Universidade de São Paulo
Instituto de Química



LIBNM
Laboratório Integrativo de Bioprospecção de Novas Moléculas

SnorMol QSAR

HQSAR · CoMFA · CoMSIA · QSAR clássico (Hansch), no navegador

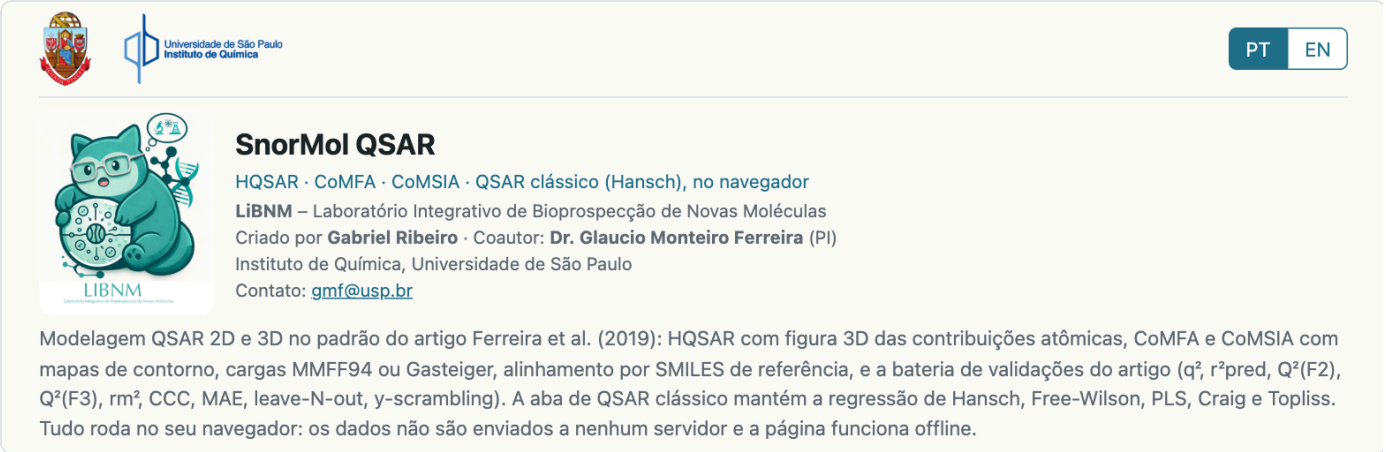
Guia completo, passo a passo

LiBNM – Laboratório Integrativo de Bioprospecção de Novas Moléculas · Instituto de Química, Universidade de São Paulo

Criado por Gabriel Ribeiro · Coautor: Dr. Glaucio Monteiro Ferreira (PI) · versão 2.1, novembro de 2026

1. O que é e o que faz

O SnorMol QSAR é uma página única (um arquivo HTML) que roda inteira no navegador, sem instalar nada e sem enviar dados a servidor algum. Ele segue o padrão do artigo *Ferreira et al., J Biomol Struct Dyn 2019* (QSAR da inibição da sirtuína 2 por anilinobenzamidas): três modelos independentes, **HQSAR** (2D, por fragmentos), **CoMFA** e **CoMSIA** (3D, por campos), todos com a mesma bateria de validação. A aba **QSAR clássico (Hansch)** mantém a regressão linear múltipla e as ferramentas da versão anterior (MiniQSAR).



Topo da página: brasão da USP e logotipo do IQ-USP, mascote SnorMol, seletor de idioma (PT/EN) e as abas dos módulos.

Aba	Para que serve
QSAR clássico (Hansch)	Dados do celecoxibe com IC50 editável, –log IC50, descritores e gráfico de dispersão; regressão com descritores (logP, π, σ, MR...), q², validação externa, Free-Wilson, PCA/PLS, Craig, Topliss, calculadoras.
Ligantes 3D	Carrega o SDF, define Y, calcula cargas (Gasteiger ou MMFF94), alinha ao composto de referência por SMILES e separa treino e teste.
HQSAR	Busca de modelos por fragmentos, validação, figura 3D com contribuições positivas e negativas.
CoMFA / CoMSIA	Campos moleculares e índices de similaridade; mapas de contorno 3D; validação completa.
Validação	Tabela comparativa dos modelos (formato da Tabela 4 do artigo), gráficos observado × previsto do teste e glossário de critérios.
Sugestões (Craig e Topliss)	Propõe trocas de substituintes a partir do composto-líder (ajuste de Hansch em Σπ e Σσ), cruza com os contornos do CoMFA/CoMSIA e as contribuições do HQSAR e indica o próximo passo do esquema de Topliss.
Guia e referências	Resumo dos passos, link para este PDF e a lista de referências.

