For reprint orders, please contact: reprints@future-science.com

Recent advances in the chemical synthesis of insulin and related peptides

Florence M Brunel¹ , John P Mayer² & Richard D DiMarchi^{*,1,3}

¹Department of Chemistry, Novo Nordisk Research Center, Indianapolis, IN 46241, USA

²Molecular, Cellular & Developmental Biology, University of Colorado, Boulder, CO 80309, USA

³Department of Chemistry, Indiana University, Bloomington, IN 47405, USA

*Author for correspondence: rdimarch@indiana.edu

“A survey of the recently reported methodologies reveals a wide range of novel chemical methods as well as refinements to some of the classical approaches.”

First draft submitted: 14 January 2020; Accepted for publication: 16 January 2020; Published online: 17 March 2020

Keywords: chain combination • chemical synthesis • directed oxidation • insulin • single-chain precursor

Insulin is a substance that has been of central importance in the treatment of diabetes as well as advancements in molecular metabolism and synthetic peptide chemistry. Its primary structure consists of a compact, two-chain peptide including a 21 amino acid A-chain with a single internal disulfide bridge (A6–A11) cross-linked to a 30 amino acid B-chain via two interchain disulfides (A7–B7 and A20–B19). In addition to the six invariant cysteines (A6, A7, A11, A20, B7 and B19), phylogenetic analysis indicates high conservation of several aliphatic residues (Leu A13, A16, B6, B11, B15 and Val B18) among the vertebrate insulins. In total, these 12 amino acids represent nearly one-quarter of the native hormone. The first chemical synthesis of insulin constituted a landmark global event, and after 50 years of steady progress, the last decade has proven to be a period of exceptional accomplishment. The present perspective highlights advances reported in the last four years while crediting earlier contributions which are referenced in more comprehensive reviews published in this most recent period [1–4].

While insulin analogs optimized for pharmacokinetic performance are now well established in clinical practice, they are all produced commercially by biosynthetic methods. Chemical synthesis offers unmatched versatility in peptide optimization and may prove essential to realization of the next class of insulin therapeutics with enhanced safety and more convenient forms of delivery. In this regard, insulin synthesis has historically served as a performance benchmark for a host of novel chemistries including enzyme-assisted semi-synthesis, regioselective disulfide bond formation and ligation of protected and native peptide fragments. The intricate disulfide pattern, and the inherent hydrophobic nature of the A-chain are the most prominent obstacles that have served to compromise synthetic efficiency [1,2]. Several chemical methodologies for the introduction of non-native elements enhancing physical and biological properties such as isosteres of the native disulfides (thioether, dicarba and diselenide), as well as the installation of additional disulfides have been recently reported. Additionally, enzymatically and chemically labile connecting-peptide surrogates that demonstrate increased synthetic versatility in structure-function studies of insulin and related peptides have also emerged. In this regard, where prior reports focused on structurally related, therapeutically relevant human, bovine and porcine insulins, the more recent work has targeted a more diverse set of insulins (guinea pig, snail, and virus derived), as well as other members of the insulin superfamily (relaxins and INSLPs).

Two basic chemical approaches have historically been used to synthesize insulin analogs. The first employs chain combination carried out using either a site-directed approach or under optimized aqueous buffer conditions. The second general approach utilizes proinsulin-like single chain insulin (SCI) precursors that are converted to two-chain products via enzymatic and more recently chemical methods. A survey of the recently reported methodologies reveals a wide range of novel chemical methods as well as refinements to some of the classical approaches.

