

## 1. Pendahuluan

*Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) atau sindrom pernafasan akut coronavirus 2 adalah virus yang menyebabkan penyakit *Coronavirus disease 19* (COVID-19) [1]. Pada bulan Desember 2019 di Wuhan China, terdapat pasien yang menderita pneumonia yang tidak diketahui penyebabnya, namun dipastikan bahwa pasien ini terinfeksi oleh virus corona baru, lalu virus itu dinamai 2019-nCoV [2]. Namun dikarenakan virus ini memiliki sindrom yang mirip dengan *Severe Acute Respiratory Syndrome Coronavirus* (SARS-CoV), maka 2019-nCoV berganti nama menjadi *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) [2]. SARS-CoV-2 merupakan virus zoonosis atau virus yang berpindah dari hewan ke manusia keempat yang muncul dalam kurun waktu 20 tahun terakhir [3]. Coronavirus secara alami dimiliki dan dibentuk secara evolusioner oleh kelelawar. Beberapa kelompok peneliti telah mengkonfirmasi bahwa ada kesamaan genetik antara SARS-CoV-2 dengan *betacoronavirus* kelelawar dari sub-genus *Sarbecovirus* [4].

COVID-19 telah menimbulkan 205.338.159 kasus yang telah dikonfirmasi, dengan angka kematian sebesar 4.333.094 kematian pertanggal 13 Agustus 2021 [5]. Penderita COVID-19 umumnya memiliki gejala seperti demam selama lebih dari 1 minggu, batuk kering, hingga sesak nafas. Tak sedikit pula dari beberapa kasus penderitanya memiliki gejala yang tidak terlihat [6].

Pada tanggal 1 Desember 2020 terdapat empat uji coba vaksin yang mendapatkan hasil sementara yang sangat baik, dan di beberapa negara vaksin yang belum dilisensikan ini diizinkan sebagai penggunaan darurat dengan alasan kesehatan masyarakat [7]. Sehingga sampai saat ini belum ada obat spesifik yang dapat menyembuhkan penderita yang terinfeksi SARS-CoV-2, dan penanganan yang dapat diberikan kepada penderita COVID-19 saat ini adalah dengan mengandalkan pemberian oksigen, cairan konservatif, dan beberapa rekomendasi perawatan berbasis bukti penanganan pada penyakit serupa yang pernah dilakukan seperti *plasma therapy*, *monoclonal antibodies*, *small molecule drug-based therapies*, dan *immune regulation therapy and interferon utilizing* [3], [8].

Terdapat beberapa protein yang menjadi kunci penting dalam penyebaran dan replika virus SARS-CoV-2 salah satunya adalah protein *Papain-like Protease* (PLpro) SARS-CoV-2 [9]–[11], PLpro juga dijadikan target penting oleh beberapa peneliti dalam langkah *drug discovery* untuk virus SARS-CoV-2, hal ini dikarenakan PLpro memiliki aktivitas deubiquitinating dan mengarahkan implikasi penting pada replikasi dan patogenesis virus [12]. Dalam langkah *drug discovery* yang menjadi masalah utama dalam menemukan obat yang sesuai adalah terkait jumlah eksperimen yang cukup banyak dan mahal biaya yang dibutuhkan untuk melakukan eksperimen-eksperimen tersebut [13].

Untuk mengatasi masalah ini, model *quantitative structure activity relationship* (QSAR) dapat digunakan untuk mengakselerasi proses menemukan obat. Cara kerja QSAR ini adalah dengan mencari keterkaitan atau korelasi antara struktur senyawa dengan bioaktivitas atau sifat fisikokimianya. Model ini memiliki kelebihan dapat mengurangi kerugian besar yang disebabkan oleh pemilihan kandidat obat dengan cara mengidentifikasi senyawa timbal [13]. Selain itu dengan menggunakan model QSAR juga dapat mengurangi eksperimen yang dilakukan dalam memilih kandidat obat [13]. Pada penelitian ini penulis memprediksi aktivitas biologikal menggunakan metode *ensemble*, dengan algoritma *Random Forest* dan *AdaBoost*.

