


 Cite this: *Chem. Commun.*, 2021, 57, 12289

 Received 6th August 2021,
Accepted 27th October 2021

DOI: 10.1039/d1cc04306d

rsc.li/chemcomm

Modular assembly and encoding strategies for dual-display DNA-encoded chemical libraries†

 Sebastian Oehler,^a Louise Plais,^a Gabriele Bassi,^a Dario Neri^{b*} and Jörg Scheuermann^{a*}

DNA-encoded chemical libraries (DELs) are increasingly being used for the discovery of protein ligands and can be constructed displaying either one or two molecules at the extremities of the two complementary DNA strands. Here, we describe that DELs, featuring the simultaneous display of two molecules, can be encoded using various types of DNA structures, which go beyond the use of conventional double-stranded DNA fragments. Specifically, we compared dual-display methodologies in hairpin, circular or linear formats in terms of polymerase chain reaction (PCR) amplifiability and performance in affinity capture selections. The methods reported in this article highlight the feasibility and modularity of the described encoding strategies and may thus further expand the scope of DNA-encoded chemistry, particularly for the identification of compounds which recognize adjacent epitopes on the surface of target proteins of interest.

DNA-encoded chemical libraries (DELs) are pools of small organic molecules individually linked to DNA tags which serve as amplifiable identification barcodes.^{1–4} In many cases, DELs are produced in a step-wise fashion (*e.g.* by split & pool methods⁵), assembling two or more building blocks into a single nascent molecule at the extremity of the DNA tag, which can be either in single- or double-stranded format.^{6–8} Alternatively, two different molecules can be displayed at the two strands of a DNA heteroduplex. Such a strategy enables the construction of very large molecular repertoires by the self-assembly of complementary sublibraries, constructed in single-stranded DNA format.^{9–12} This approach, which is referred to as encoded self-assembling chemical (ESAC) libraries, has been expanded to the use of peptide nucleic acid (PNA) structures^{13–15} and even to the construction and screening of DNA-encoded dynamic libraries.^{16–19} ESAC technology has been used to affinity mature protein ligands by coupling them to synergistic complementary fragments.

For example, carbonic anhydrase IX (CAIX) binders with affinity in the sub-nanomolar dissociation range have been discovered using ESAC technology and have successfully been used in tumor targeting applications.¹⁰ Alternatively, large combinatorial ESAC libraries facilitated the *de novo* discovery of synergistic ligand pairs (which can be coupled together), yielding high affinity ligands to targets such as AGP,¹⁰ JNK-1¹¹ and HSA.²⁰ Similarly, DNA-templated PNA libraries and dynamic libraries in dual-display format led, amongst others, to the discovery of new ligands against carbonic anhydrases,^{14,19} the phosphatase PTP1B¹³ and sirtuin3.¹⁸ The strategies reported here enable the pairing of partial and non-complementary single-stranded DELs of varying sizes and will eventually allow the construction of very large and diverse dual-pharmacophore libraries.

Here, we describe the design and implementation of three dual-display methodologies as alternative to the conventional display of molecule pairs at the extremities of complementary DNA strands. In particular, we focused on a hairpin design, on a circular construct and on a linear format which incorporates two molecules in the middle of a DNA fragment (Fig. 1). The latter approach had previously been reported for PNA libraries^{14,21} and for model studies using DNA as a molecular ruler.^{22,23}

We designed 3'-modified oligonucleotides with short coding regions which are compatible (*i.e.*, that do not form unwanted self- or hetero-dimers) with previously described single-stranded DELs or sublibraries.^{9,10,20,24} Using a hairpin design approach, we connected the new oligonucleotides to the existing sublibraries by performing a splint ligation, which yielded structures containing a 16-base pair stem (to which two different molecules were attached) and a connecting loop (comprising coding regions for the corresponding molecules). Thus, partial complementary sublibraries assembled as hairpin structures (Fig. 1A). During hairpin construction, formation of dimeric structures was observed which could be reversed to hairpins by a fast heat-cool cycle (see Fig. S1, ESI†).