Approach 1

The first route in insulin total synthesis employed random chain recombination of native A and B chains in either their reduced or *S*-sulfonate forms to initially produce the hormone in yields of not much more than 1%. Amazingly, through painstaking optimization of the reaction conditions, a highly polished version of this chemically undirected approach to disulfide formation was employed in the earliest commercial biosynthesis of human insulin at Lilly Research Laboratories [5]. The fact that the oxidative recombination can proceed efficiently in the absence of a connecting peptide demonstrates the inherent propensity of the A and B chains to fold into a thermodynamically stable conformation, one favoring the native disulfide pairing. The preference for the native conformation is presumed to be driven by noncovalent forces and specifically the highly conserved aliphatic amino acids at the core of the insulin structure. A further refined version of this early method was recently used by Metanis and coworkers to prepare an A6–A11 diselenide insulin analog [6]. While selenocysteine is a close isostere of cysteine, the selenol group has a lower pK_a and a lower oxidation potential than sulfur. The strategy involved synthesis of the A-chain with a preformed A6–A11 diselenide bond followed by chain combination with B-chain *S*-sulfonate. The authors suggested that the initial formation of the A6–A11 intramolecular bond facilitated the proper folding of the interchain disulfides resulting in a more rapid formation and higher yield of final product. The novel analog displayed insulin-like potency, and higher stability to reduction and Glu-C proteolysis. A diselenide substitution of the A7–B7 bridge of insulin was reported by Arai *et al.* [7] who used random recombination of the monosubstituted seleno A- and B-chains to achieve yields as high as 34%. This analog also demonstrated increased protease resistance, notably to insulin degrading enzyme. Its biological activity was comparable to native insulin. The diselenide bond might provide a path to sustained pharmacokinetics and potentially enhanced potency should they prove resistant to insulin-mediated intracellular proteolysis and thiol-mediated reductive bond cleavage. Separately, Arai's group published a folding method they termed as 'oxidative native chain assembly' (NCA)-based upon an extensive optimization of several reaction parameters including temperature, pH, stoichiometry and a few nonstandard additives. The optimized conditions of pH 10, in 20% ethylene glycol at -10°C with a catalytic amount of protein disulfide isomerase provided insulin in a yield as high as 48%, based upon a limiting amount of B-chain [8]. The study revealed an important mechanistic insight in the oxidative folding pathway, which confirmed the formation of the A6–A11 disulfide as the initial step to high efficiency synthesis.

While the above methods illustrate the potential to perfect classical approaches, more novel chain combination methods have been reported that use orthogonally protected cysteine pairs to ensure correct disulfide combination. This well-established strategy has been validated with a variety of protecting group schemes and continues to evolve as new cysteine protecting groups are introduced. Its versatility was recently demonstrated by Chou's group to synthesize a novel, aggregation-resistant insulin analog with an additional fourth disulfide at A22–B22 [9]. DiMarchi's group also utilized sequential disulfide bond formation to synthesize three insulin analogs with an additional disulfide (A10–B4, A10–B5 and A10–B6), for comparative purposes with snail insulin, which possesses a fourth native disulfide and a snail/human hybrid analog. The fourth and final disulfide bond was formed by dimethyl sulfoxide/trifluoroacetic acid (DMSO/TFA) oxidation of two tBu-protected cysteines [10]. Directed disulfide bond formation has also recently enabled the synthesis of several novel insulin analogs containing disulfides isosteres. For example, Karas and coworkers substituted the A6–A11 disulfide with a thioether by replacing cysteine with cystathionine [3,11]. The intrachain A6–A11 cystathionine containing A-chain was crosslinked through a B19*S*-pyridyl (*S*-Pyr) activated B-chain and a free cysteine at A20. The final A7–B7 disulfide bond was formed through iodine oxidation of the corresponding (*S*-Acm) protected cysteines. van Lierop and coworkers utilized a similar strategy to introduce a reductively-stable dicarba linkage as a replacement for the A6–A11 disulfide [12]. The dicarba isostere was formed through a ring-closing metathesis, followed by a combination of the A7(*S*-Pyr) activated A-chain with a B7(SH) and B19(*S*-Acm) B-chain. The A20-B19 disulfide was subsequently formed by iodine-mediated oxidation of the respective (*S*-Acm) protected cysteines. The poor aqueous solubility and physical handling of the A-chain is well documented and has been effectively addressed by extension with a charged solubility tag, a strategy originally introduced by Wade and co-workers [13]. Recently, Chou's group incorporated a tri-lysine tag in conjunction with a cleavable linker (4,4-dimethyl -2,6-dioxocyclohexylidene)-3-[2-(2 aminoethoxy) ethoxyl] propan-1-ol (Ddae). This 'helping hand' strategy as they termed it improved the physical properties of the A-chain, and resulted in the first synthesis of chicken insulin [14]. The same group took advantage of this method to synthesize a A6–A11 methylene thioacetal insulin [15]. Their disulfide analog demonstrated bioactivity comparable to native

insulin and enhanced structural stability as assessed by decreased propensity to physically aggregate, and resistance to enzyme treatment or thermal denaturation.