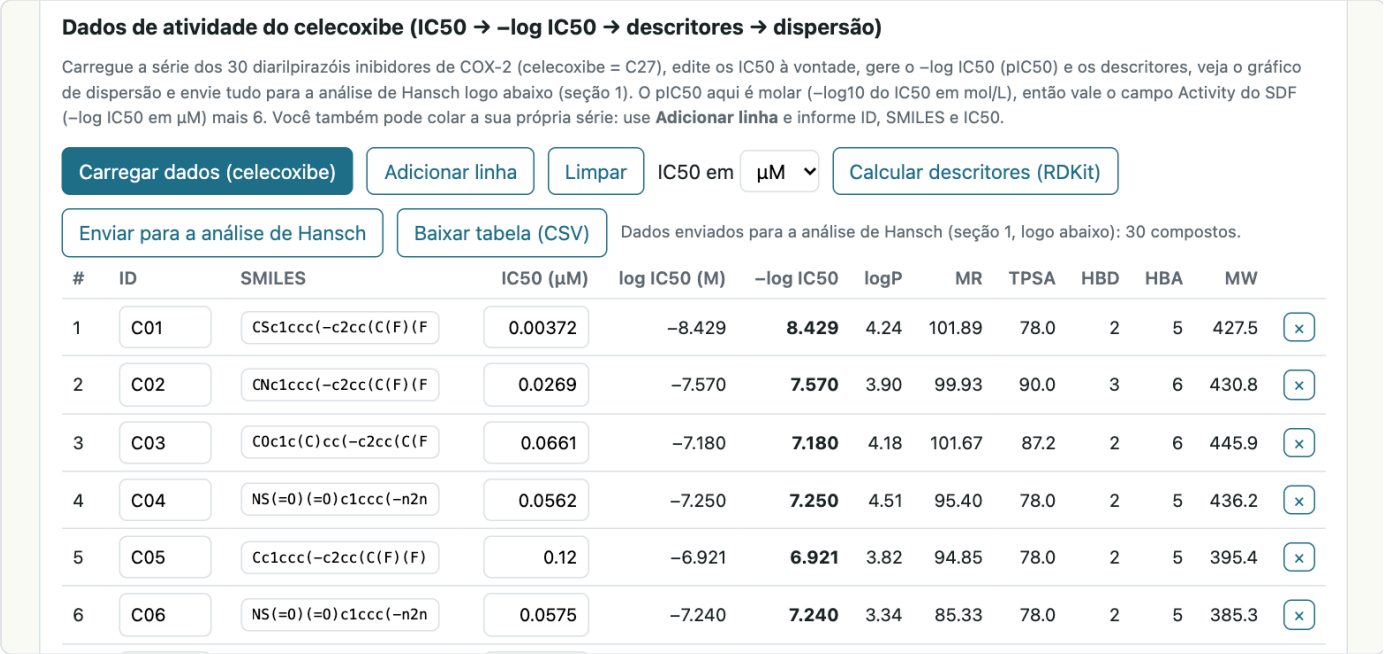
2. Antes de começar

- **Arquivo:** abra `SnorMol_QSAR.html` com dois cliques (Chrome, Edge, Firefox ou Safari atuais). A página abre no idioma do navegador; para forçar, acrescente `?lang=pt` ou `?lang=en` ao endereço.
- **Material padrão:** a página já abre com o conjunto de 30 1,5-diarilpirazóis inibidores de COX-2 (Penning et al., 1997; Molecules 2000, 5:945), alinhados ao celecoxibe (composto C27). Y = –log IC50 (µM). O campo `Set` do SDF marca 25 compostos de treino e 5 de teste. Os mesmos dados estão em `diarilpirazois_alinhados_celecoxibe.sdf` e `qitb_diarilpirazois_COX2.csv` (botões de download na aba Ligantes 3D).
- **Seu próprio SDF:** formato V2000 (não V3000), coordenadas 3D, **hidrogênios explícitos** e um campo numérico com a atividade (por exemplo `Activity`). Opcionalmente, um campo `Set` com `train` ou `test`.
- **Primeiro uso do RDKit:** as cargas MMFF94 e o alinhamento por SMILES usam o RDKit embutido na página (≈ 7 MB descompactados). Ele inicia sob demanda, na primeira vez que você aplica MMFF94 ou alinha; leva alguns segundos.

