Organic & Biomolecular Chemistry

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2019, **17**, 908

Design criteria for minimalist mimics of protein–protein interface segments†

Jaru Taechalertpaisarn, [a] Rui-Liang Lyu, [a] Maritess Arancillo, [a] Chen-Ming Lin, [a] Zhengyang Jiang, [a] Lisa M. Perez,[b] Thomas R. Ioerger[c] and Kevin Burgess *[a]

Received 20th November 2018,
Accepted 18th December 2018

DOI: 10.1039/c8ob02901f

rsc.li/obc

Small molecules that can interrupt or inhibit protein–protein interactions (PPIs) are valuable as probes in chemical biology and medicinal chemistry, but they are also notoriously difficult to develop. Design of non-peptidic small molecules that mimic amino acid side-chain interactions in PPIs ("minimalist mimics") is seen as a way to fast track discovery of PPI inhibitors. However, there has been little comment on general design criteria for minimalist mimics, even though such guidelines could steer construction of libraries to screen against multiple PPI targets. We hypothesized insight into general design criteria for minimalist mimics could be gained by comparing preferred conformations of typical minimalist mimic designs against side-chain orientations on a huge number of PPI interfaces. That thought led to this work which features nine minimalist mimic designs: one from the literature, and eight new "hypothetical" ones conceived by us. Simulated preferred conformers of these were systematically aligned with >240 000 PPI interfaces from the Protein Data Bank. Conclusions from those analyses did indeed reveal various design considerations that are discussed here. Surprisingly, this study also showed one of the minimalist mimic designs aligned on PPI interface segments more than 15 times more frequently than any other in the series (according to uniform standards described herein); reasons for this are also discussed.

Introduction

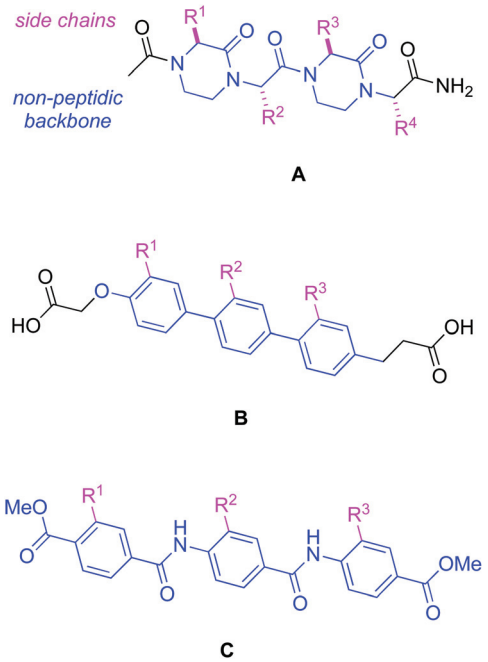
Genomic and proteomic studies are uncovering a vast PPI network that controls cell signaling processes. This information provides opportunities for development of non-peptidic small molecule probes to selectively perturb PPIs[1,2] but the design of such compounds is extremely difficult.[3] One approach to this problem is the use of "minimalist mimics",[4–6] of which **A**,[7] **B**,[8] and **C**[9] are illustrative.

Minimalist mimics are small molecules that do not have peptidic backbones, but nevertheless project amino acid side-chains in orientations that resemble protein secondary structures. The rationale for using minimalist mimics to disrupt PPIs is based on the observation that interaction energies in PPIs are mostly based on side-chain to side-chain interactions (>88%).[1,10–12] Small molecules that can present the same or similar side-chains in orientations similar to part of the protein ligand will tend to reproduce interactions between the ligands and their receptors, and perhaps bind the receptor and displace the ligand. Decades of research in this area has been driven by the assumption that if a *secondary structure* is found at a PPI interface then a minimalist mimic that resembles that particular secondary structure is a candidate to competitively inhibit the PPI.[13–15] However, this *secondary structure hypothesis* cannot be applied to PPI interfaces that do not involve secondary structure elements, and there are many of these. Furthermore, it would be advantageous to understand favorable structural characteristics of minimalist mimics that could be incorporated into libraries of these compounds used repeatedly to screen multiple PPI targets. To do this, we considered using a technique developed in our laboratory: EKO (Exploring Key Orientations).[11]

EKO (Fig. 1) was designed to evaluate minimalist mimics (specifically, those with *three* amino acid side-chains) against actual PPI interfaces. Minimalist mimics with three methyl substituents (representing Cα–Cβ vectors in three amino acid side-chains) are first run through a molecular dynamics routine (namely, Quenched Molecular Dynamics, QMD[16,17]) to determine orientations of Cα–Cβ vectors in preferred conformers (*e.g.* within 3 kcal mol^{-1} of the lowest energy one observed). The Cα–Cβ coordinates of each preferred conformer are then systematically aligned with those of every set of three side-

[a]*Department of Chemistry and Laboratory For Molecular Simulation, Texas A & M University, Box 30012, College Station, TX 77842-3012, USA. E-mail: burgess@tamu.edu*
[b]*Laboratory For Molecular Simulation, Texas A & M University, Box 30012, College Station, TX 77842-3012, USA*
[c]*Department of Computer Science, Texas A & M University, College Station, TX 77843-3112, USA*

†Electronic supplementary information (ESI) available: Detailed description of the EKO process; synthesis procedure and characterization data for mimics **1** and **5**. See DOI: 10.1039/c8ob02901f



chains in crystallographically characterized PPIs. Outputs of each overlay are expressed in terms of Root Mean Square Deviations (RMSDs, Å) of the three side chains. In our experience, overlays with RMSDs <0.35 Å based on these six coordinates represent good correspondence, and ones of <0.20 Å are excellent. This procedure is designed to test if a minimalist mimic scaffold is capable of projecting amino acid side-chains in certain orientations, irrespective of any side-chain-to-side-chain interactions. Side-chain-to-side-chain and side-chain-to-main-chain interactions influence peptide conformations in solution, but that is of little consequence here; the objective of this work is to find conformations of minimalist mimic backbones that may present side-chains to interact with a protein binding partner that naturally accommodates those side-chains in that conformation. Thus, the three methyl groups on the scaffolds can represent conformations of any set of three amino acids on the protein ligand at the interfaces of a PPI.