An alternative strategy for dual-display DEL construction consists in creating circular DNA structures (Fig. 1B). In this setting, two non-complementary sublibraries (each comprising

^a Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 3, Zürich 8093, Switzerland. E-mail: joerg.scheuermann@pharma.ethz.ch

^b Philochem AG, Libernstrasse 3, Otelfingen, 8112, Switzerland. E-mail: dario.neri@pharma.ethz.ch

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1cc04306d



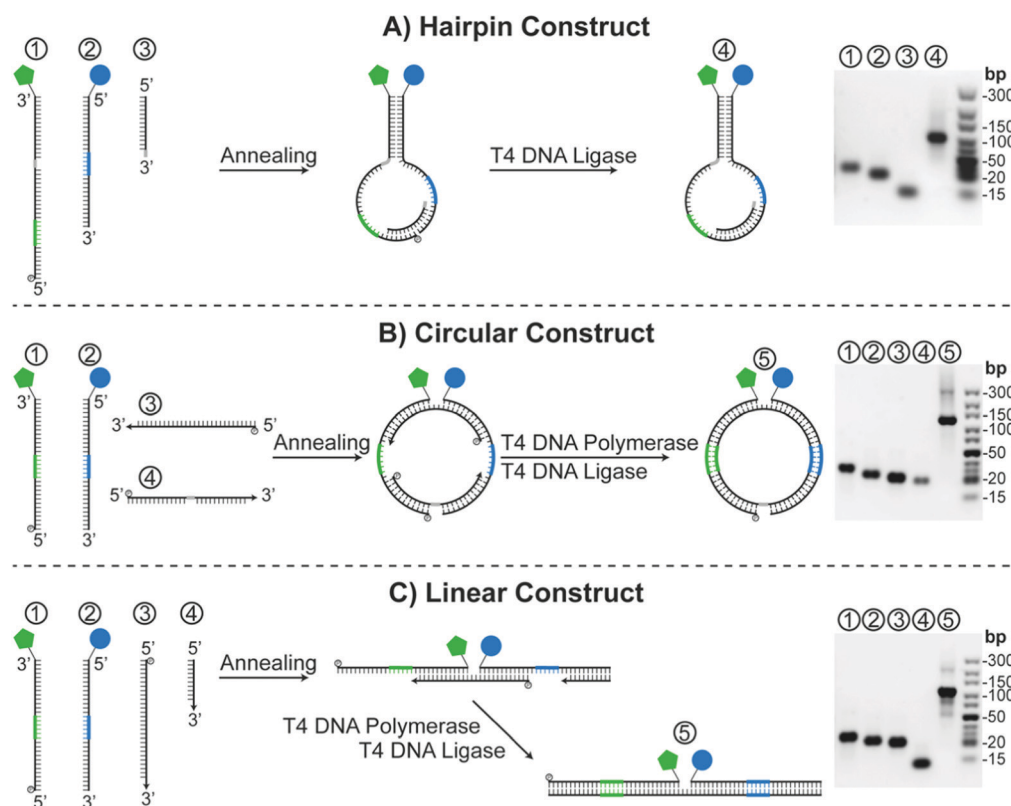


Fig. 1 Modular encoding strategies for the one-pot construction of dual-pharmacophore libraries. Assembly of partially complementary libraries led to hairpin structures (A). Non-complementary sublibraries were assembled with two relay primers to yield circular constructs (B) while the use of one relay primer and one terminal primer resulted in linear formats (C). The formation of the library constructs could be followed on agarose gel as shown on the right. bp = base pairs.

molecules attached to the extremity of single-stranded DNA fragments containing a coding region), could be converted into a circular structure by the use of two relay primers. One of the two components of the circular structure consists of an intact circular DNA moiety (inner circle in Fig. 1B). Formation of the intact circle resulted in a stable complex with the sublibraries for the dual-display of building blocks while functioning as template to encode the respective library pairs.

A linear design can also be implemented by anchoring two sets of chemically modified single-stranded DNA fragments onto a complementary DNA template. The linear template could be formed by the use of one relay and one terminal primer so that the molecule pairs are presented in the middle of the linear construct (Fig. 1C).

