & Raschellà G, *Cell Death Disc,* 3 (2017) 1.
153. Lee M S, Gippert G P, Soman K V, Case D A & Wright P E, *Science,* 245 (1989) 635.
154. Pavletich N P & Pabo C O, *Science,* 252 (1991) 809.
155. Elrod-Erickson M, Rould M A, Nekludova L & Pabo C O, *Structure,* 4 (1996) 1171.
156. Kim C A & Berg J M, *Nat Struct Biol,* 3 (1996) 940.
157. Sera T, *Adv Drug Delivery Rev,* 61 (2009) 513.
158. Neuhaus D, *Prog Nucl Magn Reson Spectrosc,* 130 (2022) 62.
159. Chen K & Rajewsky N, *Nat Rev Genet,* 8 (2007) 93.
160. Todeschini A.-L, Georges A & Veitia R A, *Trends Genet,* 30 (2014) 211.
161. Melloul D, Marshak S & Cerasi E, *Diabetologia,* 45 (2002) 309.
162. Krawczyk M & Reith W, *Tissue antigens,* 67 (2006) 183.
163. Perou C M, Sørlie T, Eisen M B, Van De Rijn M, Jeffrey S S, Rees C A, Pollack J R, Ross D T, Johnsen H & Akslen L A, *Nature,* 406 (2000) 747.
164. Pandolfi P P, *Oncogene,* 20 (2001) 3116.
165. Miller J, McLachlan A & Klug A, *EMBO J,* 4 (1985) 1609.
166. Beerli R R, Segal D J, Dreier B & Barbas III C F, *Proc Nati Acad Sci U S A,* 95 (1998) 14628.
167. Beerli R R, Dreier B & Barbas III C F, *Proc Nati Acad Sci,* 97 (2000) 1495.
168. Lund C V, Popkov M, Magnenat L & Barbas III C F, *Mol Cell Biol,* 25 (2005) 9082.
169. Bartsevich V V & Juliano R, *Mol Pharmacol* 58 (2000) 1.
170. Zhang L, Spratt S K, Liu Q, Johnstone B, Qi H, Raschke E E, Jamieson A C, Rebar E J, Wolffe A P & Case C C, *J Biol Chem,* 275 (2000) 33850.
171. Liu P-Q, Rebar E J, Zhang L, Liu Q, Jamieson A C, Liang Y, Qi H, Li P-X, Chen B & Mendel M C, *J Biol Chem,* 276 (2001) 11323.
172. Rebar E J, Huang Y, Hickey R, Nath A K, Meoli D, Nath S, Chen B, Xu L, Liang Y & Jamieson A C, *Nat Med,* 8 (2002) 1427.
173. Snowden A W, Zhang L, Urnov F, Dent C, Jouvenot Y, Zhong X, Rebar E J, Jamieson A C, Zhang H S & Tan S, *Cancer Res,* 63 (2003) 8968.
174. Yu J, Lei L, Liang Y, Hinh L, Hickey R P, Huang Y, Liu D, Yeh J L, Rebar E & Case C, *FASEB J,* 20 (2006) 479.
175. Price S A, Dent C, Duran-Jimenez B, Liang Y, Zhang L, Rebar E J, Case C C, Gregory P D, Martin T J & Spratt S K, *Diabetes,* 55 (2006) 1847.
176. Dai Q, Huang J, Klitzman B, Dong C, Goldschmidt-Clermont P J, March K L, Rokovich J, Johnstone B, Rebar E J & Spratt S K, *Circulation,* 110 (2004) 2467.
177. Sera T & Uranga C, *Biochem,* 41 (2002) 7074.
178. Kang Y.-A, Shin H.-C, Yoo J Y, Kim J.-H, Kim J.-S & Yun C.-O, *Mol Ther,* 16 (2008) 1033.
179. Jouvenot Y, Ginjala V, Zhang L, Liu P, Oshimura M, Feinberg A, Wolffe A, Ohlsson R & Gregory P, *Gene Ther,* 10 (2003) 513.
180. Falke D, Fisher M, Ye D & Juliano R, *Nuc Acids Res,* 31 (2003) e10.
181. Bartsevich V V, Miller J C, Case C C & Pabo C O, *Stem Cells,* 21 (2003) 632.

182. Tan S, Guschin D, Davalos A, Lee Y.-L, Snowden A W, Jouvenot Y, Zhang H S, Howes K, McNamara A R & Lai A, *Proc Nati Acad Sci,* 100 (2003) 11997.