To date, EKO has been used in one way. A researcher inputs a series of minimalist mimic designs, then uses EKO to determine which can adopt side-chain conformations found at a particular PPI interface. If there is an encouraging degree of fit, then the researcher identifies precisely which interface residues the mimic overlaid on. To explore that virtual lead experimentally, it is then necessary to prepare the minimalist mimic scaffold with the specified side-chains, and assay these to find out if they perturbed the parent PPI. This approach has been successfully used in our laboratory to evaluate inhibitors, and this has led to compounds that have been experimentally proven to perturb the HIV-1 protease dimer,[11] ones that enhance oligomerization of antithrombin,[18] and others that inhibit PCSK9·LDLR interaction[19] (see ESI† for an overview of this published work).

This paper features a new use of EKO: developing *general* design criteria for superior minimalist mimics. To do that, preferred conformations of one minimalist mimic are aligned with vast numbers (*e.g.* >240 000) of PPI interfaces. Mimics that overlay very frequently in this process must resemble interface segments in some way, while ones that do not are inferior interface mimics. This process may be performed without regard to any particular secondary structures found at interfaces.

As a platform to explore this new use of EKO in establishing general design criteria for minimalist mimics, we conceived eight mimics, **1–8**, that had not been prepared or reported previously, but which seemed like plausible designs. Structures **1–8** were intuitively designed to present side-chains in orientations that resemble peptides and proteins *in general*, not biased towards any particular secondary structure. The putative helical mimic, oxopiperazine **A** designed by Arora *et al.*,[7] was also included for comparison.

Results and discussion

Data accumulated from EKO analyses

Fig. 3a plots the number of EKO alignments (referred to as "hits" later for simplicity) with RMSD < 0.35 Å for each hypothetical mimic, averaged over all stereoisomers. Mimic **1** gave significantly more hits than any other mimic. Mimics **1–4** gave more hits than our "control" **A**, whereas **5–8** gave less. Thus, **1–4** and **A** are fine interface mimics, and mimic **1** is a privileged design.

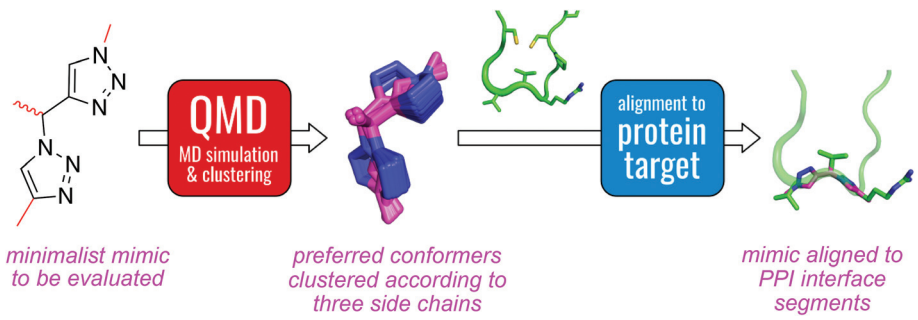


Fig. 1 Workflow of Exploring Key Orientations (EKO).

Fig. 3b breaks down the interface overlays according to stereochemistry of the mimics. All the mimics in Fig. 2 hypothetically could be obtained *via* synthesis from amino acids. It is tempting to assume the corresponding LLL-configurations should give the most overlays on PPI interfaces since natural proteins are derived exclusively from L-amino acids, but Fig. 3b shows this is *not* a valid assumption. Different stereochemical configurations can compensate for the unnatural backbones of minimalist mimics relative to the parent polyamide systems, to place side-chains in favorable orientations.

We set out to rationalize the observations outlined above, but first it was necessary to test if they were just artifacts of the relative flexibilities of the mimics. The considerations used to do this is described in the following section.

The influence of relative flexibility on the EKO data

There are two simple quantitative indicators of flexibilities of minimalist mimics; both are indicated in Fig. 2. The first is "significant degrees of freedom", defined by the number of freely-rotating σ-bonds and by tertiary amides that flip between *cis* and *trans* conformations. The second indicator of conformational flexibility is based on assessment of preferred conformations generated by QMD simulation. The conformers are clustered based on Cα–Cβ orientations of the three side chains at the end of each QMD run. More flexible molecules tend to give more clusters in this process.

Mimics **A** and **1–8** were chosen partly because, in our estimation, they are more conformationally constrained than peptides; however, it is necessary to be mindful of their *relative* flexibilities. Flexible compounds may have more preferred conformers, thus have a better chance of aligning themselves with interface segments. Consequently, EKO evaluations are biased towards more flexible compounds, provided all other factors are equal, simply because a "match" is more likely to be found by screening more conformational partners. By examining the data, we found this was not the dominant consideration for minimalist mimic design. The following discussion elaborates on this conclusion.

