Identifying new anti-cancer drugs by multi-tasking QSAR approaches targeting G-quadruplex DNA

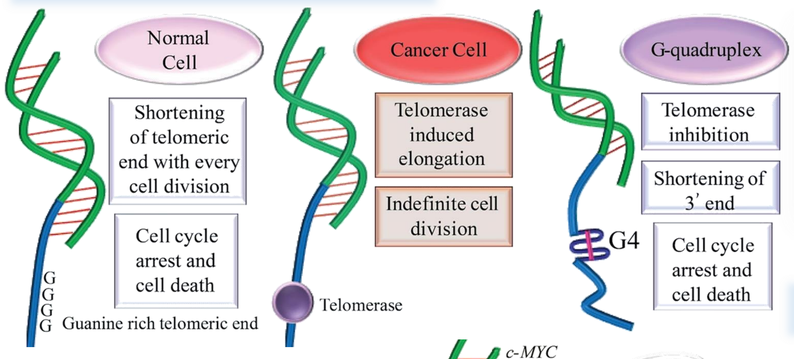
Jyotsna Ambure[1], Pravin Ambure[2], Eva Serrano-Candelas[2], Rafael Gozalbes[1,2]

1. MolDrug AI Systems SL, 2. ProtoQSAR SL,

Centro Europeo de Empresas Innovadoras (CEEI), Parque Tecnológico de Valencia, Spain

INTRODUCTION

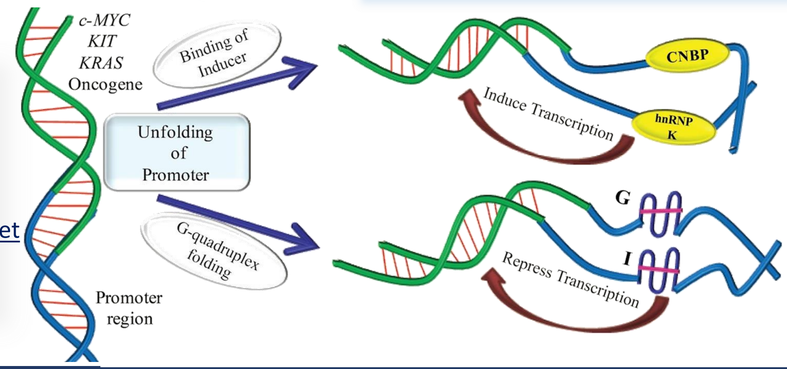
G-quadruplex(G4) at telomere



- G4 is formed in a guanine rich DNA at telomeric region and at promoters of various oncogenes
- Ligand induced stabilization of Telomeric-G4 regulates cell cycle and
- Promoter-G4 stabilization regulates the expression of respective oncogenes

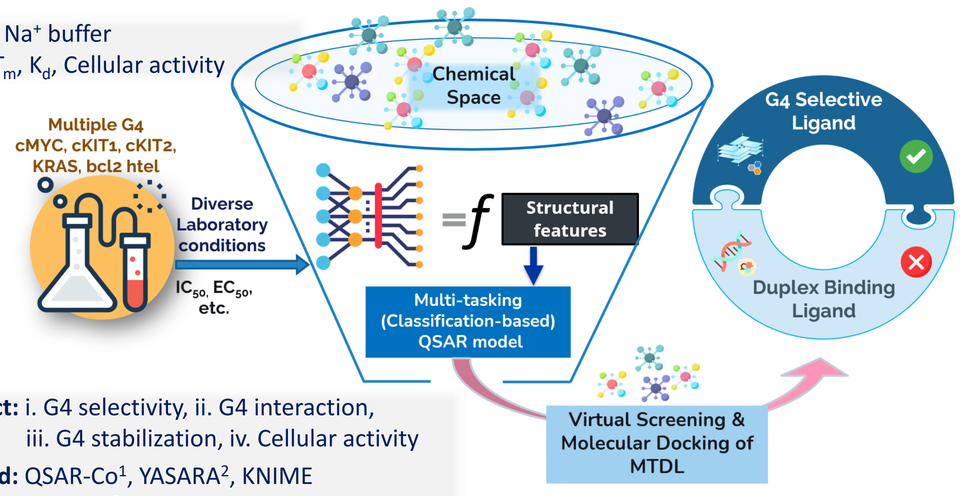
- Cancer is a complex disease with multifactorial gene deregulation
- Simultaneously targeting multiple G4s associated with multiple oncogenes is requisite
- Thus 'stabilization of G4s with multi-target directed ligands (MTDL)' is crucial in targeted cancer therapy