183. Gräslund T r, Li X, Magnenat L, Popkov M & Barbas C F, *J Biol Chem,* 280 (2005) 3707.
184. Beltran A, Parikh S, Liu Y, Cuevas B, Johnson G, Futscher B & Blancafort P, *Oncogene,* 26 (2007) 2791.
185. Onori A, Desantis A, Buontempo S, Di Certo M G, Fanciulli M, Salvatori L, Passananti C & Corbi N, *Biochem Cell Biol,* 85 (2007) 358.
186. Mattei E, Corbi N, Di Certo M G, Strimpakos G, Severini C, Onori A, Desantis A, Libri V, Buontempo S & Floridi A, *PLoS One,* 2 (2007) e774.
187. Yokoi K, Zhang H S, Kachi S, Balaggan K S, Yu Q, Guschin D, Kunis M, Surosky R, Africa L M & Bainbridge J W, *Mol Ther,* 15 (2007) 1917.
188. Reynolds L, Ullman C, Moore M, Isalan M, West M J, Clapham P, Klug A & Choo Y, *Proc Nati Acad Sci,* 100 (2003) 1615.
189. Segal D J, Gonçalves J, Eberhardy S, Swan C H, Torbett B E, Li X & Barbas C F, *J Biol Chem,* 279 (2004) 14509.
190. Kim Y.-S, Kim J.-M, Jung D.-L, Kang J.-E, Lee S, Kim J S, Seol W, Shin H.-C, Kwon H S & Van Lint C, *J Biol Chem* 280 (2005) 21545.
191. Papworth M, Moore M, Isalan M, Minczuk M, Choo Y & Klug A, *Proc Nati Acad Sci,* 100 (2003) 1621.
192. Zimmerman K A, Fischer K P, Joyce M A & Tyrrell D L J, *J Virol,* 82 (2008) 8013.
193. Nataraj D, Srinivasan N, Sowdhamini R & Ramakrishnan C, *Curr Sci,* 69 (1995) 434.
194. Chou K C *Biopolymers* 42 (1997) 837.
195. Pavone V, Gaeta G, Lombardi A, Nastri F, Maglio O, Isernia C & Saviano M, *Biopolymers,* 38 (1996) 705.
196. Shepherd N E, Hoang H N, Abbenante G & Fairlie D P, *J Am Chem Soc,* 127 (2005) 2974.
197. Zimm B H & Bragg J, *J Chem Phy,* 31 (1959) 526.
198. Nick Pace C, Trevino S, Prabhakaran E & Martin S J, *Phil Trans R Soc B,* 359 (2004) 1225.
199. Ghadiri M R & Fernholz A K, *J Am Chem Soc,* 112 (1990) 9633.
200. Ruan F, Chen Y & Hopkins P B, *J Am Chem Soc,* 112 (1990) 9403.
201. Zhang Q, Jiang F, Zhao B, Lin H, Tian Y, Xie M, Bai G, Gilbert A M, Goetz G H & Liras S, *Sci Rep,* 6 (2016) 38573.
202. Scholtz J M & Baldwin R L, *Annu Rev Biophys Biomol Struct,* 21 (1992) 95.
203. Tsou L K, Tatko C D & Waters M L, *J Am Chem Soc,* 124 (2002) 14917.
204. Albert J S & Hamilton A D, *Biochem,* 34 (1995) 984.
205. Karle I L & Balaram P, *Biochem,* 29 (1990) 6747.
206. Augspurger J, Bindra V, Scheraga H & Kuki A, *Biochem,* 34 (1995) 2566.
207. Toniolo C, Crisma M, Formaggio F & Peggion C, *Pept Sci* 60 (2001) 396.
208. Shamala N, Nagaraj R & Balaram P, *J Chem Soc Chem Comm,* 1978 (22) 996.
209. Blackwell H E & Grubbs R H, *Angew Chem Int Ed,* 37 (1998) 3281.
210. Blackwell H E, Sadowsky J D, Howard R J, Sampson J N, Chao J A, Steinmetz W E, O'Leary D J & Grubbs R H, *J Org Chem,* 66 (2001) 5291.