Mimics **4**, **6** and **8** have both the most significant degrees of freedom in the series, and number of conformational clusters; these are the most flexible mimics. Conversely, **1**, **5**, and **7**, are the least flexible ones according to the two indicators discussed above. Comparing to Fig. 3 reveals the *most* flexible mimic (**8**) gave the *least* hits while (**1**), being one of the *least* flexible ones, gave significantly *more* hits than any other; this is opposite to one might expect if the numbers of overlays at interfaces segments were governed predominantly by conformational flexibility. Structural parameters other than flexibility must be dominant in determining the numbers of hits since there is no apparent correlation between compound flexibility and number of hits. That conclusion cleared the way to search for other structural features that favor good minimalist mimic designs.

Preferred structural characteristics for interface mimicry

The privileged interface mimic **1** has the same "chain periodicity" as three continuous amino acids, *i.e.* each side-chain is separated by two atoms in the main-chain. In the most stable conformation detected after QMD simulation, the sequential Cα–Cα atoms were ~3.7 Å apart (Fig. 4a), whereas for a peptide in most ideal secondary structures they are only slightly longer at ~3.8 Å. This mimic has three chiral centers and no planar

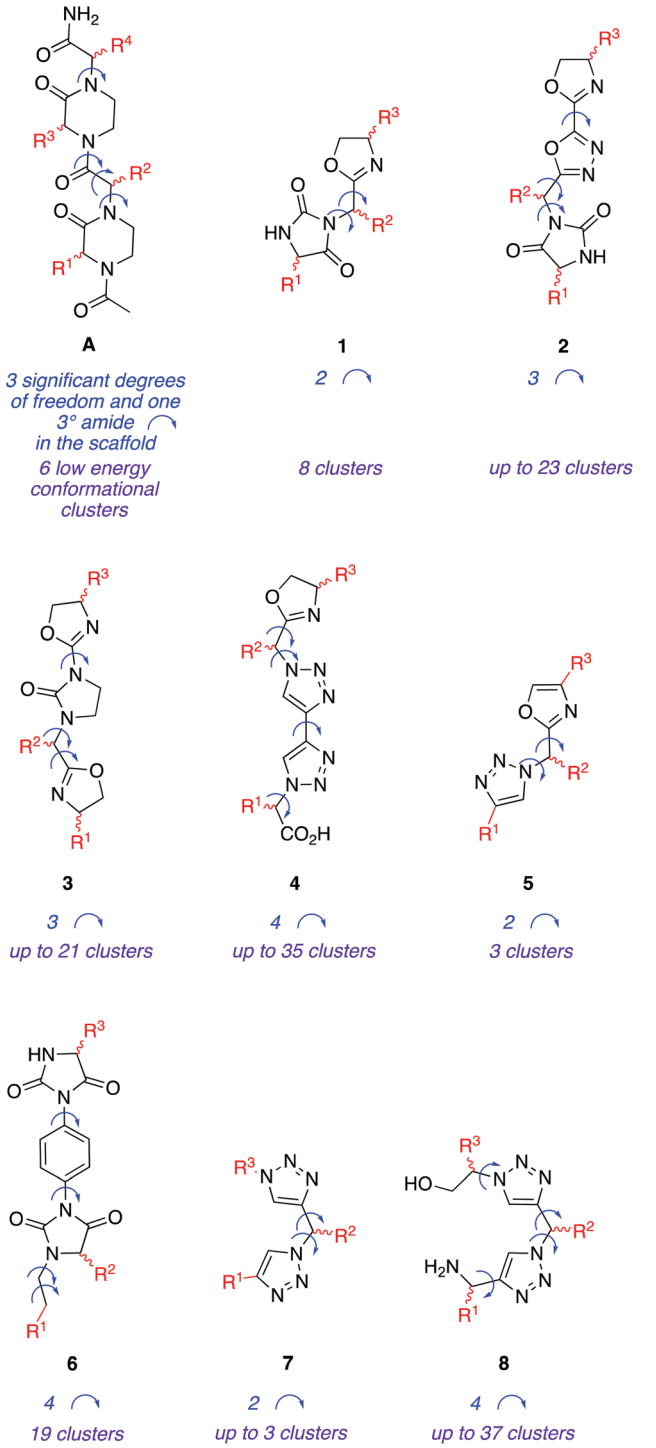


Fig. 2 Minimalist mimics featured in this study. Two indicators of compound flexibility are annotated below each structure. Since EKO was developed to evaluate minimalist mimics with three side-chains, we used only R^1, R^2, and R^4 of **A** in this work.