Approach 2

The second general approach used in insulin chemical synthesis is characterized by various single chain synthetic intermediates that are convertible to two-chain peptides by enzymatic or chemical methods. The intrinsic attraction in the single-chain approach is the ability to conduct native oxidative intramolecular folding. The limitation is often the requirement to subsequently convert to a two-chain, insulin-like structure. Enzymatic processing is inherently limited by sequence specificity and structural accessibility in the site of cleavage. Most frequently a single, lysine or arginine at the B-chain C-terminus has served this purpose for selective cleavage with LysC or trypsin. In this regard, trypsin was employed in concert with carboxypeptidase B in the first commercial synthesis of biosynthetic human insulin from proinsulin at the Lilly Research Laboratories, demonstrating the ability to selectively process at two sites (B-chain C-terminus and A-chain N-terminus), with minimal cleavage at the single internal ArgB22 site [16]. More complete understanding of the role of the C-peptide in promoting native disulfide formation eventually suggested that the full connecting peptide sequence was unnecessary. Consequently, a linear fusion of the C-terminus of B-chain to the N-terminus of A-chain through a precursor 37 amino acids shorter than proinsulin in which native residues B28–30 are replaced by a single lysine yielded what Zaykov and colleagues termed DesDi insulin [17]. It was found to fold efficiently, even in the case of several historically challenging insulin analogs and to easily convert to the two-chain form by a single-proteolytic step. The versatility of this approach was demonstrated by a comprehensive disulfide scan we completed recently [18]. Employing state-of-the-art resin-based peptide assembly we prepared a series of 49 or 50 residue peptides with no modifications to the amide backbone, in which two Cys(Acm) were strategically introduced at various positions. Twenty insulin analogs in the single-chain DesDi form were prepared and folded with an efficiency comparable to the native sequence. Once enzymatically converted to the two-chain form, the fourth non-native disulfide was installed through iodine-mediated oxidation of the (S-Acm) protected cysteines. Attempts to form the four disulfides simultaneously were less successful, particularly in case of analogs with lower bioactivity. This indicates that once the native insulin structure is formed the molecule retains flexibility to accommodate an additional constraining tether. This is an example of where recent advances in synthetic methodology have enabled our ability to identify analogs, that while fully active in their native three disulfide form, are reduced to 1% of that potency when constrained by a fourth disulfide. These results may offer a basis for the design of a glucose-responsive insulin provided the deactivating disulfide can be replaced with a glucose responsive tether.

