xGen™ for drug-discovery

Fitting of complex ligand conformational ensembles to X-ray electron density maps

Himani Tandon[1], Ajay N. Jain[1], Ann E. Cleves[1], Matthew Segall[1]

[1]*Optibrium Ltd, Cambridge, United Kingdom,* ***email: info@optibrium.com***

Introduction

Conventional ligand fitting and refinement in X-ray electron density maps relies on single conformers and B-factors, that often yields ligands with unrealistically high conformational strain. **xGen™** is a **real-space ligand fitting** and **refinement method** that **balances** electron density fit with ligand conformational strain. It is applicable to small molecules and macrocyclic peptides alike. It produces **occupancy-weighted ensembles** yielding substantially **reduced strain energies** compared to deposited structures.

Applying the xGen method to over 3,000 protein-ligand complexes revealed that strain estimates calculated using PDB ligand coordinates were unusually high. It further showed that strain increases superlinearly with ligand size and established a strong inverse correlation between ligand efficiency and per-atom strain, demonstrating strain as a predictive factor in drug design.

The xGen™ method

Input
(Ligand structure + electron density)

Core strategy: Conformational search using a reward term for electron density overlap. Each conformer expanded into a trio with re-minimisation:

- Balanced conformer (**slate**) – compromise between energy and fit
- Energy-favoured surrogate conformer (**yellow**) – minimises strain with positional restraints
- Density-favoured conformer (**orange**) – maximises density fit

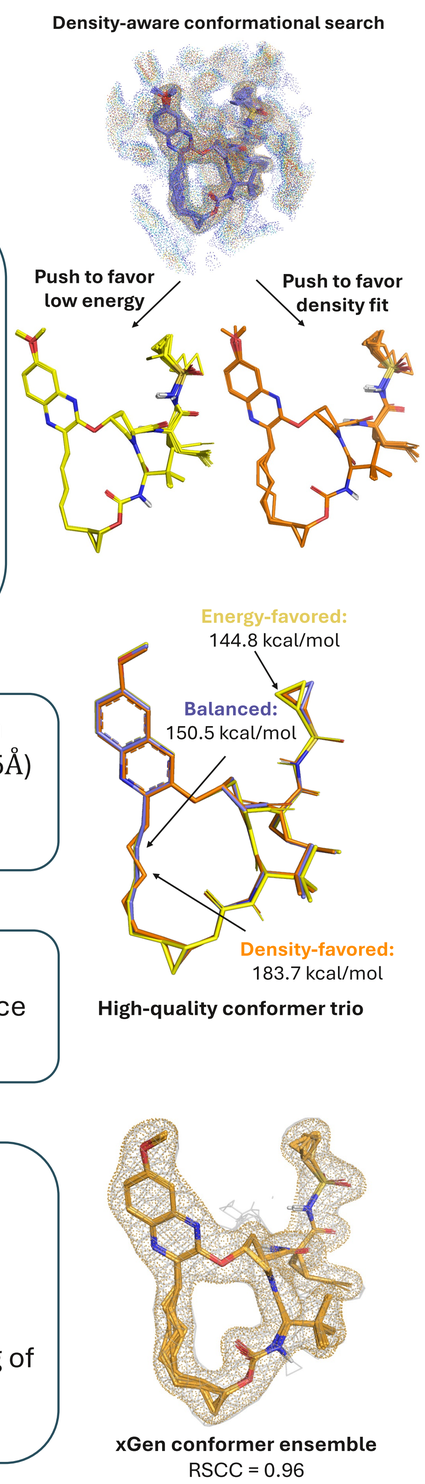
Quality filtering: A high-quality trio identified based on

- Geometric compactness of trio (scaled RMSD ≤0.65Å)
- Near-optimal density fit (≥90%)
- Low-energy (≤ 3.0 kcal/mol window)

Ensemble building: Conformers from qualifying trios optimised for occupancy weights to minimise real-space R-factor (RSR), without atom-specific B-factors

The result:

- Explicit conformational heterogeneity with lower strain (*difference between surrogate conformer energy and global minimum energy*)
- Improved density fit
- Can be used for both refinement and *de novo* fitting of small molecules to large macrocycles, including peptides