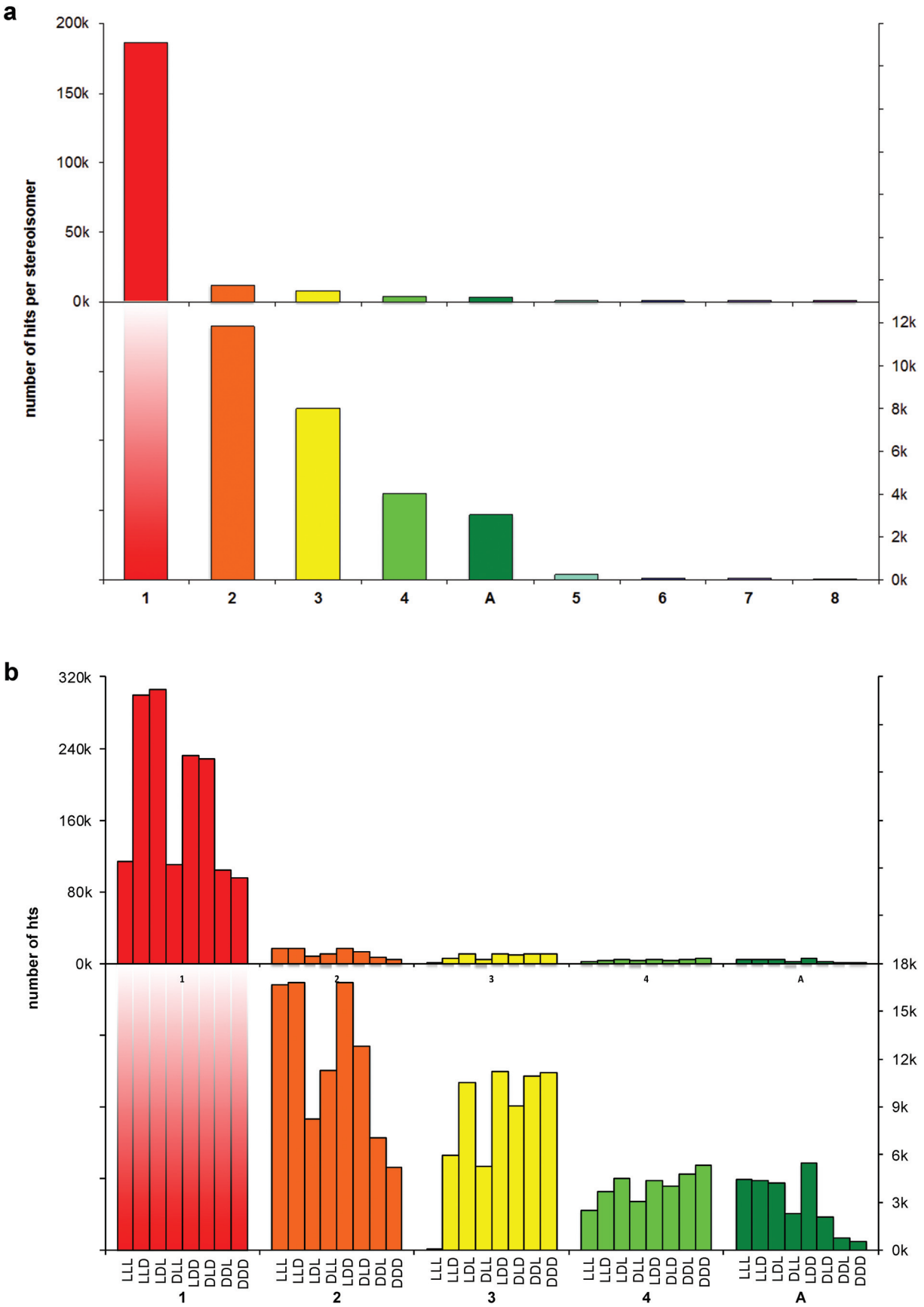


Fig. 3 Number of EKO hits with RMSD < 0.35 Å: (a) averaged over all stereoisomers for **1–8** and **A**; (b) breakdown to each stereoisomer for **1–4** and **A**.