While enzymatic conversion remains an important, well established method, recent trends have favored chemical means to convert single-chain synthetic intermediates to two-chain insulins. Chemical conversions provide unique advantages as they not only eliminate the need for an enzyme and a unique proteolytic site, they also enable non-native crosslink-strategies. Such a chemically convertible insulin linker was demonstrated as early as 1976 by Carpenter and associates [19]. Carbonylbis(L-methionine p-nitrophenyl ester) was used to cross-link the A-chain terminal amine to the sidechain of LysB29 in porcine insulin. The linear reduced analog was successfully refolded, and porcine insulin was regenerated through cyanogen bromide cleavage of the two methionine residues anchoring the crosslink. The concept was applied in a more elegant, traceless chemical crosslink by Kent's group in 2010 who designed an 'ester precursor' to synthesize insulin lispro [20]. The synthetic intermediate was assembled by native chemical ligation and the two chains joined by a GluA4–ThrB30 ester bond. Following native folding the two-chain lispro insulin was obtained by selective alkaline saponification of the single ester. Recently, Kent's group further evaluated several methods in constructing the GluA4–ThrB30 ester intermediate to demonstrate that native chemical ligation was the optimal synthetic route [21]. Bode and coworkers have applied their KAHA (α -ketoacid-hydroxylamine) ligation chemistry to the syntheses of the M2 metabolite of glargine as well as mouse, guinea pig and human insulin [22]. The site chosen for the keto-acid/5-oxaproline KAHA ligation was TyrB26–ThrB27 resulting in a homoserine (Hse) residue at B27 following the concomitant O→N acyl shift and traceless removal of the prosthetic C-peptide under alkaline treatment. An additional element of this versatile approach was the use of a base-labile arginine solubilizing handle to enhance A-chain aqueous solubility. The Bode research group subsequently integrated their KAHA and KAT (potassium acyltrifluoroborate) ligation methods to yield a photo-protected hydroxylamine amino acid [23]. Through this synthetic intermediate a series of site-specific glargine-like insulin conjugates were further prepared.

DiMarchi's group has made extensive use of dipeptide extensions that are optimized to cleave via diketopiperazine (DKP) cyclization to selectively remove a non-native crosslink. In the first demonstration, chemically synthesized human A- and B-chains were crosslinked at the analogous sites first employed by Carpenter [19], but now through a novel terephthalaldehyde oxime ligation [24]. The non-native linear precursor was folded and converted to the two-chain form by a two-step aqueous cleavage that employed elevated temperature at pH 7, followed by room temperature alkaline treatment at pH 11 [24]. This synthetic strategy is agnostic to the specific peptide sequence, a feature demonstrated by successful application to two well recognized insulin-like peptides of sizable synthetic difficulty, relaxin-2 and insulin-like peptide-5 [25]. The attraction of the dipeptide-based DKP cyclization as a synthetic step was significantly advanced through the introduction of a traceless, symmetrical linker between the N-termini of the A- and B-chains that was released in a single chemical step [26]. The A- and B-chains were prepared with site-specific isoacyl bonds, that had first been separately demonstrated by Liu and associates in 2014 to improve chain-synthesis. The N-termini were extended beyond the non-native dipeptides with identical linkers that served to enhance solubility but more importantly provided conformational flexibility that facilitated high yield native folding, following ligation. The non-native linker was removed by a 2-h, pH 7 aqueous treatment at 56°C. This approach is unique as it demonstrates for the first time noncanonical tethering through the two N-termini, which aligns the synthesis with conventional methods in solid phase-based synthesis that are used to extend the N-terminus. The versatility of the method was demonstrated by successful application to insulin, lispro insulin, a four-disulfide insulin analog, two insulin penicillamine analogs and relaxin.

Conclusion

The evolution of insulin therapy over the last century can be divided into two periods. The initial phase focused on enhanced purity with an emphasis in obtaining the human sequence which was eventually made possible predominantly by rDNA technology sixty years following the discovery of the hormone. The last 40 years has been a journey largely employing a combination of biosynthetic and semisynthetic methods to discover insulin analogs with more optimal pharmacokinetic properties. So, it seems appropriate to ask whether a third phase focused on enhanced pharmacodynamic performance might be enabled by the newly enriched chemical methods. The advances in synthetic chemistry have greatly expanded our ability to more fully interrogate insulin structure and empower it with enhanced physical and biological functions. In many ways this has been a primary justification for the historical commitment to establishing these synthetic capabilities that we now possess.

Insulin is unquestionably a miraculous substance but a challenging drug with an inherently narrow therapeutic index that requires dose titration. It is also in a relative medicinal sense a complex and fragile macromolecule that requires refrigeration and administration most often by subcutaneous injection. In a more academic sense, there are provocative questions and opportunities that pertain to whether a more tissue-specific insulin, or one that is receptor isoform specific or biased in its intracellular action might constitute a superior therapeutic agent in treatment of insulin-dependent forms of diabetes. Finally, will such chemistry provide a route to discovery of an insulin that might address the public demand for a less costly form of therapy? While each of these questions is worthy of consideration it seems prudent not to overtax novel chemical methods with expectations that can't be immediately fulfilled.