G4 at promoter of oncogenes



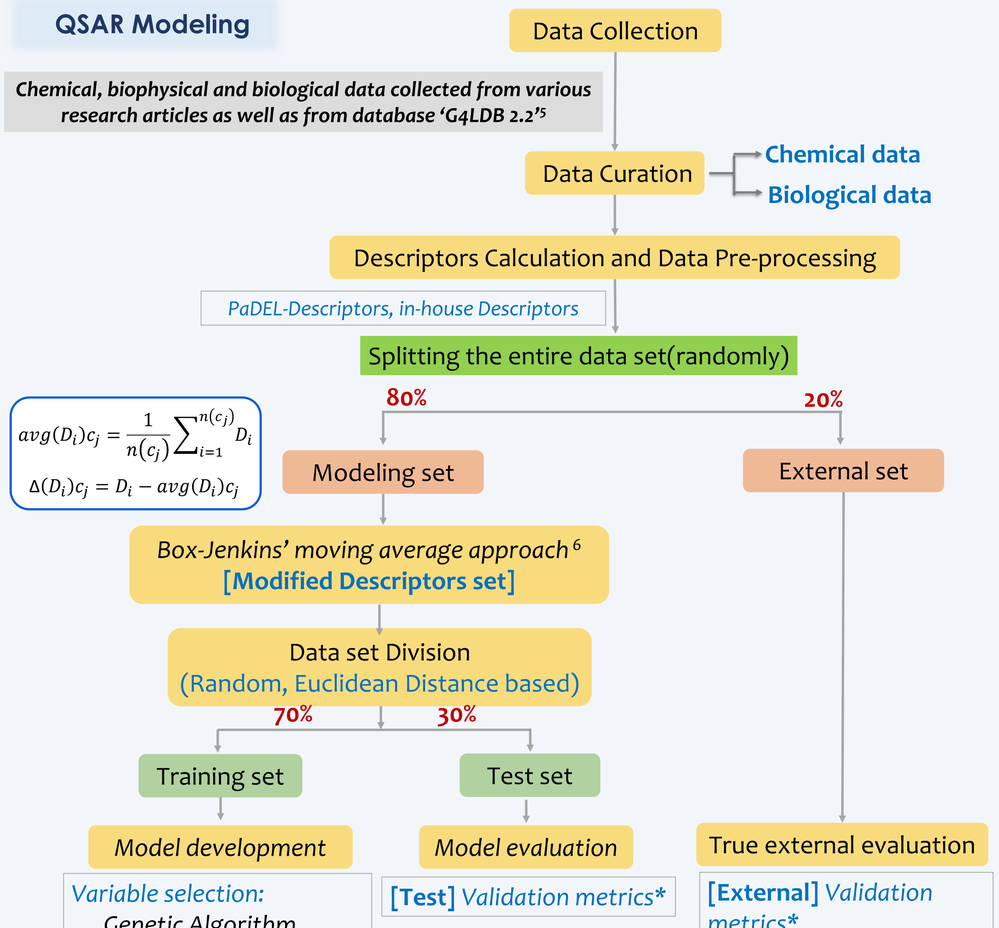
STUDY OUTLINE

- **Buffer:** K^+ and Na^+ buffer
- **End points:** ΔT_m, K_d, Cellular activity



- **Models predict:** i. G4 selectivity, ii. G4 interaction, iii. G4 stabilization, iv. Cellular activity
- **Programs Used:** QSAR-Co[1], YASARA[2], KNIME workflow[3], *In-house* python scripts, PaDEL-Descriptor[4]

QSAR Modeling

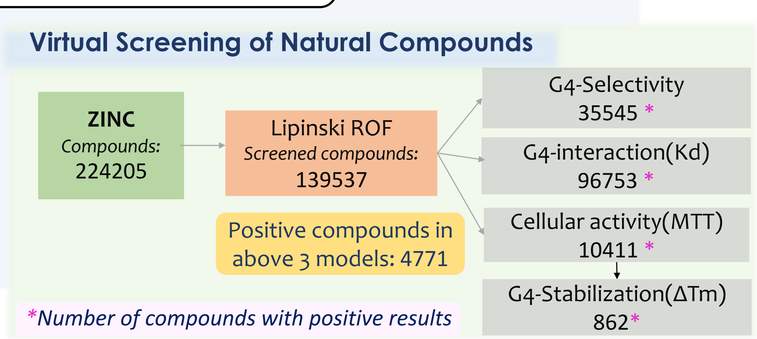


$$avg(D_i)c_j = \frac{1}{n(c_j)}\sum_{i=1}^{n(c_j)} D_i$$

$$\Delta(D_i)c_j = D_i - avg(D_i)c_j$$

*Accuracy, Precision, Specificity, Selectivity, F-score, Mathews Correlation Coefficient