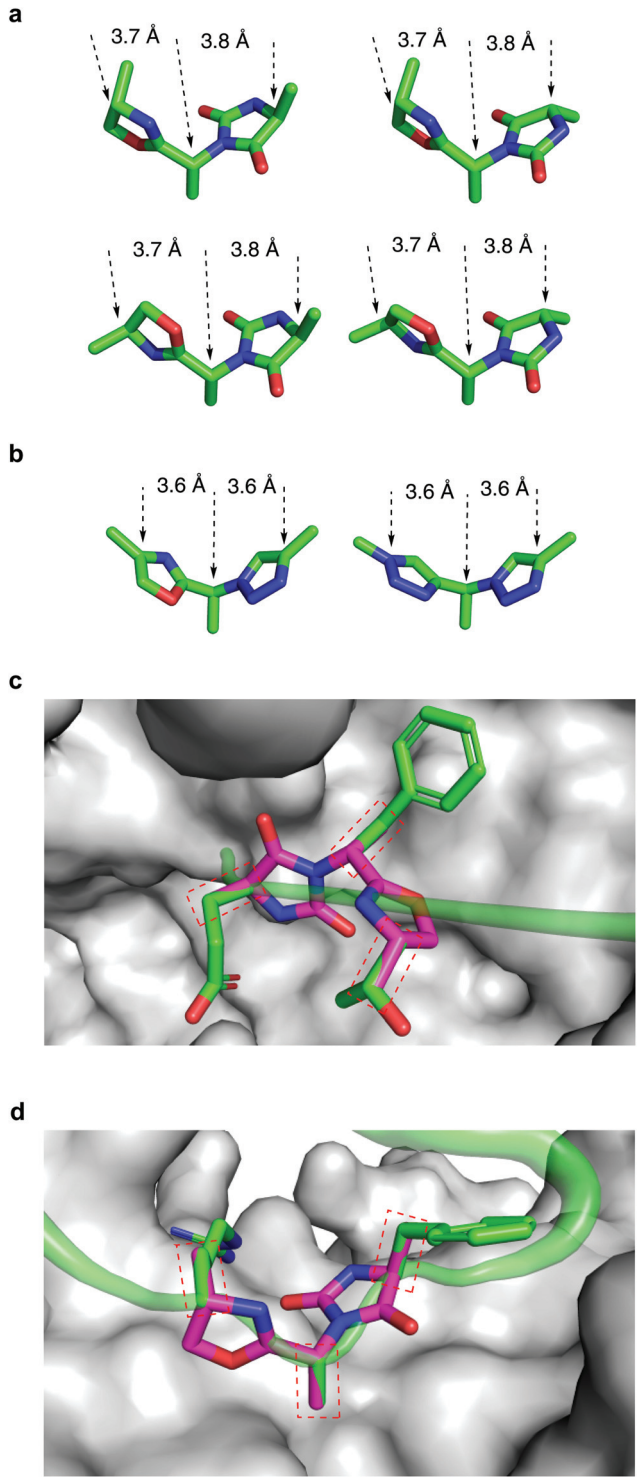


Fig. 4 (a) Cα–Cα separations in LDL-**1aaa**; four out of eight conformational clusters are shown to illustrate; data are the same for the other stereoisomers of **1**. (b) Cα–Cα separations in mimic **5** (left) and **7** (right); only one conformational cluster is shown for each mimic. (c) LDL-**1aaa** aligned on Thr34, Phe35, and Glu36 of yeast RNA polymerase II subunit RPB11 (PDB 1TWF, RMSD 0.10 Å). (d) LDL-**1aaa** aligned on Arg150, Ala151, and Phe152 of cyclin-dependent kinase 2 (PDB 2JGZ, RMSD 0.05 Å). The proteins (or protein subunits) that **1** aligned on are shown as green ribbons, while the partner proteins were shown as grey solvent-accessible surfaces. Only a small segment of each green ribbon is shown for clarity.

heterocyclic rings, so it is expected to extend chains in three dimensions that will vary widely amongst the eight (2^3) possible stereoisomers.

Mimic **1** is unusual because minimalist mimics that display three side-chains in orientations that align with three *consecutive* peptidic side-chains are rare in the literature. The only well-known example of this is the Smith–Hirschmann β-strand mimic.[20,21] Two examples of **1** aligned on actual PPI interfaces are shown in Fig. 4c and d. The first shows how LDL-**1aaa** aligns on a antiparallel β-sheet of an RNA polymerase subunit

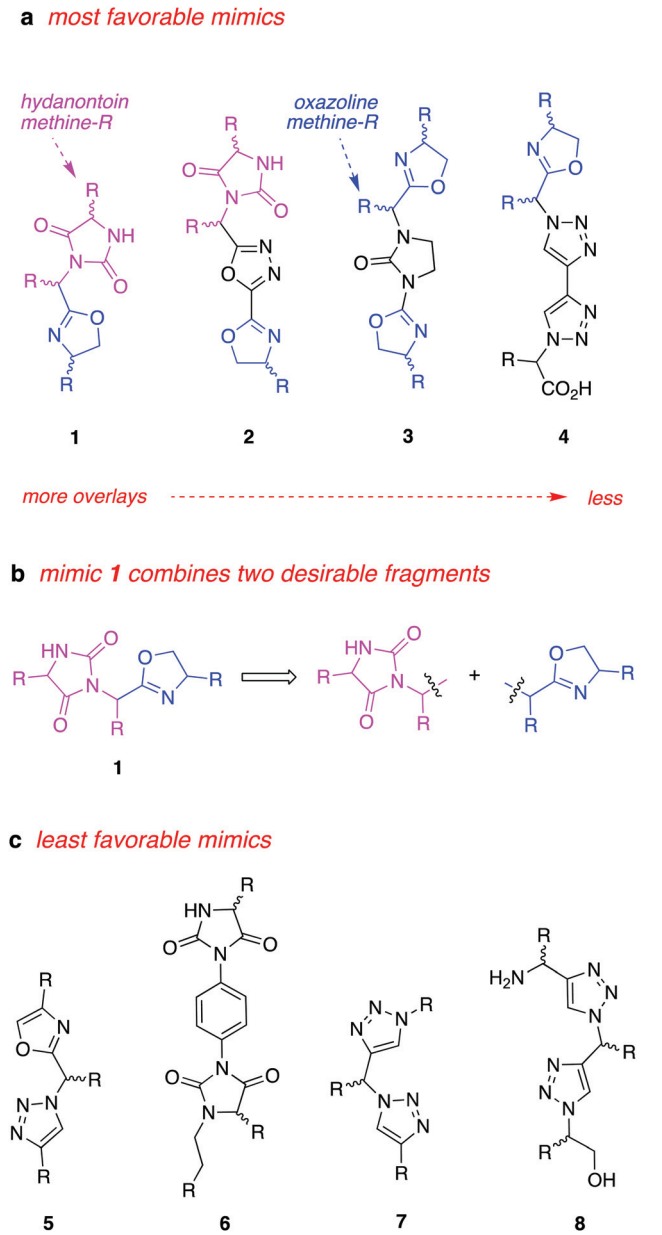


Fig. 5 (a) Mimics most likely to align on PPI interfaces contain oxazoline or hydantoin R-methine fragments. (b) Privileged mimic **1** features oxazoline and hydantoin R-methine fragments. (c) Less favored mimics for interface mimicry do not contain oxazoline or hydantoin R-methine fragments, and are richer in planar aromatic structures.

at Thr34, Phe35, and Glu36. In the second, the same compound aligns on a loop in cyclin-dependent kinase 2 (CDK2) at Arg150, Ala151, and Phe152, interfaced with cyclin-B1. In both cases, mimic **1** aligns on three consecutive residues.

Differences between mimics **1**, **5** and **7** are interesting because these scaffolds are structurally similar but give significantly different number of overlays on PPI interfaces. Each of these compounds has two five membered rings joined by a CH(R) group, and the relative disposition of side-chains resembles continuous sets of three amino acids. Nevertheless, **1** is an incomparably better mimic of interface segments than **5** and **7** (from Fig. 3). We considered three possible reasons for this: (i) separation of Cα–Cα distances in the preferred conformers; (ii) stereochemical diversity; and, (iii) structural topography. Survey of favored conformers of **5** and **7** revealed their Cα–Cα distances tend to be ~3.6 Å (Fig. 4b); this is shorter than for mimic **1** and for tripeptides (~3.7 and 3.8 Å, respectively), but this difference does not seem significant enough to account for the disparity, so (i) is likely not the dominant factor. Parameters (ii) and (iii) are related, but not identical. By "stereochemical diversity" we mean the numbers of stereoisomers available in the data set. However, we disfavor this as being a dominant factor because the data in Fig. 3b is plotted per stereoisomer, and still **1** gives far more close overlays than **5** and **7**. "Structural topography" in the context of (iii) means the overall shape of the molecular scaffold. Of **1**, **5** and **7**, the latter two are composed solely of planar, aromatic rings, and only have one chiral center each. It appears the local planarity of mimics **5** and **7** is not conducive to interface mimicry, but the topography of **1** is.