211. Pal S, *Chem Sel,* 8 (2023) e202204207.
212. Cabezas E & Satterthwait A C, *J Am Chem Soc,* 121 (1999) 3862.
213. Kalia J & Raines R T, *Angew Chem Int Ed,* 47 (2008) 7523.
214. Patgiri A, Jochim A L & Arora P S, *Acc Chem Res,* 41 (2008) 1289.
215. Schmidt B & Hermanns J, *Curr Org Chem,* 10 (2006) 1363.
216. Miller S E, Kallenbach N R & Arora P S, *Tetrahedron,* 68 (2012) 4434.
217. Wang D, Chen K, Dimartino G & Arora P S, *Org Biomol Chem,* 4 (2006) 4074.
218. Vernall A J, Cassidy P & Alewood P F, *Angew Chem Int Ed,* 48 (2009) 5675.
219. Wang L, Coric P, Zhu K, Liu W.-Q, Vidal M, Bouaziz S & Broussy S, *Org Biomol Chem,* 16 (2018) 459.
220. Wang L, Liu W.-Q, Saraux N, Vidal M & Broussy S, *Tetrahedron Lett,* 56 (2015) 2456.
221. Pal S, Banerjee S, Kumar A & Prabhakaran E N, *ACS Omega,* 5 (2020) 13902.
222. Pal S & Prabhakaran E N, *Eur J Org Chem,* 2021 (2021) 1714.
223. Chin D.-H, Woody R W, Rohl C A & Baldwin R L, *Proc Nati Acad Sci,* 99 (2002) 15416.
224. Narita M, Sode K & Ohuchi S, *Bull Chem Soc Jpn,* 72 (1999) 1807.
225. Dasgupta B, Pal L, Basu G & Chakrabarti P, *Proteins: Struct Funct Bioinf,* 55 (2004) 305.
226. Dames S A, Martinez-Yamout M, De Guzman R N, Dyson H J & Wright P E, *Proc Nati Acad Sci,* 99 (2002) 5271.
227. Moreira I S, Fernandes P A & Ramos M J, *Proteins: Struct Funct Bioinf,* 68 (2007) 803.
228. Kong E, Pike A & Hubbard R, *Biochem Soc Trans,* 31 (2003) 56.
229. Heinlein C A & Chang C, *Endocr Rev,* 25 (2004) 276.
230. Burstein E S, Spalding T A & Brann M R, *J Biol Chem,* 271 (1996) 2882.
231. Kim C, Schmidt T, Cho E.-G, Ye F, Ulmer T S & Ginsberg M H, *Nature,* 481 (2012) 209.
232. Armstrong A H, Chen J, McKoy A F & Hecht M H, *Biochem,* 50 (2011) 4058.
233. Nakagawa S H & Tager H S, *Biochem,* 31 (1992) 3204.
234. Zhang D.-W, Cole S P & Deeley R G, *J Biol Chem,* 277 (2002) 20934.
235. Ponsel D & Bruss V, *J Virol,* 77 (2003)416.
236. De Araujo A D, Hoang H N, Kok W M, Diness F, Gupta P & Hill T A, *Int Ed,* 53 (2014) 6965.
237. Pal S, Banerjee S & Prabhakaran E N, *FEBS Lett,* 595 (2021) 2942.
238. Pal S, Banerjee S & Prabhakaran E N, *J Phys Chem, A* 124 (2020) 7478.
239. Pal S & Prabhakaran E N, *Tetrahedron Lett,* 59 (2018) 2515.
240. Reddy S S, Pal S, Ghosh S & Prabhakaran E N, *Chem Bio Chem,* 22 (2021) 2111.
241. Jamieson A C, Wang H & Kim S-H, *Proc Nati Acad Sci,* 93 (1996) 12834.
242. Isalan M, Klug A & Choo Y A, *Nat Biotechnol,* 19 (2001) 656.
243. Nagai K, Nakaseko Y, Nasmyth K & Rhodes D, *Nature,* 332 (1988) 284.
244. Zhao D & Huang Z, *Bioinorg Chem Appl,* 2016 (2016) 8206854.

245. Zhao D.-X, Ding Z.-C, Liu Y.-Q & Huang Z.-X, *Protein Expression Purif*, 53 (2007) 232.
246. Jiang L, Zhang B, Wang G, Wang K & Xiao X, *Mol Cell Biochem,* 328 (2009) 137.
247. Young T S, Ahmad I, Yin J A & Schultz P G, *J Mol Biol,* 395 (2010) 361.
248. Ryu Y & Schultz P G, *Nat Methods,* 3 (2006) 263.
249. Kilburn L B, Khuong-Quang D.-A, Hansford J R, Landi D, van der Lugt J, Leary S E, Driever P H, Bailey S, Perreault S & McCowage G, *Nat Med,* 30 (2024) 207.
250. Tkacik E, Li K, Gonzalez-Del Pino G, Ha B H, Vinals J, Park E, Beyett T S & Eck M J, *J Biol Chem,* 299 (2023) 5.
251. González-Sales M, Lennox A L, Huang F, Pamulapati C, Wan Y, Sun L, Berry T, Kelly Behrs M, Feller F & Morcos P N, *CPT Pharmacom Syst Pharm ,*2024.
252. Effat H, Abosharaf H A & Radwan A M, *Sci Rep,* 14 (2024) 2824.
253. Pérez-Arnaiz C, Busto N, Leal J M & García B, *J Phys Chem B,* 118 (2014) 1288.