NON-CONFIDENTIAL

An efficient algorithm to find matched pairs of a peptide

Noel O'Boyle, Chris de Graaf
noel.oboyle@soseiheptares.com

Disclaimer

The material that follows is a presentation of general background information about Sosei Group Corporation and its subsidiaries (collectively, the “Company”) as of the date of this presentation. This material has been prepared solely for informational purposes and is not to be construed as a solicitation or an offer to buy or sell any securities and should not be treated as giving investment advice to recipients. It is not targeted to the specific investment objectives, financial situation or particular needs of any recipient. It is not intended to provide the basis for any third party evaluation of any securities or any offering of them and should not be considered as a recommendation that any recipient should subscribe for or purchase any securities.

The information contained herein is in summary form and does not purport to be complete. Certain information has been obtained from public sources. No representation or warranty, either express or implied, by the Company is made as to the accuracy, fairness, or completeness of the information presented herein and no reliance should be placed on the accuracy, fairness, or completeness of such information. The Company takes no responsibility or liability to update the contents of this presentation in the light of new information and/or future events. In addition, the Company may alter, modify or otherwise change in any manner the contents of this presentation, in its own discretion without the obligation to notify any person of such revision or changes.

This presentation contains “forward-looking statements,” as that term is defined in Section 27A of the U.S. Securities Act of 1933, as amended, and Section 21E of the U.S. Securities Exchange Act of 1934, as amended. The words “believe”, “expect”, “anticipate”, “intend”, “plan”, “seeks”, “estimates”, “will” and “may” and similar expressions identify forward looking statements. All statements other than statements of historical facts included in this presentation, including, without limitation, those regarding our financial position, business strategy, plans and objectives of management for future operations (including development plans and objectives relating to our products), are forward looking statements. Such forward looking statements involve known and unknown risks, uncertainties and other factors which may cause our actual results, performance or achievements to be materially different from any future results, performance or achievements expressed or implied by such forward looking statements. Such forward looking statements are based on numerous assumptions regarding our present and future business strategies and the environment in which we will operate in the future. The important factors that could cause our actual results, performance or achievements to differ materially from those in the forward looking statements include, among others, risks associated with product discovery and development, uncertainties related to the outcome of clinical trials, slower than expected rates of patient recruitment, unforeseen safety issues resulting from the administration of our products in patients, uncertainties related to product manufacturing, the lack of market acceptance of our products, our inability to manage growth, the competitive environment in relation to our business area and markets, our inability to attract and retain suitably qualified personnel, the unenforceability or lack of protection of our patents and proprietary rights, our relationships with affiliated entities, changes and developments in technology which may render our products obsolete, and other factors. These factors include, without limitation, those discussed in our public reports filed with the Tokyo Stock Exchange and the Financial Services Agency of Japan. Although the Company believes that the expectations and assumptions reflected in the forward-looking statements are reasonably based on information currently available to the Company's management, certain forward looking statements are based upon assumptions of future events which may not prove to be accurate. The forward looking statements in this document speak only as at the date of this presentation and the company does not assume any obligations to update or revise any of these forward statements, even if new information becomes available in the future.

This presentation does not constitute an offer, or invitation, or solicitation of an offer, to subscribe for or purchase any securities. Neither this presentation nor anything contained herein shall form the basis of any contract or commitment whatsoever. Recipients of this presentation are not to construe the contents of this summary as legal, tax or investment advice and recipients should consult their own advisors in this regard.