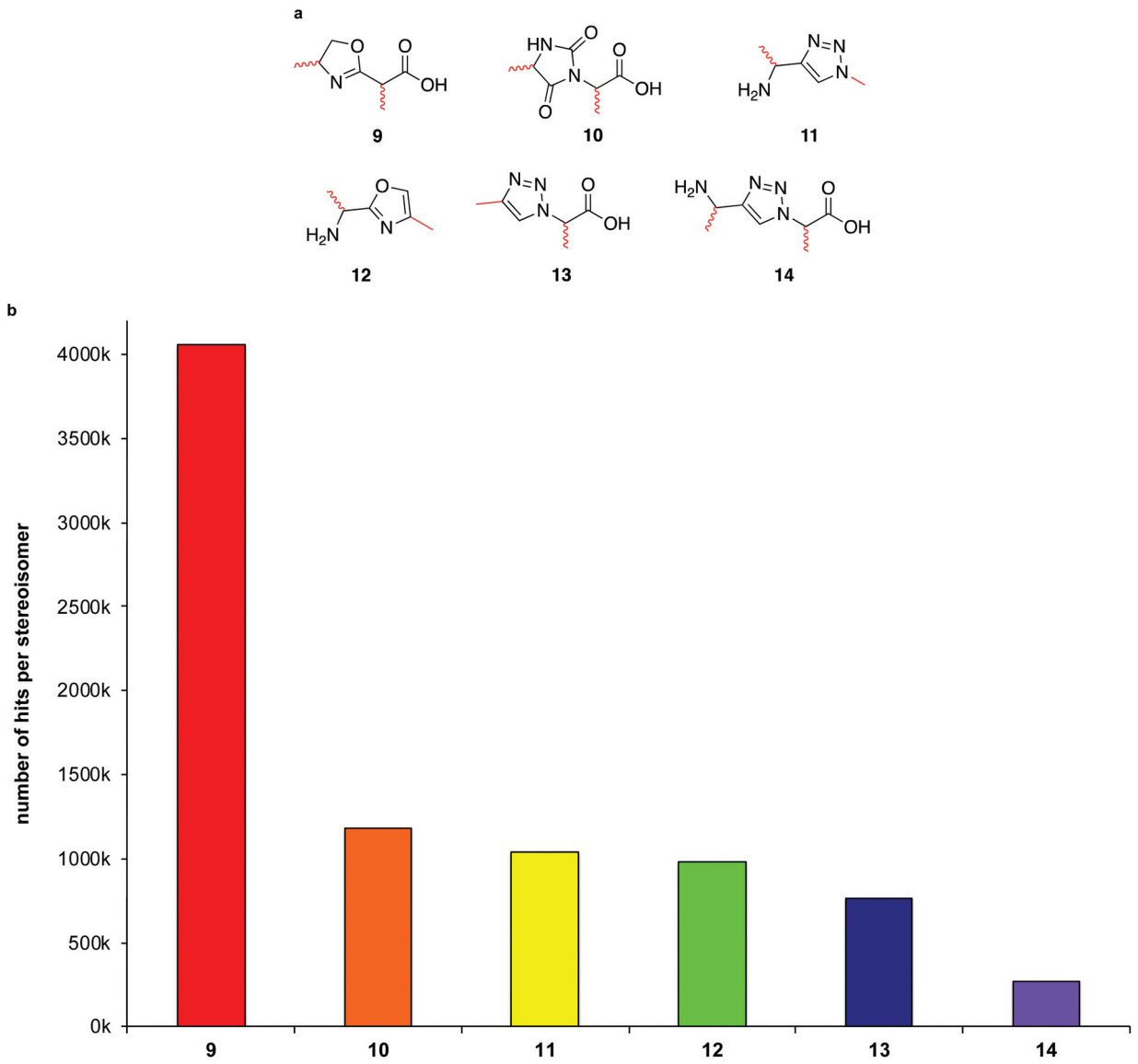


Fig. 6 (a) Dipeptide mimics **9–14** that were evaluated on >240 000 PPI interfaces using the same procedure described above for tripeptide mimics. (b) Average number of hits per stereoisomer are shown.

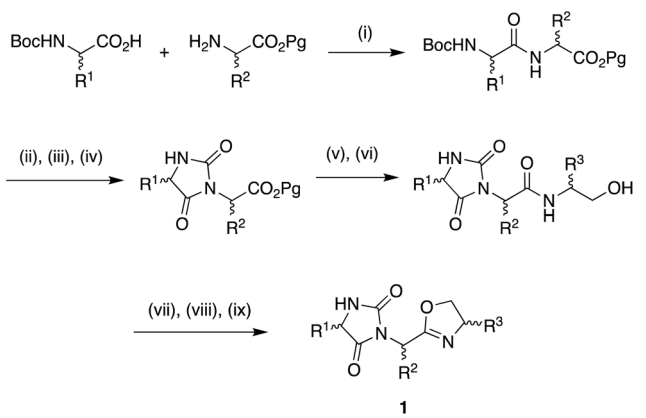
The observations outlined above indicate that there is something special about some molecular topographies that are inherently conducive to interface mimicry. Review of the structures reveal that the four most favorable mimics all contain one of two "dipeptide mimic fragments", based on hydantoin or oxazoline fragments in Fig. 5a, and colored purple and blue respectively. Consequently, we decided to analyze these fragments in more detail.