Virtual Screening of Natural Compounds



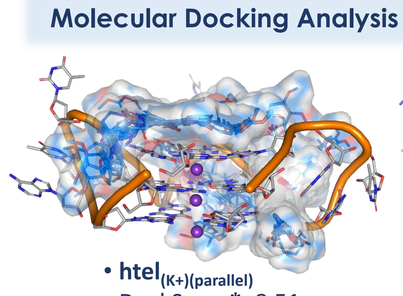
*Number of compounds with positive results

RESULTS

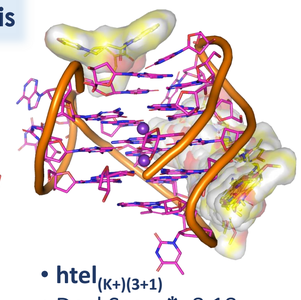
Multi-tasking QSAR Modeling

Model Category	Parameters	Train	Train 10 fold CV	Test	Ext	Ext in-AD
I. G4 Selectivity Descriptor: 15 *Random Forest*	**No. of Compounds**	P: 493 N: 175 Total: 668	P: 493 N: 175 Total: 668	P: 127 N: 39 Total: 166	P: 159 N: 58 Total: 217	P: 156 N: 50 Total: 209
	Accuracy %	100	87.13	86.75	90.78	91.26
	Precision%	100	88.32	91.34	92.64	92.59
	Sensitivity%	100	95.13	91.34	94.97	96.15
	Specificity%	100	64.57	71.79	79.31	76
	F-measure	1	0.92	0.91	0.94	0.94
	MCC	1	0.65	0.63	0.76	0.75
II. G4 Interaction (K_d) Threshold Positive: K_d <1μM Negative: K_d ≥1μM Descriptor: 10 *LDA*	**No. of Compounds**	P: 100 N: 95 Total: 195	P: 100 N: 95 Total: 195	P: 21 N: 27 Total: 48	P: 29 N: 25 Total: 54	P: 23 N: 16 Total: 39
	Accuracy %	92.82	89.23	85.42	83.33	89.74
	Precision%	93	88.35	75	88.46	95.24
	Sensitivity%	93	91	100	79.31	86.96
	Specificity%	92.63	87.37	74.07	88	93.75
	F-measure	0.93	0.90	0.86	0.82	0.91
	MCC	0.86	0.78	0.74	0.67	0.80
III. G4 Stabilization (ΔT_m) Threshold Positive: $\Delta T_m \geq 15^oC$ Negative: $\Delta T_m < 15^oC$ Descriptor: 10 *Random Forest*	**No. of Compounds**	P: 498 N: 394 Total: 892	P: 498 N: 394 Total: 892	P: 129 N: 93 Total: 222	P: 154 N: 121 Total: 275	P: 119 N: 85 Total: 204
	Accuracy %	99.89	86.66	90.54	85.45	93.14
	Precision%	99.80	86.65	91.54	89.58	96.46
	Sensitivity%	100	89.96	92.25	83.77	91.60
	Specificity%	99.75	82.49	88.17	87.60	95.29
	F-measure	0.999	0.88	0.92	0.87	0.94
	MCC	0.998	0.73	0.80	0.71	0.86
IV. Cellular Activity (MTT, MTS, CCK8) Threshold Positive: IC50 <10μM Negative:IC50 ≥10μM Descriptor: 10 *Random Forest*	**No. of Compounds**	P: 244 N: 219 Total: 463	P: 244 N: 219 Total: 463	P: 71 N: 44 Total: 115	P: 79 N: 65 Total: 144	P: 74 N: 65 Total: 139
	Accuracy %	100	87.26	84.35	83.33	82.73
	Precision%	100	87.70	90.77	84.81	83.78
	Sensitivity%	100	88.07	83.10	84.81	83.78
	Specificity%	100	86.36	86.36	81.54	81.54
	F-measure	1	0.88	0.87	0.85	0.84
	MCC	1	0.74	0.68	0.66	0.65