High-quality conformer trio

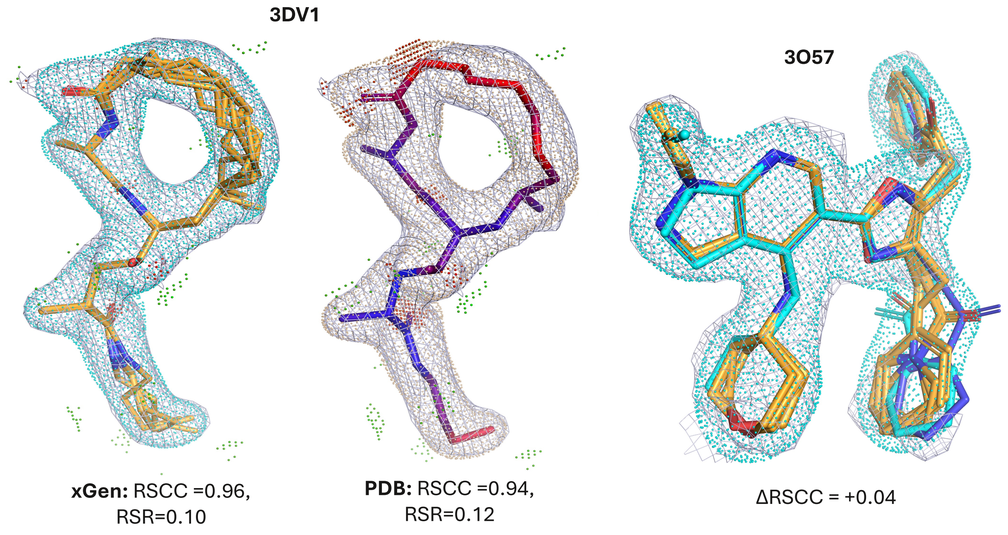
xGen conformer ensemble
RSCC = 0.96

Refinement and *de novo* fitting using xGen

xGen ensembles achieve better density fits and reduce the ligand strain of deposited PDB models by ~50% for both refinement and *de novo* fitting. Average strain for:

150 macrocycles: 3.7 kcal/mol *vs* 6.8 kcal/mol

76 non-macrocycles: 2.5. kcal/mol *vs* 4.2 kcal/mol



xGen: RSCC =0.96, RSR=0.10

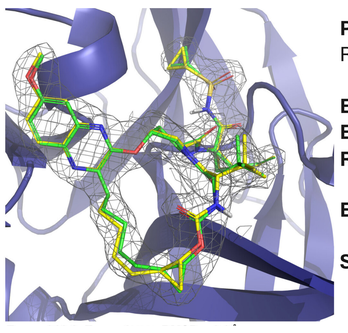
PDB: RSCC =0.94, RSR=0.12

ΔRSCC = +0.04

Real-space refinement of macrocycles (left) with 3DV1 shown. xGen ensemble (orange) vs. PDB reference coordinates (coloured by B-factors) showing improved RSCC/RSR.
De novo fitting (right) with 3O57 shown. xGen ensemble (orange) captures both primary (cyan) and alternate (dark blue) PDB conformers with improvement in RSCC.

Ligand strain, size, and efficiency relationships

Applying xGen to ~3000 protein-ligand complexes revealed that strain energies calculated using deposited PDB ligand structures are artifactually high.