The construction of hairpin formats by pairing partial complementary sublibraries or circular and linear constructs with non-complementary sublibraries was developed in a one-pot procedure which could be followed by agarose gel (see Fig. 1). Construct formation could also be observed by LC-MS analysis (see Fig. S2–S6, ESI[†]). Furthermore, the circular and linear templates comprised one EcoRI restriction site each. In theory, the single cleavage of circular constructs should yield one linear product while the linear template should be cleaved into two parts. Indeed, EcoRI digestion led to the predicted cleavage patterns as visualised by agarose gel (see Fig. S7, ESI[†]).

The described encoding strategies are based on a code-region independent assembly. This should allow not only combining libraries with different sequences but also with different numbers of coding regions, thus expanding the combinatorial scope to the use of sublibraries encoding more than one building block. For the hairpin structure, pairing of a 3'-modified sublibrary comprising one code with a partial complementary 5'-modified sublibrary with two ("1 + 2") coding regions revealed the expected ligation product by LC-MS analysis (see Fig. S9, ESI[†]). The 3'-modified non-complementary sublibrary could be expanded *via* splint ligation with a second coding region (see Fig. S10, ESI[†]). Therefore, one-pot assembly and formation of the circular and linear constructs could be tested in "1 + 2", "2 + 1" and "2 + 2" formats. Agarose gel analysis, EcoRI digestion and LC-MS analysis revealed the expected products (see Fig. S11–S25, ESI[†]).

We characterized the possibility to PCR amplify the different dual-display DEL formats by using different amounts of DNA input, ranging from 10^{10} to 0 molecules. Fig. 2 shows that the hairpin structure (A) revealed a higher threshold at around 10^5 DNA molecules as input while circular (B) and linear (C) constructs could be amplified down to 10^3 DNA molecules.

PCR amplifiability (10 to 100-fold) decreased for constructs consisting of pairs of sublibraries with more than one code (see Fig. S26–S28, ESI[†]). Notably, the productive amplification at



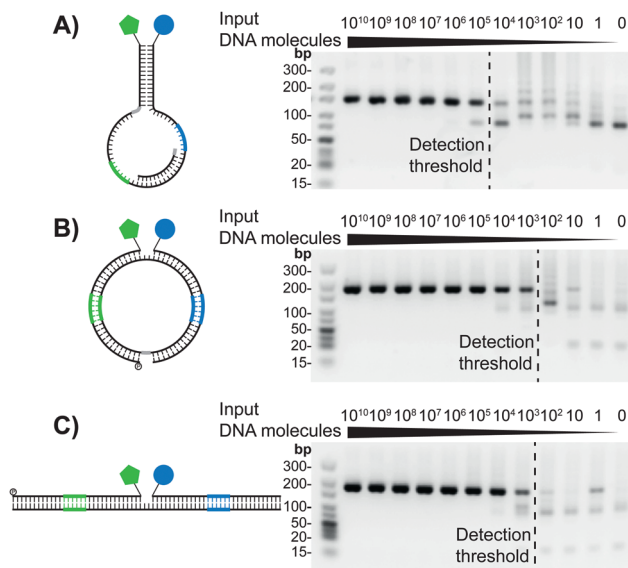


Fig. 2 PCR amplifiability of the dual-display libraries in hairpin (A), circular (B) and linear (C) format. The PCR amplification thresholds are highlighted by dashed lines.

low DNA inputs as observed in Fig. 2 is an important feature to allow encoding and ensure readability on a real library scale.