A century of advances in insulin science has established much improved therapy that elevates the level of efficacy and safety that must be exceeded by any further change in insulin structure. An example in how these new methods might be employed can be witnessed most recently by a report from Wade and associates to prepare an insulin analog with an additional cysteine at the N-terminal B-chain that following glycosylation displayed appreciable bioactivity devoid of fibril formation and with improved plasma stability [27]. Alternatively, the previously discussed report from Brunel *et al.* illustrates one approach toward the chemical optimization of the hormone in pursuit of glucose-responsive insulin analogs [18]. Might the discovery of insulin analogs of much enhanced *in vivo* potency emerge? Other peptides such as parathyroid hormone, glucagon-like peptide agonists and calcitonin are therapeutically used at a fraction of insulin's dose. The yearly dose of parathyroid hormone in treatment of maturity onset osteoporosis is approximately the daily dose of insulin in similarly aged, obese patients with diabetes. Potency has relevance to cost of a drug but also to drug delivery where non-injectable forms, such as oral delivery are of exceedingly low bioavailability.

It is an exciting time where insulin synthetic chemistry has advanced to a point where it challenges us to put it to suitable medicinal use. It also presents the prospect in discovery of insulin analogs that might help in a

chemical-biology sense to better define the interaction dynamics with insulin receptor isoforms and subsequent intracellular processing of its complex with engineered ligands.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Financial & competing interests disclosure

F Brunel and R DiMarchi are Novo Nordisk employees. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1. Liu F, Zaykov AN, Levy JJ, Dimarchi RD, Mayer JP. Chemical synthesis of peptides within the insulin superfamily. *J. Pept. Sci.* 22(5), 260–270 (2016).
2. Moroder L, Musiol HJ. Insulin-from its discovery to the industrial synthesis of modern insulin analogues. *Angew Chem. Int. Ed. Engl.* 56(36), 10656–10669 (2017).
3. Hossain MA, Wade JD. Novel methods for the chemical synthesis of insulin superfamily peptides and of analogues containing disulfide isosteres. *Acc. Chem. Res.* 50(9), 2116–2127 (2017).
4. Zaykov AN, Mayer JP, Dimarchi RD. Pursuit of a perfect insulin. *Nat. Rev. Drug Discov.* 15(6), 425–439 (2016).
5. Chance RE, Kroeff EP, Hoffmann JA, Frank BH. Chemical, physical, and biologic properties of biosynthetic human insulin. *Diabetes Care* 4(2), 147–154 (1981).
6. Weil-Ktorza O, Rege N, Lansky S *et al.* Substitution of an internal disulfide bridge with a diselenide enhances both foldability and stability of human Insulin. *Chemistry* 25(36), 8513–8521 (2019).
7. Arai K, Takei T, Okumura M *et al.* Preparation of selenoinsulin as a long-lasting insulin analogue. *Angew. Chem. Int. Ed. Engl.* 56(20), 5522–5526 (2017).
8. Arai K, Takei T, Shinozaki R *et al.* Characterization and optimization of two-chain folding pathways of insulin via native chain assembly. *Commun. Chem.* 1(1), 26 (2018).
9. Xiong X, Blakely A, Karra P *et al.* Novel four-disulfide insulin analog with high aggregation stability and potency. *Chem. Sci.* 11, 195–200 (2020).
10. Wu F, Mayer JP, Gelfanov VM, Liu F, Dimarchi RD. Synthesis of four-disulfide insulin analogs via sequential disulfide bond formation. *J. Org. Chem.* 82(7), 3506–3512 (2017).
11. Karas JA, Patil NA, Tailhades J *et al.* Total chemical synthesis of an intra-A-chain cystathionine human insulin analogue with enhanced thermal stability. *Angew. Chem. Int. Ed. Engl.* 55(47), 14743–14747 (2016).
12. van Lierop B, Ong SC, Belgi A *et al.* Insulin in motion: the A6–A11 disulfide bond allosterically modulates structural transitions required for insulin activity. *Sci. Rep.* 7(1), 17239 (2017).
13. Hossain MA, Belgi A, Lin F *et al.* Use of a temporary “solubilizing” peptide tag for the Fmoc solid-phase synthesis of human insulin glargine via use of regioselective disulfide bond formation. *Bioconjug. Chem.* 20(7), 1390–1396 (2009).
14. Disotuar MM, Petersen ME, Nogueira JM, Kay MS, Chou DH. Synthesis of hydrophobic insulin-based peptides using a helping hand strategy. *Org. Biomol. Chem.* 17(7), 1703–1708 (2019).
15. Zheng N, Karra P, Vandenberg MA *et al.* Synthesis and characterization of an A6–A11 methylene thioacetal human insulin analogue with enhanced stability. *J. Med. Chem.* 62(24), 11437–11443 (2019).
16. Galloway JA, Hooper SA, Spradlin CT *et al.* Biosynthetic human proinsulin. Review of chemistry, *in vitro* and *in vivo* receptor binding, animal and human pharmacology studies, and clinical trial experience. *Diabetes Care* 15(5), 666–692 (1992).
17. Zaykov AN, Mayer JP, Gelfanov VM, Dimarchi RD. Chemical synthesis of insulin analogs through a novel precursor. *ACS Chem. Biol.* 9(3), 683–691 (2014).
18. Brunel FM, Mayer JP, Gelfanov VM *et al.* A disulfide scan of insulin by [3+1] methodology exhibits site-specific influence on bioactivity. *ACS Chem. Biol.* 14(8), 1829–1835 (2019).
19. Busse WD, Carpenter FH. Synthesis and properties of carbonylbis(methionyl)insulin, a proinsulin analogue which is convertible to insulin by cyanogen bromide cleavage. *Biochemistry* 15(8), 1649–1657 (1976).
20. Sohma Y, Hua QX, Whittaker J, Weiss MA, Kent SB. Design and folding of [GluA4(ObetaThrB30)]insulin (“ester insulin”): a minimal proinsulin surrogate that can be chemically converted into human insulin. *Angew. Chem. Int. Ed. Engl.* 49(32), 5489–5493 (2010).