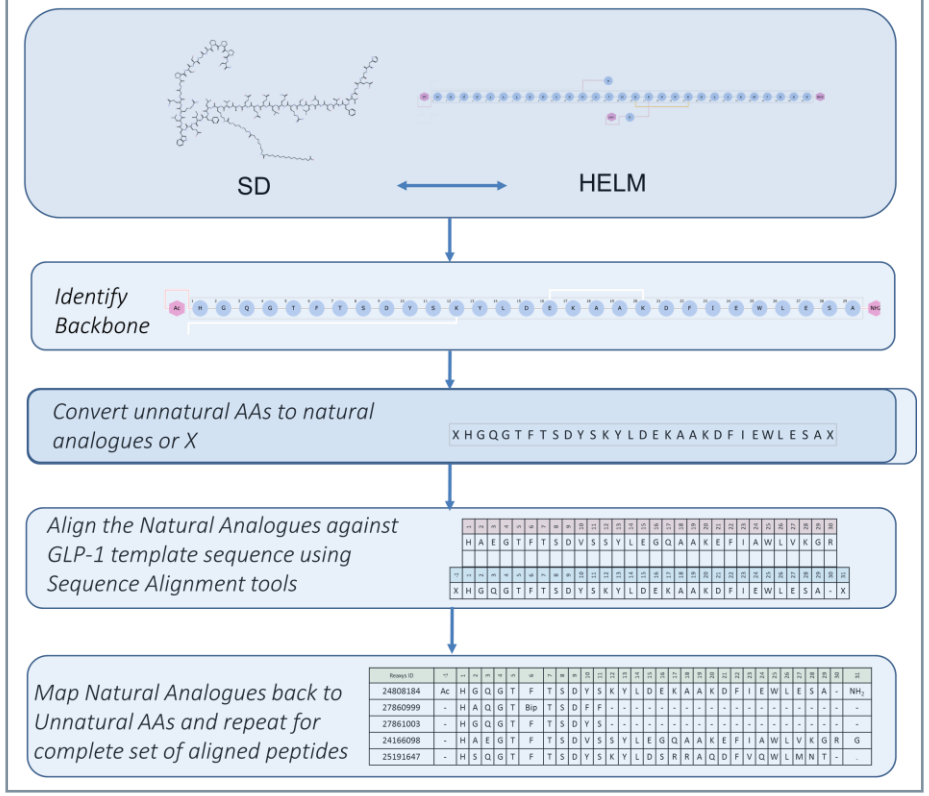
This presentation and its contents are proprietary confidential information and may not be reproduced, published or otherwise disseminated in whole or in part without the Company’s prior written consent. These materials are not intended for distribution to, or use by, any person or entity in any jurisdiction or country where such distribution or use would be contrary to local law or regulation.

This presentation contains non-GAAP financial measures. The non-GAAP financial measures contained in this presentation are not measures of financial performance calculated in accordance with IFRS and should not be considered as replacements or alternatives profit, or operating profit, as an indicator of operating performance or as replacements or alternatives to cash flow provided by operating activities or as a measure of liquidity (in each case, as determined in accordance with IFRS). Non-GAAP financial measures should be viewed in addition to, and not as a substitute for, analysis of the Company's results reported in accordance with IFRS.

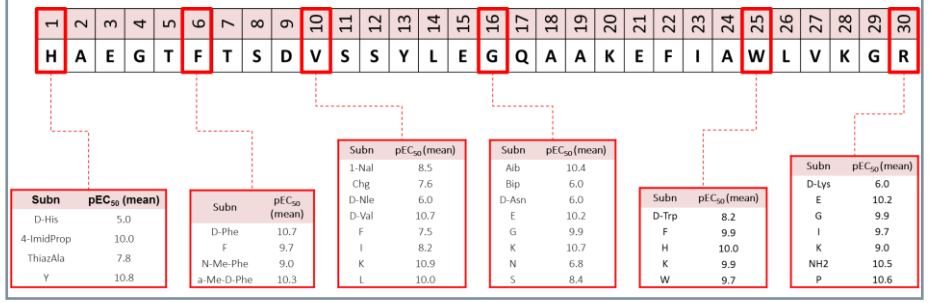
References to "FY" in this presentation for periods prior to 1 January 2018 are to the 12-month periods commencing in each case on April 1 of the year indicated and ending on March 31 of the following year, and the 9 month period from April 1 2017 to December 31 2017. From January 1 2018 the Company changed its fiscal year to the 12-month period commencing in each case on January 1. References to "FY" in this presentation should be construed accordingly.