In order to demonstrate that the dual-display formats were compatible with library construction and affinity selection procedures, we generated small molecular assemblies consisting of five encoded building blocks on the 3'-modified sub-libraries which were assembled with a previously described 5'-modified DEL comprising 553 members.^{10,24} Screening of this particular 5'-modified sublibrary in dual-display ESAC format had previously led to the identification of an affinity-matured CAIX ligand of acetazolamide, "AAZ⁺", and an AGP ligand pair.¹⁰ The respective 3'-coupled counterparts, acetazolamide and a furan derivative (AGP ligand fragment), were encoded on the 3'-sublibraries including two additional molecules and also acetylated DNA, which was used at 100 times excess, serving as artificial selection background (building block structures see Fig. S29, ESI[†]). Thereby, the 100-fold presence of 553 library members (5'-sublibrary paired with 3'-acetylated DNA) enabled to emulate a library size of 57⁵512 members with only 2⁷765 individual pharmacophore pairs. Prior knowledge of the comprised binder pairs allowed us to compare the performance of the three described dual-display formats with the classical ESAC approach. The ESAC library was constructed *via* Klenow polymerization as reported previously.¹⁰

Fig. 3 shows that screening of ESAC, hairpin, circular and linear library constructs against CAIX led to the preferential enrichment of acetazolamide comprising library members (CodeA:2). With acetazolamide as constant 3'-coupled building block (Fig. 3, framed plane CodeA:2), enriched dots representing individual library members could reveal potential binder pairs. Only in ESAC and hairpin format, the known bisphenol fragment (CodeB:493) which led to "AAZ⁺" was identified. Selection of circular and linear constructs resulted in the

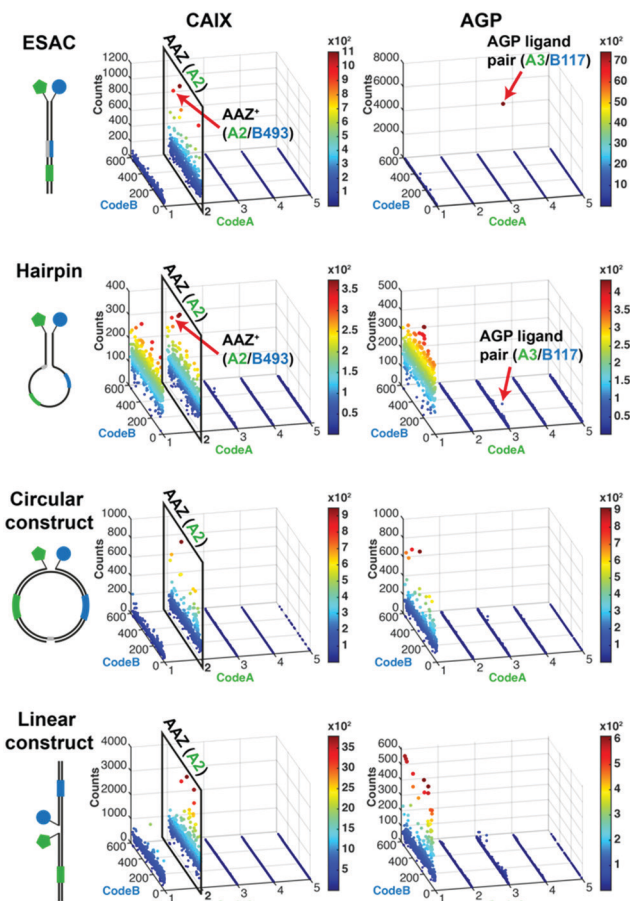


Fig. 3 Comparison of different dual-pharmacophore library formats: ESAC, hairpin, circular and linear constructs. Fingerprints of the libraries (553 × 5 building blocks with acetylated DNA = CodeA:1 in 100-fold excess) after selection against CAIX (left) and AGP (right). Each library member is represented by a dot for which the z-axis and color illustrates the normalized sequence count. Preferentially enriched library members were identified as CAIX binders (acetazolamide (AAZ) = CodeA:2 and "AAZ⁺" = A2/B493) and AGP ligand pair (A3/B117). Building block structures and fingerprints of selection duplicates are given in Fig. S29–S41 (ESI[†]).

preferential enrichment of unrelated ligand pairs which could derive from synergistic binding events with different sites on the protein in comparison to fragments identified with ESAC or hairpin libraries. Off-DNA hit resynthesis or focussed libraries will be necessary to validate these findings.¹⁵ The AGP ligand pair (A3/B117) was highly enriched for the ESAC library, at lower level also for the hairpin construct but not for circular or linear library formats (see Fig. 3). Decreased enrichments of expected ligands leads to higher background, visible as CodeA:1 plane due to the 100-fold spike-in of the respective 3'-acetylated DNA. As expected, the CodeA:1 plane can also be observed in the fingerprints of the libraries before selections and after selections against "empty" beads (no protein immobilized, see Fig. S30, S33, S36 and S39, ESI[†]).