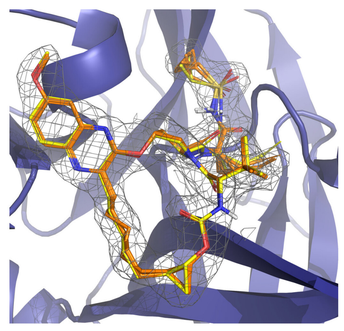


PDB 3SUE, Deposited
RSCC=0.95, RSR =0.12

E_{PDB} = 316.8,
E_{surr} = 154.7,
RMSD = 0.3Å

E_{gmin} = 138.6,

Strain = 16.1 kcal/mol



PDB 3SUE, xGen Ensemble
RSCC=0.96, RSR =0.11

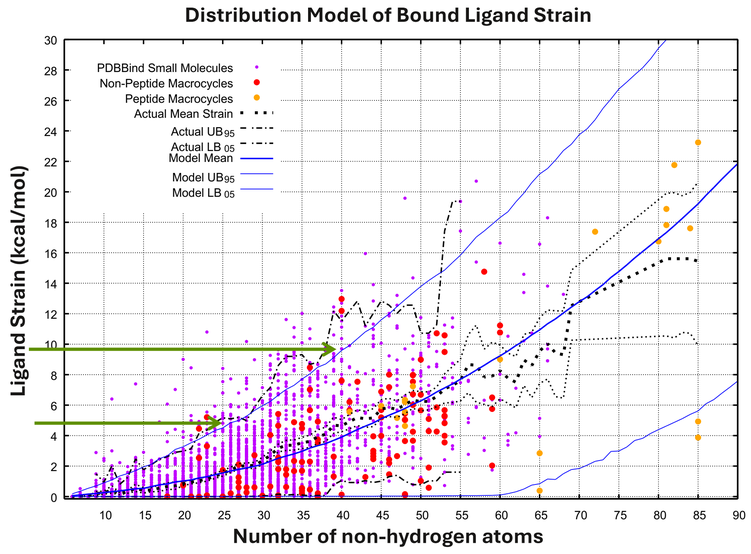
E_{xGen} = 154.3-183.7,
E_{surr} = 142.5-144.6,
RMSD = 0.3-0.4Å

E_{gmin} = 138.6,

Strain = 3.9 kcal/mol

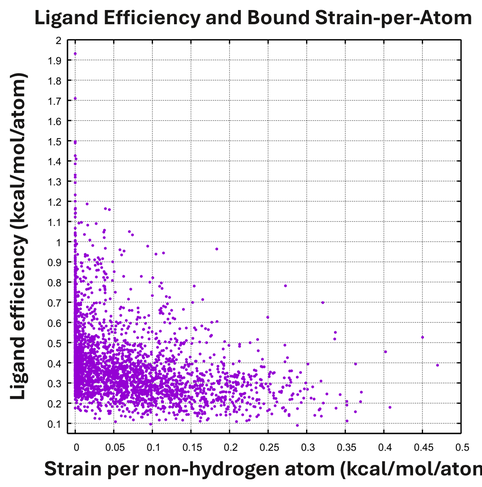
Grazoprevir-NS3/4A protease (3SUE). PDB ligand (green) fits the electron density well **(left)** (RSCC = 0.95) but show high strain (16.1 kcal/mol) calculated as difference between surrogate conformer (yellow) energy (E_{surr}) and global minimum conformer energy (E_{gmin}). xGen ensemble (orange) maintains fit quality (improved RSCC/RSR) while reducing strain by 75% to 3.9 kcal/mol **(right)**.

Ligand strain increases superlinearly with molecular size, following a predictable distribution.



The distributional model provides practical upper bounds for conformational search protocols and design strategies. For example, 4.5 kcal/mol for 25 atoms and 9.4 kcal/mol for 40 atoms (green arrows).

There is also a strong inverse relationship between ligand efficiency i.e., how tightly a ligand binds for its size, and ligand strain-per-atom ($\tau = -0.35$, $p \ll 0.001$).



High-efficiency ligands seldom have high per-atom bound conformational strain, whereas low efficiency ligands have variable strain estimates. This further highlights the importance of minimising the strain during ligand optimisation.

Conclusions

- **xGen offers a paradigm shift for ligand modelling,** producing physically realistic conformer ensembles for ligands
- **Ensemble-based fitting yields ligands with lower strain estimates, suggesting greater biological relevance**
- **Ligand strain is superlinear and is a predictive factor for drug design and optimisation:** If a ligand has high strain relative to expected distribution, aim to optimise its geometry and if it already has low strain, improve protein-ligand interaction footprint

References

Jain AN, Cleves AE, *et al* (2020), *J. Med. Chem.63 (18);* Jain AN, Brueckner AC, *et al* (2023) *J Med Chem* 66(3)

Acknowledgements

Merck: Alexander C. Brueckner, Mikhail Reibarkh, and Edward C. Sherer

Optibrium: Mario Öeren and Kyle Kroeck

Learn more on our webpage