Structural Cheminformatics based GPCR Peptide Ligand Design

Sequence Alignment and Annotation workflow



Residue level analysis of Peptide properties

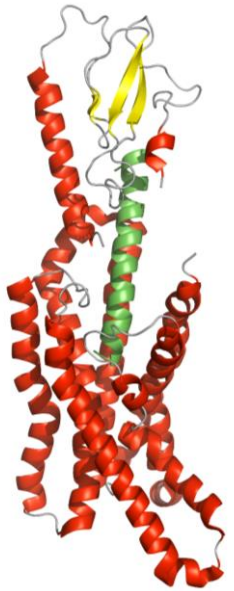


HELM

Chemaxon Biomolecular Toolkit

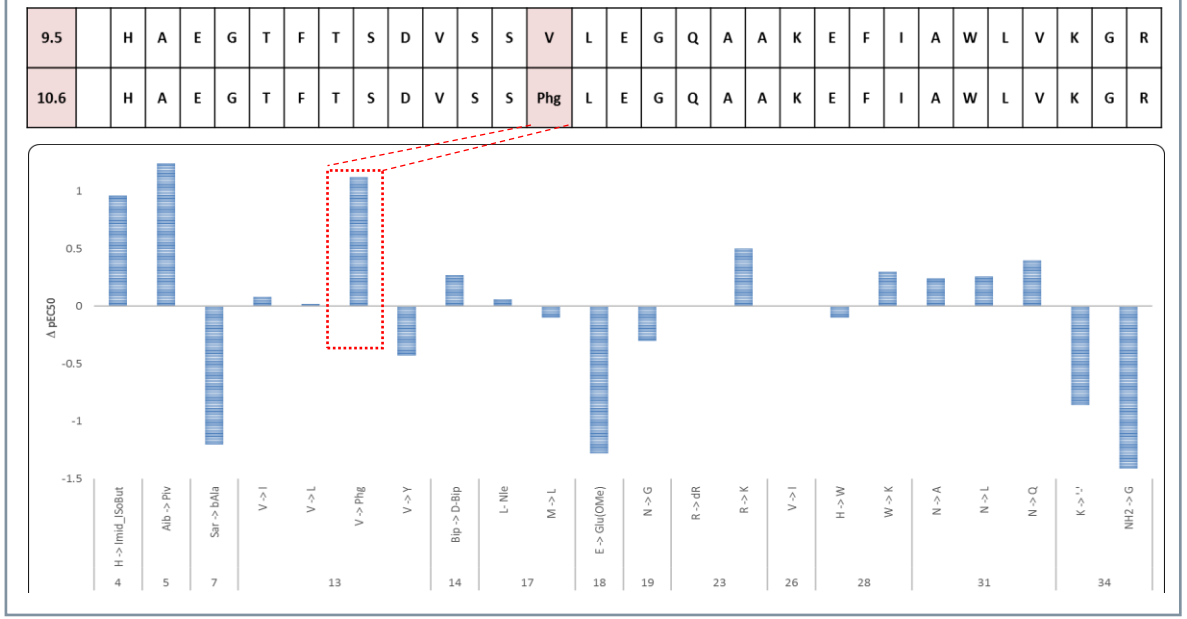
RDKit

sosei HEPTARES

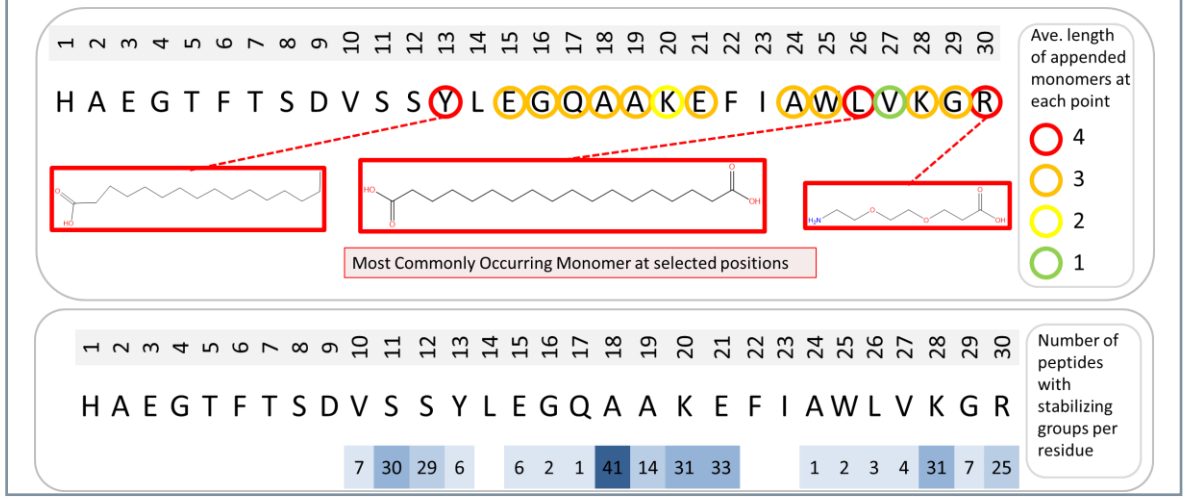


Structure-Based Peptide Ligand design

Property Modulation within Matched Peptide Pairs



Property Modulation within Matched Peptide Pairs



Davenport, Scully, De Graaf, Brown, Maguire (2020) Nature Rev Drug Discov 19: 389

Introduction

- "Making a hash of it: The advantage of selectively leaving out structural information"
  - https://www.nextmovesoftware.com/talks.html - 258th ACS National Meeting, Aug 2019
- Describes 'molhash'
  - Generate molecular hashes useful for finding pairs of tautomers, mesomers, regioisomers, etc.
  - https://nextmovesoftware.github.io/molhash/introduction.html
  - Using hashes can be an efficient alternative to brute-force algorithms
- Incorporated into RDKit (since 2019.09)
  - Example of use by Takayuki Serizawa: https://iwatobipen.wordpress.com/2019/10/27/a-new-function-of-rdkit201909-rdkit-chemoinformatics/
- The problem: finding all matched pairs (in terms of sequence) within a dataset of peptides
  - Singly-substituted peptides
  - Shout-out to related work: "Matched Peptides: Tuning Matched Molecular Pair Analysis for Biopharmaceutical Applications" – Julian Fuchs, Bernd Wellenzohn, Nils Weskamp, Klaus Liedl – *JCIM*. **2015**, *55*, 2315.