Penelitian terkait *drug discovery* dengan menggunakan *Quantitative Structure Activity Relationships* (QSAR) ini telah banyak dilakukan karena dapat membantu peneliti mengidentifikasi ada atau tidaknya bahan berbahaya, dan dalam kasus tertentu dapat digunakan sebagai pengganti hasil uji hewan [14]. Telah banyak penelitian *drug discovery* yang menggunakan QSAR sebagai model untuk mengakselerasi penemuan obat. Salah satunya penelitian yang dilakukan oleh Ghaleb dkk [15]. Penelitian tersebut dilakukan observasi untuk melihat *pyridine N-Oxide* sebagai penghambat potensial untuk SARS-CoV-2 dengan menggunakan 3D-QSAR, *molecular docking modeling*, dan ADMET *screening* [15]. Database yang digunakan terdiri dari 110 senyawa *pyridine N-Oxide* sebagai antigen virus, yang mana dataset ini dibagi menjadi dua yaitu, data latih sebanyak 90 senyawa dan data tes sebanyak 20 yang dipilih menggunakan seleksi acak [15]. Lalu database ini akan digunakan untuk membangun model 3D-QSAR dan untuk menganalisis sifat fisikokimianya [15]. Hasil dari penelitian ini menunjukkan bahwa senyawa *pyridine N-Oxide* lebih baik untuk menghambat SARS-CoV-2 dibanding *chloroquine* dan *hydroxychloroquine* [15].

Selain itu, penelitian terkait lainnya melakukan penelitian untuk mencari penghambat enzim protease mirip 3C (3CLpro) terhadap penyakit SARS-CoV dengan menggunakan QSAR sederhana. Pada penelitian ini peneliti mengembangkan model 2D-QSAR berbasis *multiple linear regression* (MLR) untuk

mengidentifikasi struktur penting yang paling berpengaruh terhadap penghambatan enzim protease pada penyakit SARS-CoV-2. Model ini telah dibangun dan dikembangkan oleh peneliti dengan 8 deskriptor yang signifikan [16]. Tujuan pengembangan model yang sederhana ini adalah agar model dapat digunakan dengan mudah untuk memprediksi dan menyaring senyawa eksternal [16]. Kesimpulan dan hasil dari penelitian ini adalah model yang divalidasi dapat memprediksi aktivitas penghambatan baru terhadap enzim 3CLpro, dan fitur dari analisis 2D-QSAR dan *molecular docking modeling* dapat digunakan untuk desain penghambat baru, dan daftar penghambat prioritas yang didapatkan dari *virtual screening* dapat digunakan untuk pengembangan analisis dan pencarian penghambat yang baru [16].

Penelitian lainnya juga dilakukan oleh Abdul Amin dkk. yang diterbitkan pada tahun 2020, disini peneliti melakukan *drug discovery* dengan menggunakan QSAR berbasis *monte carlo*, *virtual screening*, dan *molecular docking study* untuk meneliti molekul *in-house* sebagai penghambat *papain-like protease* (PLpro) [17]. Pada penelitian ini terdapat dua pendekatan yang telah dilakukan oleh penulis yang pertama pendekatan berbasis ligan yang mana didalamnya dilakukan pengklasifikasian dengan menggunakan QSAR berbasis *data mining* yang diambil dari berbagai jenis penghambat senyawa PLpro, dan dilakukan juga *virtual screening* dengan basis QSAR untuk mengidentifikasi molekul *in-house* pada pendekatan dengan menggunakan basis ligan, lalu yang selanjutnya pada penelitian ini juga melakukan pendekatan menggunakan basis struktur [17]. Data yang digunakan pada penelitian ini sebanyak 91 molekul, dan penelitian ini menggunakan 3 deskriptor yang berbeda yaitu deskriptor berbasis SMILES, deskriptor berbasis grafik, dan deskriptor berbasis hibrida. Penelitian tersebut menemukan bahwa molekul *in-house* mempunyai potensi untuk dapat digunakan sebagai *seed* pada *drug design* dan optimasi untuk SARS-CoV-2 PLpro, dan setelah dilakukan studi *in vitro* dan *in vivo* yang ekstensif didapan hasil juga bahwa hasil *virtual screening* yang dilakukan pada molekul *in-house* dapat dijadikan pilihan pengobatan untuk penyakit COVID-19 [17].

Penelitian yang dilakukan oleh penulis disini adalah memprediksi aktivitas biologikal menggunakan metode *ensemble learning*, dengan deskriptor yang digunakan pada penelitian ini adalah *fingerprint*, dan jenis *fingerprint* yang digunakan pada penelitian ini adalah *PubChem* dan *Extended*. *Ensemble learning* dipilih sebagai metode pada penelitian ini dikarenakan memiliki kemampuan generalisasi yang kuat, sehingga dapat mendorong model dasar untuk memiliki hasil yang lebih baik dan menghasilkan nilai prediksi yang kuat [18]. Beberapa hal yang ingin dicapai pada penelitian ini adalah membangun model prediksi QSAR yang dapat mengklasifikasikan data *Papain-like Protease* (PLpro) SARS-CoV-2, menilai kinerja algoritma yang dipilih dalam penelitian ini yaitu *AdaBoost* dan *Random Forest*, dan mengetahui pengaruh jenis *fingerprint* terhadap model yang dibangun.