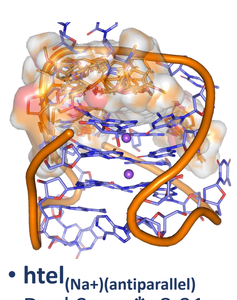
Molecular Docking Analysis



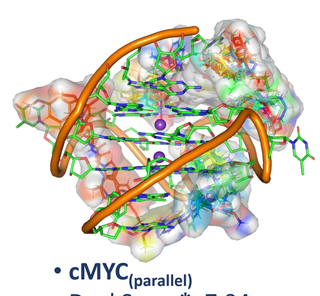
- **htel$_{(K+)(parallel)}$**
- DockScore*: 8.51
- 3'end, Groove



- **htel$_{(K+)(3+1)}$**
- DockScore*: 8.18
- Groove

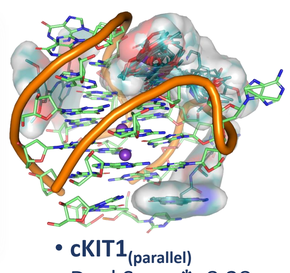


- **htel$_{(Na+)(antiparallel)}$**
- DockScore*: 8.91
- 5'end, Groove

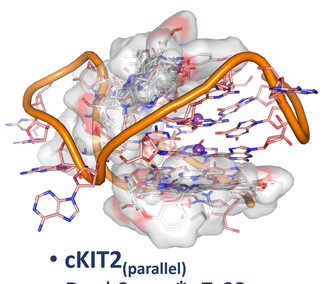


- **cMYC$_{(parallel)}$**
- DockScore*: 7.94
- 5'end, 3'end, Groove

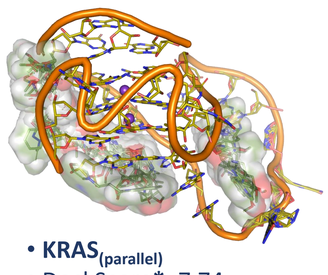
- Screened **MTDLS** bind with multiple G4s
- Docking scores are in energetically favorable range
- MTDLS majorly show end stacking and groove binding



- **cKIT1$_{(parallel)}$**
- DockScore*: 8.28
- 3'end, Groove



- **cKIT2$_{(parallel)}$**
- DockScore*: 7.68
- 5'end, 3'end, Groove



- **KRAS$_{(parallel)}$**
- DockScore*: 7.74
- 5'end, 3'end, Groove

DockScore: Average docking score*

CONCLUSIONS

- This work represents **the very first computational study** for identifying **MTDLs** against **multiple G4s**
- **Multi-tasking** QSAR models predict **G4 selectivity, G4 interaction, G4 stabilization, and Cellular activity** thus offering an integrated evaluation of ligands for G4 activity.
- The models have high discriminatory power and are robust (the accuracy of all the models is >90% for the training sets and >80% for the external sets)
- Molecular docking of screened molecules suggests the feasible stable binding of the ligands with multiple G4s

Study in Pipeline

- **Multi-tasking** QSAR models will be **validated** by ***in-vitro*** experiments such as UV-Vis, CD spectroscopies, FRET melting assay, and MTT assay
- Screened molecules with positive response in *in-vitro assays* will be studied further with molecular dynamics simulation to analyze the binding interactions and binding free energies *etc.*
- Refined QSAR models will be deployed in the form of a standalone software

ACKNOWLEDGEMENT

European Commission

H2020 MSCA-IF 2020

MOLDRUG AI Systems S.L. Sistemas de Inteligencia Artificial Aplicadas al Diseño de Moléculas

ICMOL INSTITUTO DE CIENCIA MOLECULAR UNIVERSITAT DE VALENCIA

PosterPresentations.com flaticon

REFERENCES

1. QSAR-Co: J. Chem. Inf. Model., 2019 59 (6), 2538-2544, DOI: 10.1021/acs.jcim.9b00295
2. YASARA: Bioinform., 2014 30 (20), 2981–2982, DOI:10.1093/bioinformatics/btu426
3. KNIME workflow: https://sites.google.com/site/dtclabdc/
4. PaDEL-descriptor: J. Comput. Chem., 2011 32, 1466-1474, DOI:10.1002/jcc.21707
5. G4LDB: Nucleic Acids Res., 2022 50(D1), D150-D160, DOI:10.1093/nar/gkab952
6. Mol. Divers., 2015;19(2): 347-56, DOI:10.1007/s11030-015-9571-9