Find words that differ by a single letter – brute force

- For each word in a list of English words, convert each letter to each of the other letters of the alphabet in turn
  - Check whether there is a match in the list

- "isotropic": 225 substitutions tested
  - **asotropic**
  - **bsotropic**
  - **csotropic**
  - **dsotropic**
  - **esotropic** ⇦
  - **fsotropic**
  - **...**
  - **iaotropic**
  - **ibotropic**
  - **icotropic**

- Not an efficient approach
  - Consider the extension to two letters changed – 22500
- Does not handle letters or other characters not initially considered
  - diacritics (naïve, piñata, ångström), Greek characters, capital letters, hyphens

Find words that differ by a single letter – using a hash

- For each word in a list of English words, create hashes where each letter in turn is replaced by an asterisk
  - Collate words using the hashes: `collate = defaultdict(list); collate[myhash].append(word)`

- "isotropic":
  - ***sotropic**
  - **i*otropic**
  - **is*tropic**
  - **iso*ropic**
  - **isot*opic**
  - **isotr*pic**
  - **isotro*ic**
  - **isotrop*c**
  - **isotropi***
- "inotropic":
  - ***notropic**
  - **i*otropic**
  - **in*tropic**
  - **ino*ropic**
  - **inot*opic**
  - **inotr*pic**
  - **inotro*ic**
  - **inotrop*c**
  - **inotropi***
- "esotropic":
  - ***sotropic**
  - **e*otropic**
  - **es*tropic**
  - **eso*ropic**
  - **esot*opic**
  - **esotr*pic**
  - **esotro*ic**
  - **esotrop*c**
  - **esotropi***
- Collated by hash:
  - ***sotropic**: [isotropic, esotropic]
  - **i*otropic**: [isotropic, inotropic]
  - **is*tropic**: [isotropic]
  - **iso*ropic**: [isotropic]
  - …30 other hashes with a single entry
- For each word in a list of length ≥ 2, use that word to collate other words in the same list:
  - isotropic: [inotropic, esotropic]
  - inotropic: [isotropic]
  - esotropic: [isotropic]

Find sequences that differ by a single residue – using a hash

- For each sequence in a list of peptides, create hashes where each residue in turn is replaced by an asterisk
  - Collate sequences using the hashes: `collate = defaultdict(list); collate[myhash].append(seq)`

- "isotropic":
  - ***sotropic**
  - **i*otropic**
  - **is*tropic**
  - **iso*ropic**
  - **isot*opic**
  - **isotr*pic**
  - **isotro*ic**
  - **isotrop*c**
  - **isotropi***
- "inotropic":
  - ***notropic**
  - **i*otropic**
  - **in*tropic**
  - **ino*ropic**
  - **inot*opic**
  - **inotr*pic**
  - **inotro*ic**
  - **inotrop*c**
  - **inotropi***
- "esotropic":
  - ***sotropic**
  - **e*otropic**
  - **es*tropic**
  - **eso*ropic**
  - **esot*opic**
  - **esotr*pic**
  - **esotro*ic**
  - **esotrop*c**
  - **esotropi***
- Collated by hash:
  - ***sotropic**: [isotropic, esotropic]
  - **i*otropic**: [isotropic, inotropic]
  - **is*tropic**: [isotropic]
  - **iso*ropic**: [isotropic]
  - …30 other hashes with a single entry
- For each sequence in a list of length ≥ 2, use that sequence to collate other sequences in the same list:
  - isotropic: [inotropic, esotropic]
  - inotropic: [isotropic]
  - esotropic: [isotropic]

sosei HEPTARES

Some tweaks

- Support doubly-substituted matched pairs
  - “isotropic”: `**otropic`, `*s*tropic`, etc.
- Consider one-residue extension/deletion as a matched pair:
  - As well as the nine “isotropic” hashes shown earlier, also include `*isotropic` and `isotropic*` (no other changes needed)
- Use a list, instead of representing the peptide sequence as a string
  - Residues can be represented by arbitrary strings, such as three-letter codes, modified residues, or internal names: e.g. [“Ile”, “N(Me)Ser(O-Ac)”, …]
  - Generate a hash by replacing one of the entries with a “*” (copy the list first), and convert to a string with “-”.join(mylist).