3. Passo a passo

Passo 0. QSAR clássico: dados do celecoxibe, IC50, –log IC50 e dispersão (aba “QSAR clássico (Hansch)”)

No alto da aba há o cartão **Dados de atividade do celecoxibe**. O botão **Carregar dados (celecoxibe)** traz a série dos 30 diarilpirazóis (celecoxibe = C27) com ID, SMILES e **IC50**, calcula o **log IC50** e o **–log IC50 (pIC50)** e os descritores (logP, MR, TPSA, HBD, HBA, massa molar; RDKit, Crippen/Lipinski) e já envia tudo para a seção 1 da análise de Hansch. A página faz isso sozinha ao abrir; o botão repete o carregamento e desfaz edições.



Dados de atividade do celecoxibe (IC50 → –log IC50 → descritores → dispersão)

Carregue a série dos 30 diarilpirazóis inibidores de COX-2 (celecoxibe = C27), edite os IC50 à vontade, gere o –log IC50 (pIC50) e os descritores, veja o gráfico de dispersão e envie tudo para a análise de Hansch logo abaixo (seção 1). O pIC50 aqui é molar (–log10 do IC50 em mol/L), então vale o campo Activity do SDF (–log IC50 em µM) mais 6. Você também pode colar a sua própria série: use **Adicionar linha** e informe ID, SMILES e IC50.

Carregar dados (celecoxibe) | Adicionar linha | Limpar | IC50 em µM | Calcular descritores (RDKit)

Enviar para a análise de Hansch | Baixar tabela (CSV) | Dados enviados para a análise de Hansch (seção 1, logo abaixo): 30 compostos.

#	ID	SMILES	IC50 (µM)	log IC50 (M)	–log IC50	logP	MR	TPSA	HBD	HBA	MW	
1	C01	CSc1ccc(-c2cc(C(F)(F	0.00372	−8.429	**8.429**	4.24	101.89	78.0	2	5	427.5	×
2	C02	CNc1ccc(-c2cc(C(F)(F	0.0269	−7.570	**7.570**	3.90	99.93	90.0	3	6	430.8	×
3	C03	COc1c(C)cc(-c2cc(C(F	0.0661	−7.180	**7.180**	4.18	101.67	87.2	2	6	445.9	×
4	C04	NS(=O)(=O)c1ccc(-n2n	0.0562	−7.250	**7.250**	4.51	95.40	78.0	2	5	436.2	×
5	C05	Cc1ccc(-c2cc(C(F)(F)	0.12	−6.921	**6.921**	3.82	94.85	78.0	2	5	395.4	×
6	C06	NS(=O)(=O)c1ccc(-n2n	0.0575	−7.240	**7.240**	3.34	85.33	78.0	2	5	385.3	×

Tabela editável: o IC50 pode ser alterado na própria célula (a unidade vem do seletor “IC50 em”); log IC50 e pIC50 são recalculados na hora.

Editar: clique na célula do IC50 (ou do ID e do SMILES) e digite; o pIC50 e o gráfico se atualizam. **Adicionar linha** inclui um composto novo (ID, SMILES, IC50) e **×** remove uma linha; depois de mexer nos SMILES, use **Calcular descritores (RDKit)** e **Enviar para a análise de Hansch**. **Baixar tabela (CSV)** salva a tabela editada. O pIC50 aqui é molar (–log10 do IC50 em mol/L): vale o campo Activity do SDF (–log IC50 em µM) mais 6.