21. Dhayalan B, Mandal K, Rege N *et al.* Scope and limitations of Fmoc chemistry SPPS-based approaches to the total synthesis of insulin lispro via ester insulin. *Chemistry* 23(7), 1709–1716 (2017).
22. Boross GN, Shimura S, Besenius M *et al.* Facile folding of insulin variants bearing a prosthetic C-peptide prepared by alpha-ketoacid-hydroxylamine (KAHA) ligation. *Chem. Sci.* 9(44), 8388–8395 (2018).
23. Boross GN, Schauenburg D, Bode JW. Chemoselective derivatization of folded synthetic insulin variants with potassium acyltrifluoroborates (KATs). *Helv. Chim. Acta* 102(2), e1800214 (2019).
24. Thalluri K, Kou B, Gelfanov V, Mayer JP, Liu F, Dimarchi RD. Biomimetic synthesis of insulin enabled by oxime ligation and traceless “C-peptide” chemical excision. *Org. Lett.* 19(3), 706–709 (2017).
25. Thalluri K, Kou B, Yang X *et al.* Synthesis of relaxin-2 and insulin-like peptide 5 enabled by novel tethering and traceless chemical excision. *J. Pept. Sci.* 23(6), 455–465 (2017).
26. Thalluri K, Mayer JP, Chabenne JR, Gelfanov V, Dimarchi RD. Synthesis of disulfide-rich heterodimeric peptides through an auxiliary N, N-crosslink. *Nat. Commun.* 1(1), (2018).
27. Hossain MA, Okamoto R, Karas JA *et al.* Total chemical synthesis of a non-fibrillating human glycoinsulin. *J. Am. Chem. Soc.* 142(3), 1164–1169 (2020).