THE USE OF DEEP DOCKING FOR AUTOMATED, CONSENSUS-BASED HIT IDENTIFICATION IN DRUG DISCOVERY

ARTEM CHERKASOV UBC
MOLECULAR DOCKING – MAJOR DRUG DISCOVERY TOOL

WET-LAB HIGH-THROUGHPUT SCREENING HIT RATE IS \(~0.03\%)\), DOCKING-SUPPORTED HIT RATE IS 5-20\%
CONVENTIONAL MOLECULAR DOCKING WORKFLOW

TARGET PROTEIN/TARGET SITE

DOCKING DATABASE

~5-10 MILLION MOLECULES PER TARGET

EXPERIMENTAL ASSAYS

DRUG CANDIDATES FOR SCREENING

IC50 = 12
R² = 0.9554
2015
no A.I.
5M Molecules Screened
CHEMICAL SPACE REMAINS INACCESSIBLE TO DRUG DISCOVERY

DOCKING CANNOT KEEP UP WITH EXPLODING CHEMICAL SPACE
CURRENTLY ENAMINE RS DATABASE CONTAINS 38 B MOLECULES
DOCKING MISSES OUT 99.9% OF ALREADY AVAILABLE MOLECULES
TOTAL NUMBER OF POSSIBLE DRUG-LIKE MOLECULES : $10^{60} - 10^{100}$
Progressive docking: a hybrid QSAR/docking approach for accelerating in silico high throughput screening.

Cherkasov A, Ban E, Li Y Fallahi M, Hammond GL.

Progressive Docking 1.0
WHAT IF WE PREDICT DOCKING SCORES WITHOUT DOCKING??

Molecule → Fingerprint → Algorithm
DEEP DOCKING

TARGET PROTEIN/TARGET SITE

DOCKING DATABASE

1.4 BILLION MOLECULES PER TARGET

DEEP DOCKING

100s X FASTER

DRUG CANDIDATES

Final enrichment

0 1,000 2,000 3,000 4,000

Protein

AR ERα PPARγ CAMKK2 CDK6 VEGF R2 ADORA2A TBX2R AT1R Nav1.7 GLIC GABA α1 L
DEEP DOCKING PERFORMANCE ON 12 MAJOR DRUG TARGETS

Predicted high scoring molecules augment the training set of the model (1% in total).

Active/inactive cut-off to is made more stringent at every iteration.

NR of molecules predicted as virtual hits after each iteration is reduced.
WE HAVE IDENTIFIED 25+ POTENTIALLY DRUGGABLE SITES ON VIRAL PROTEINS OF SARS-COV-2. SELECTED 3CL PROTEASE AS MAIN TARGET FOR INITIAL DEEP DOCKING WITH 1.4B ZINC15 COMPOUNDS
DEEP DOCKING FOR SARS-CoV-2 DRUG DISCOVERY

1.4B ZINC15 MOLECULES

GPU-AUTODOCK (NVIDIA)

DOCKING DATABASE

SARS-COV-2 3CL PROTEASE

3CL PRO INHIBITORS

DEEP DOCKING

Predict scores with QSAR models
DEEP DOCKING IDENTIFIED 585 POTENTIAL 3CL PRO INHIBITORS

DOCKING SCORES OF TOP 1,000 CANDIDATES SIGNIFICANTLY BETTER THAN OF KNOWN BENCHMARKS

IDENTIFIED 585 UNIQUE SCAFFOLDS FOR SARS-COV-2 3CL PRO, NOT SHARED WITH KNOWN PROTEASE INHIBITORS.

IDENTIFIED 30+ ACTIVES, SOME CONFIRMED BY INDEPENDENT LABS
30+ inhibitors of 3CL pro enzyme are confirmed active

Our first publication with initial drug candidates against COVID19 appeared as early as Feb 19, 2020.

Out of 585 predicted compounds, 30+ active (5%).