Analysis of dipeptide building blocks

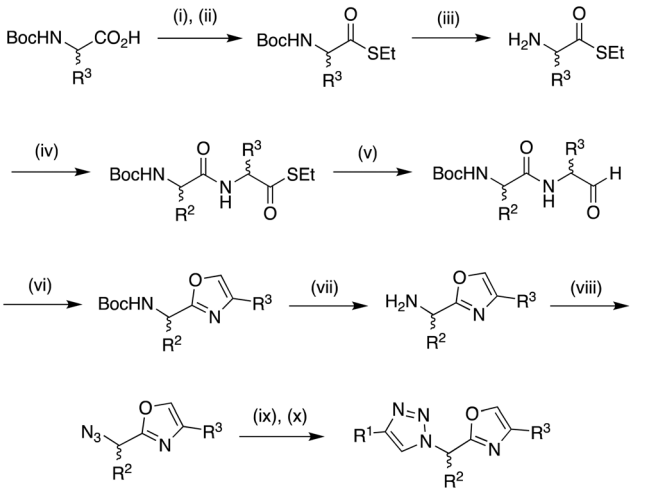
We decided to compare fragments **9–14** (Fig. 6a) that most closely correspond to adjacent amino acids in mimics **1–8**. Specifically, we were interested in how frequently preferred conformations of these overlay on our database of >240 000 crystallized PPI interfaces. The data collected is shown in Fig. 6b.

The oxazoline fragment **9** was superior to the hydantoin **10**, while fragments **10**, **11**, and **12** gave similar numbers of hits to each other. Somewhat surprisingly then, even though the hydantoin **10**, like amino acids, has more three-dimensional features than the flat triazole **11** and oxazole **12**, that did not make a significant difference.

Fragment **14** was featured in our work on minimalist mimics,[22–24] and in "triazolamers" by Arora *et al.*[25–27] However, it turned out to be the *least* suitable of the dipeptide fragments considered based on EKO evaluation. In the light of previous discussion on privileged mimic **1**, it is now evident that the two side chains in this fragment do not share the same "chain periodicity" of natural peptides as those in **1** does: they are too far apart to mimic i, $i + 1$ side-chain combinations, yet too close to align on i, $i + 2$. When samples of the hits for **14** were visualized, they tended to be on non-adjacent side-chains, *i.e.* the two side-chains align to two residues in different strands in a β-sheet, or different helices in a coiled coil, *etc.*, usually spanned by more than a hundred residues.



Scheme 1 Synthesis of mimic **1**. (i) EDC, HOBt, NMM; (ii) deprotection, see ESI;† (iii) 4-$NO_2C_6H_4OCOCl$, $NaHCO_3$; (iv) Et_3N; (v) deprotection; (vi) EDC, HOBt, NMM, $H_2NCH(R^3)CH_2OH$; (vii) MsCl, DMAP, Et_3N; (viii) iPr_2NEt, DMAP, pyridine; (ix) deprotection, if side chains contain protecting groups. Pg, protecting group.



Scheme 2 Synthesis of mimic **5**. (i) EtOCOCl, Et_3N; (ii) EtSH, DMAP; (iii) deprotection; (iv) EDC, HOBt, NMM, $BocHNCH(R^2)CO_2H$; (v) 10% Pd/C, Et_3SiH; (vi) C_2Cl_6, PPh_3, Et_3N; (vii) deprotection; (viii) TfN_3, $CuSO_4$, K_2CO_3; (ix) $R^1C{\equiv}CH$, Cu, $CuSO_4$; (x) deprotection, only if side chains contain protecting groups.

Synthesis of mimics 1 and 5

We have established synthesis routes of mimics **1** and **5** using amino acids as starting materials (Schemes 1 and 2). To show these mimics are easily accessible, six and seven compounds with actual amino acid side chains were synthesized for mimics **1** and **5**, respectively. See ESI† for detailed synthesis procedures and characterization data.

Conclusion

Insight into general design criteria can be gained by comparing preferred conformations of typical minimalist mimic designs with side-chain residues in a very large number of PPI interfaces using EKO. The overriding conclusion from this work on **1–8** is that oxazoline methine-fragments are excellent dipeptide mimics, hydantoin methine-fragments are good ones, and the combination of these in one molecule gives the privileged structure **1**. Mimic **1** gives vastly more hits than any other mimics we conceived, likely because it combines two favorable dipeptide mimics, with optimal Cα–Cα separation, 3D topology, and structural flexibility. These principles can be extended to designs of other privileged dipeptide mimics, which can be then combined to give larger minimalist mimics. An excellent dipeptide mimic lays the foundation for design of effective minimalist mimics containing three side-chains. It follows that these will only be useful to overlay on i, $i + 1$, $i + n$ peptide fragments in interface segments: mimicry of amino acids disposed at more distal separations will require a different focus on that problem.

In general, minimalist mimics should be able to present amino acid side chains at separations that resemble peptides in reasonable conformations. Orientations vectors that are

allowed for these side-chains in the preferred conformations should encompass those presented by peptides in favorable φ,ψ-dispositions. There is a tension in design criteria because minimalist mimics must also have only a few significant degrees of freedom (ideally three or less) to avoid high entropic penalties on binding proteins, but less flexibility favors overlay on less structures, all other factors being equal. Allowing flexibility to obtain a good fit is less important than purposeful design of minimalist mimics based on other factors; some of the least favored mimics studied here are the most flexible. We maintain that *purposeful* design of minimalist mimics cannot be achieved by chemical intuition alone; it is a process that requires molecular dynamic simulations and an algorithm to facilitate systematic alignment procedures. For minimalist mimics containing amino acid derived fragments, L-stereochemistry is *not* vital, so investment in D-building blocks can be well-justified when making libraries of minimalist mimics derived from amino acids.