The presentation of two molecules at the extremity of a DNA heteroduplex for ESAC and hairpin formats implies similar dual-display geometries which might be reflected by the



coinciding fingerprints shown in Fig. 3. For both, expected pharmacophore pairs were enriched with lower signal-to-noise ratio for the hairpin construct. In contrast, fingerprints of circular and linear libraries revealed enrichment of CAIX ligands but not CAIX and AGP ligand pairs (see Fig. 3). Different enrichment patterns might reflect different binding interactions or encoding artefacts. To address this issue, we implemented an electrophoretic mobility shift assay with the AGP ligand pair (A3/B117) assembled in ESAC, hairpin, circular and linear format (see Fig. S42, ESI†). Titration of AGP resulted in detectable binding events for the ESAC and hairpin format with an apparent dissociation constant (K_D) around 160 nM, while no binding could be observed for ligand presentation in circular or linear format. Thus, decreased signal-to-noise ratio with the hairpin library might derive from assembly or sequencing artefacts. Circular and linear libraries did not reveal an enrichment of the known AGP ligand pair, which is in line with the respective absence of a band-shift in the electrophoretic mobility shift assay, indicating a low affinity of these display formats (Fig. S42, ESI†). Reduced affinity might derive from different spatial arrangements. The topography of building block presentation at the extremity of a DNA heteroduplex in comparison to templated assemblies remains to be investigated. For both, spatial proximity as well as avidity or synergistic binding effects have been exploited in the background of DNA-templated chemistry, DNA-controlled distancing and DNA- or PNA-encoded chemical libraries.^{9,14,15,22,25,26} Notably, templated assembly of sublibraries might introduce additional flexibility between building block presenting sublibraries.²² In consequence, one problem could be spatial distortion or coiling effects so that the building blocks face different directions. Furthermore, libraries were constructed at a fixed distance of two nucleotides in the bridging region between the pharmacophore-carrying sublibrary ends. As previously reported, the affinity of fragment pairs that bind to adjacent pockets on a target protein of interest depends on the use of suitable linkers^{15,27} for which optimal distancing may be an important factor. In theory, modular pairing of non-complementary libraries in circular and linear format could be exploited to identify distance-dependent binder pairs on DNA-encoded library level by varying the number of nucleotides within the bridging region.

We describe three dual-display strategies which allow to pair different sublibraries in hairpin, circular and linear formats. One-pot assembly and encoding yielded dual-pharmacophore libraries which could efficiently be amplified *via* PCR. Selection of small libraries against AGP and CAIX in comparison to the classical ESAC approach revealed similarities between ESAC and hairpin libraries with a lower signal-to-noise ratio for the latter. For the circular and linear libraries, individual CAIX ligands could be enriched but not previously identified synergistic AGP or CAIX fragment pairs, suggesting a different ligand display geometry. The modularity and structural diversity of the new encoding strategies might provide useful tools for the

combinatorial pairing of different sublibraries, to identify new dual-pharmacophore ligands.

We gratefully acknowledge funding by the Swiss National Science Foundation (grant 310030_182003/1), the European Research Council (ERC, grant agreement 670603) and the Innosuisse – Swiss Innovation Agency, (48350.1 IP-LS).

Detailed information about materials and methods used within this study is given in the ESI.†

Conflicts of interest

Dario Neri is a cofounder and shareholder of Philochem AG (Otfelfingen, Switzerland) and Philogen S.p.A. (Siena, Italy). The other authors disclosed no potential conflicts of interest.