Wet-lab screening hit rate is usually ~0.03%.
DEEP DOCKING ENABLES ROUTINE SCREENING >1B LIBRARIES

LARGER DOCKING LIBRARIES YIELD BETTER AND MORE HITS (LYU ET AL, NATURE, 2019)

PUBLICLY AVAILABLE CHEMICAL LIBRARIES KEEP EXPLODING: ZINC20 (1.6B), ENAMINE REAL (1.6B)

FEW METHODS PUBLISHED AFTER OUR FEB20 PUBLICATION ON SCREENING 1B+ ULTRA LARGE LIBRARIES

<table>
<thead>
<tr>
<th>METHOD</th>
<th>DATABASE</th>
<th>REQUIRED TIME</th>
<th>RESOURCES</th>
<th>DOCKING PROGRAM</th>
<th>TARGET</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPENEYE ORION</td>
<td>REAL</td>
<td>1 DAY</td>
<td>45,000 CORES</td>
<td>FRED</td>
<td>PNP/HSP90</td>
<td><a href="HTTPS://WWW.EYESOPE.N.COM/ORION">HTTPS://WWW.EYESOPE.N.COM/ORION</a></td>
</tr>
<tr>
<td>AUTODOCK-GPU</td>
<td>REAL</td>
<td>1 DAY</td>
<td>27,600 GPU</td>
<td>AUTODOCK-GPU</td>
<td>SARS-COV-2 MPRO</td>
<td>ACHARYA ET AL, CHEMRXIV, 2020</td>
</tr>
<tr>
<td>VIRTUALFLOW</td>
<td>REAL</td>
<td>4 WEEKS</td>
<td>8,000 CORES</td>
<td>QUICKVINA, VINA, ...</td>
<td>KEAP1-NRF2 INTERACTION</td>
<td>GORGULLA ET AL, NATURE, 2020</td>
</tr>
<tr>
<td>DEEP DOCKING</td>
<td>ZINC15</td>
<td>5 WEEKS</td>
<td>60 CORES, 4 GPU</td>
<td>FRED, GLIDE</td>
<td>MULTIPLE TARGETS</td>
<td>GENTILE ET AL, CENTRAL SCIENCE, 2020</td>
</tr>
</tbody>
</table>
Stringent consensus docking as hitlist filters
SELECTED 1700 CANDIDATE MRO INHIBITORS (ROWS) FROM 26 HITLISTS (COLUMNS)

GOAL: COMPLETELY AUTOMATED SELECTION. NO HUMAN INSIGHT
High throughput screening (HTS) IQFS assay

- Contracted Bienta (Enamine Biological Services) for HTS
- Used similar substrate
- Automated 384-well plate screening
- GC376 IC50 is similar to Jean’s lab result
INTEGRATED EVALUATION PIPELINE FOR SARS-CoV-2 3CL PRO

MAIN PROTEASE CATALYTIC ASSAY

CELLULAR ASSAYS CONDUCTED IN CL3/BSL3 UBC FINDER

DOUBLE-STRANDED RNA ASSAY: MEASURES SARS-COV-2 VIRAL RNA DURING LIFE CYCLE

NUCLEOCAPSID ASSAY: MEASURES SARS-COV-2 VIRAL PROTEIN PRODUCTION DURING LIFE CYCLE

PLAQUE ASSAY: MEASURES INFECTION/BUDDING OF SARS-COV-2 VIRUS DURING LIFE CYCLE
40B DD hit rate (5 μM; 10% inhibition scored as active)

162 actives/1,387 evaluated = 11.7% hit rate
in 2005 on the emergence of SARS-1 outbreak we created an integrated scientific platform PREPARE-2005 aiming to rapidly respond to future SARS-like infectious threats
WE PROPOSE TO ESTABLISH GLOBAL PATHOGEN - DEFENSE SYSTEM ON THE BASIS OF **PREPARED-2023** PROJECT

**INFRASTRUCTURE INCLUDES ALL LATEST TECHNOLOGICAL ADVANCEMENTS**

**AI CAN SPEED UP CADD CORE WORKFLOW 100-S FOLDS**

**CAN ALSO SIGNIFICANTLY IMPROVE THE OVERALL WORKFLOW PERFORMANCE**
Acknowledgements

• **Cherkasov Lab**
  - Dr. Art Cherkasov
  - Dr. Francesco Gentile
  - Dr. Fuqiang Ban
  - Dr. Michael Llamosa
  - Dr. Michael Hsing
  - Dr. Eric Leblanc
  - Dr. Carl Perez
  - Helene Morin
  - Hazem Mslati
  - James Gleave
  - Jean Charle Yaacoub
  - Jane Foo
  - Mohit Pandey
  - Mariia Radaeva
  - Jiaying You

• **Young Lab**
  - Dr. Robert Young
  - Dr. Michael Bielecki
  - Dr. Jason Smith

• **Strynadka Lab**
  - Dr. Natalie Strynadka
  - Dr. Liam Worrall
  - Dr. Jaeyong Lee

• **Jean Lab**
  - Dr. François Jean
  - Dr. Tirosh Shapira
  - Dr. Andrea Omstead
  - Ivan Villanueva
  - Rory Long

[Logos and affiliations]