Conflicts of interest

The authors declare no competing financial interests.

Acknowledgements

Financial support for this project was provided by the National Science Foundation (CHE-1608009), National Institutes of Health (HL126346), The Robert A. Welch Foundation (A-1121), DoD BCRP Breakthrough Award (BC141561), and CPRIT (RP150559 and RP170144). We thank Dr Lisa M. Perez for useful discussions. NMR instrumentation at Texas A&M University was supported by a grant from the National Science Foundation (DBI-9970232) and the Texas A&M University System.

References

1 L.-G. Milroy, T. N. Grossmann, S. Hennig, L. Brunsveld and C. Ottmann, *Chem. Rev.*, 2014, **114**, 4695–4748.
2 M. A. Kuenemann, O. Sperandio, C. M. Labbe, D. Lagorce, M. A. Miteva and B. O. Villoutreix, *Prog. Biophys. Mol. Biol.*, 2015, **119**, 20–32.
3 S. Milhas, B. Raux, S. Betzi, C. Derviaux, P. Roche, A. Restouin, M.-J. Basse, E. Rebuffet, A. Lugari, M. Badol, R. Kashyap, J.-C. Lissitzky, C. Eydoux, V. Hamon, M.-E. Gourdel, S. Combes, P. Zimmermann, M. Aurrand-Lions, T. Roux, C. Rogers, S. Muller, S. Knapp, E. Trinquet, Y. Collette, J.-C. Guillemot and X. Morelli, *ACS Chem. Biol.*, 2016, **11**, 2140–2148.
4 H. Yin and A. D. Hamilton, *Protein Secondary Structure Mimetics as Modulators of Protein-Protein and Protein-Ligand Interactions, in Chemical Biology*, Wiley-VCH-VCH Verlag, 2008.
5 E. Ko, J. Liu and K. Burgess, *Chem. Soc. Rev.*, 2011, **40**, 4411–4421.
6 E. Ko, J. Liu, L. M. Perez, G. Lu, A. Schaefer and K. Burgess, *J. Am. Chem. Soc.*, 2011, **133**, 462–477.
7 P. Tosovska and P. S. Arora, *Org. Lett.*, 2010, **12**, 1588–1591.
8 B. P. Orner, J. T. Ernst and A. D. Hamilton, *J. Am. Chem. Soc.*, 2001, **123**, 5382–5383.
9 M. K. P. Jayatunga, S. Thompson and A. D. Hamilton, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 717–724.
10 L. L. Conte, C. Chothia and J. Janin, *J. Mol. Biol.*, 1999, **285**, 2177–2198.
11 E. Ko, A. Raghuraman, L. M. Perez, T. R. Ioerger and K. Burgess, *J. Am. Chem. Soc.*, 2013, **135**, 167–173.
12 M. Aeluri, S. Chamakuri, B. Dasari, S. K. R. Guduru, R. Jimmidi, S. Jogula and P. Arya, *Chem. Rev.*, 2014, **114**, 4640–4694.
13 M. J. Perez de Vega, M. Martin-Martinez and R. Genzalez-Muniz, *Curr. Top. Med. Chem.*, 2007, **7**, 33–62.
14 A. M. Watkins and P. S. Arora, *Eur. J. Med. Chem.*, 2014, **94**, 480–488.
15 M. Pelay-Gimeno, A. Glas, O. Koch and T. N. Grossmann, *Angew. Chem., Int. Ed.*, 2015, **54**, 8896–8927.
16 B. M. Pettitt, T. Matsunaga, F. Al-Obeidi, C. Gehrig, V. J. Hruby and M. Karplus, *Biophys. J.*, 1991, **60**, 1540–1544.
17 S. D. O'Connor, P. E. Smith, F. Al-Obeidi and B. M. Pettitt, *J. Med. Chem.*, 1992, **35**, 2870–2881.
18 D. Xin, A. Holzenburg and K. Burgess, *Chem. Sci.*, 2014, **5**, 4914–4921.
19 J. Taechalertpaisarn, B. Zhao, X. Liang and K. Burgess, *J. Am. Chem. Soc.*, 2018, **140**, 3242–3249.
20 A. B. Smith III, A. K. Charnley and R. Hirschmann, *Acc. Chem. Res.*, 2011, **44**, 180–193.
21 W. A. Loughlin, J. D. A. Tyndall, M. P. Glenn and D. P. Fairlie, *Chem. Rev.*, 2004, **104**, 6085–6118.
22 Q. Shi, A. T. Nguyen, Y. Angell, D. Deng, C.-R. Na, K. Burgess, D. D. Roberts, F. C. Brunicardi and N. S. Templeton, *Gene Ther.*, 2010, **17**, 1085–1097.
23 D. Chen, F. Brahimi, Y. Angell, Y.-C. Li, J. Moscowicz, H. U. Saragovi and K. Burgess, *ACS Chem. Biol.*, 2009, **4**, 769–781.
24 Y. Angell and K. Burgess, *Chem. Soc. Rev.*, 2007, **36**, 1674–1689.
25 A. L. Jochim, S. E. Miller, N. G. Angelo and P. S. Arora, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 6023–6026.
26 N. G. Angelo and P. S. Arora, *J. Org. Chem.*, 2007, **72**, 7963–7967.
27 N. G. Angelo and P. S. Arora, *J. Am. Chem. Soc.*, 2005, **127**, 17134–17135.