Notes and references

- 1 S. Brenner and R. A. Lerner, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 5381–5383.
- 2 R. A. Goodnow, C. E. Dumelin and A. D. Keefe, *Nat. Rev. Drug Discovery*, 2017, **16**, 131–147.
- 3 D. Neri and R. A. Lerner, *Annu. Rev. Biochem.*, 2018, **87**, 479–502.
- 4 A. Gironde-Martínez, E. J. Donckele and F. Samain, *et al.*, *ACS Pharmacol. Transl. Sci.*, 2021, **4**, 1265–1279.
- 5 Á. Furka, F. Sebestyén and M. Asgedom, *et al.*, *Int. J. Pept. Protein Res.*, 1991, **37**, 487–493.
- 6 G. Bassi, N. Favalli and S. Oehler, *et al.*, *Biochem. Biophys. Res. Commun.*, 2020, **533**, 223–229.
- 7 L. Mannocci, Y. Zhang and J. Scheuermann, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17670–17675.
- 8 M. A. Clark, R. A. Acharya and C. C. Arico-Muendel, *et al.*, *Nat. Chem. Biol.*, 2009, **5**, 647–654.
- 9 S. Melkko, J. Scheuermann and C. E. Dumelin, *et al.*, *Nat. Biotechnol.*, 2004, **22**, 568–574.
- 10 M. Wichert, N. Krall and W. Decurtins, *et al.*, *Nat. Chem.*, 2015, **7**, 241–249.
- 11 G. Zimmermann, U. Rieder and D. Bajic, *et al.*, *Chem. – Eur. J.*, 2017, **23**, 8152–8155.
- 12 J. Scheuermann and D. Neri, *Curr. Opin. Chem. Biol.*, 2015, **26**, 99–103.
- 13 S. Barluenga, C. Zambaldo and H. A. Ioannidou, *et al.*, *Bioorg. Med. Chem. Lett.*, 2016, **26**, 1080–1085.
- 14 J. P. Dagher, M. Ciobanu and S. Alvarez, *et al.*, *Chem. Sci.*, 2011, **2**, 625–632.
- 15 J. P. Dagher, C. Zambaldo and M. Ciobanu, *et al.*, *Chem. Sci.*, 2015, **6**, 739–744.
- 16 G. Li, W. Zheng and Z. Chen, *et al.*, *Chem. Sci.*, 2015, **6**, 7097–7104.
- 17 F. V. Reddavid, W. Lin and S. Lehnert, *et al.*, *Angew. Chem., Int. Ed.*, 2015, **54**, 7924–7928.
- 18 Y. Zhou, C. Li and J. Peng, *et al.*, *J. Am. Chem. Soc.*, 2018, **140**, 15859–15867.
- 19 F. V. Reddavid, M. Cui and W. Lin, *et al.*, *Chem. Commun.*, 2019, **55**, 3753–3756.
- 20 G. Bassi, N. Favalli and M. Vuk, *et al.*, *Adv. Sci.*, 2020, **7**, 2001970.
- 21 K. Gorska, K.-T. Huang and O. Chaloin, *et al.*, *Angew. Chem., Int. Ed.*, 2009, **48**, 7695–7700.
- 22 H. Eberhard, F. Diezmann and O. Seitz, *Angew. Chem., Int. Ed.*, 2011, **50**, 4146–4150.
- 23 F. Abendroth, A. Bujotzek and M. Shan, *et al.*, *Angew. Chem., Int. Ed.*, 2011, **50**, 8592–8596.
- 24 C. E. Dumelin, J. Scheuermann and S. Melkko, *et al.*, *Bioconjugate Chem.*, 2006, **17**, 366–370.
- 25 Z. J. Gartner and D. R. Liu, *J. Am. Chem. Soc.*, 2001, **123**, 6961–6963.
- 26 K. I. Sprinz, D. M. Tagore and A. D. Hamilton, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3908–3911.
- 27 M. Bigatti, A. Dal Corso and S. Vanetti, *et al.*, *ChemMedChem*, 2017, **12**, 1748–1752.