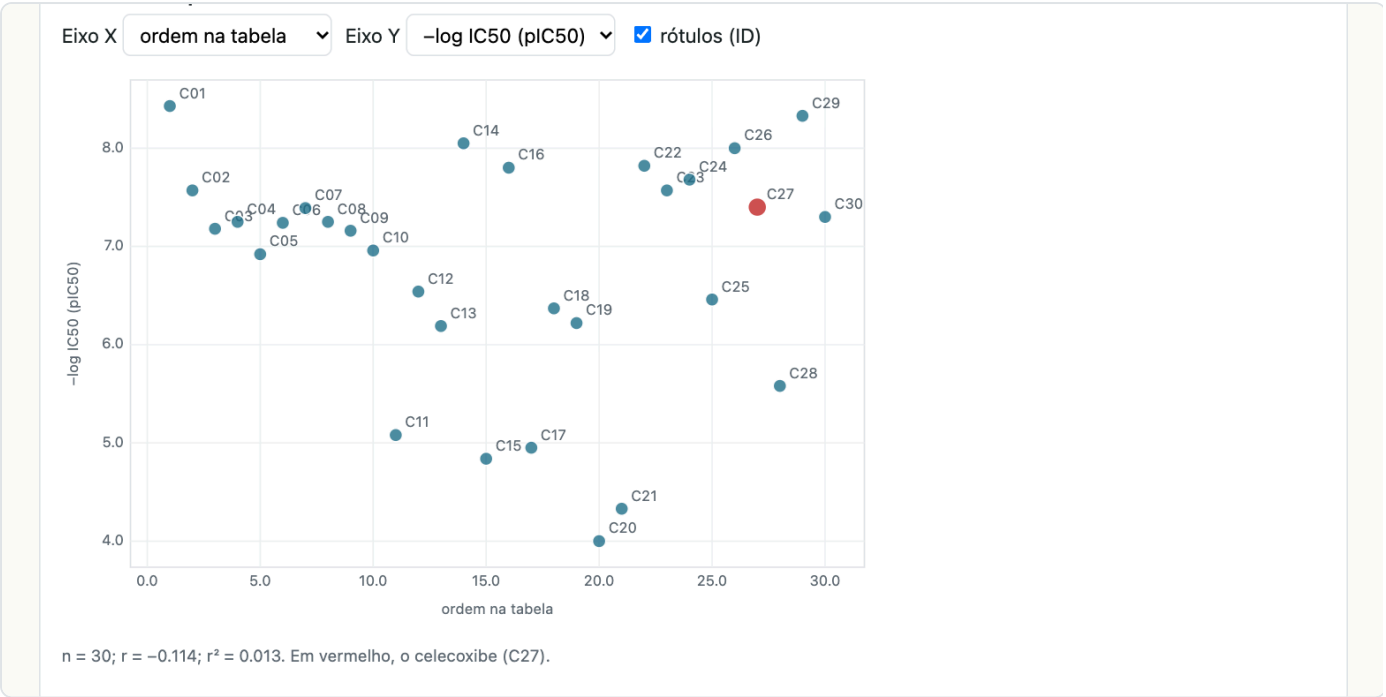
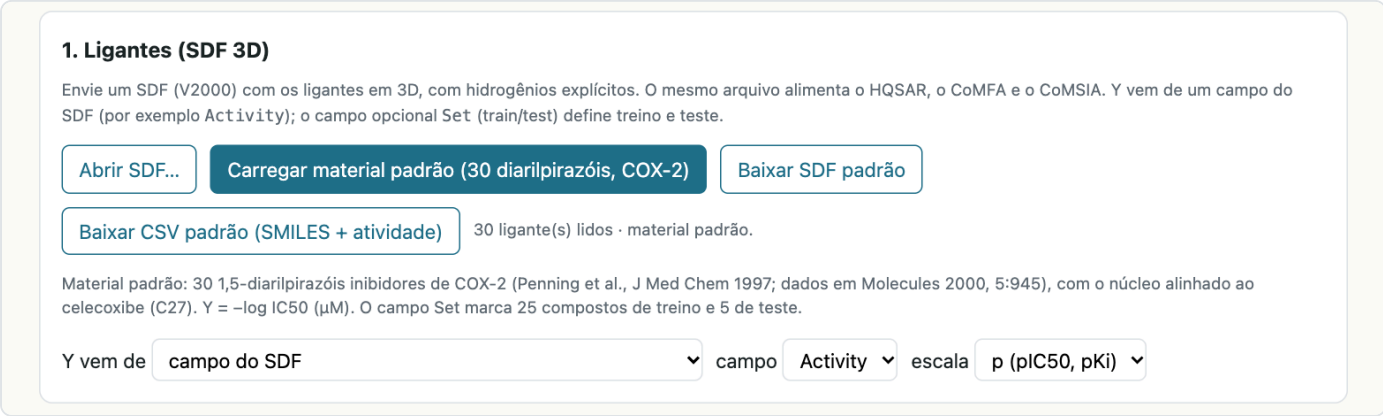


Gráfico de dispersão: escolha os eixos (ordem, IC50, log IC50, pIC50 ou um descritor). O celecoxibe aparece em vermelho; r e r² vêm abaixo do gráfico.

Os cartões seguintes da aba (regressão, validação externa, Free-Wilson, PCA/PLS, Craig, Topliss) usam os dados enviados. Os módulos 3D (HQSAR, CoMFA, CoMSIA) não dependem dessa tabela: usam o SDF da aba Ligantes 3D.

Passo 1. Carregar os ligantes (aba “Ligantes 3D”)

Clique em **Carregar material padrão** ou em **Abrir SDF...**. A linha de informação mostra quantos ligantes foram lidos e avisa se algum não tem hidrogênios. Em “Y vem de”, escolha o campo do SDF e a escala: *p* se o valor já é –log (pIC50, pKi) ou a unidade de concentração (nM, µM...), caso em que a página converte para –log.



Cartão 1: carregamento do SDF, escolha do campo de atividade e da escala.

> Se o Y estiver na aba “QSAR clássico”, escolha “coluna Y da matriz” e use nomes de moléculas iguais aos IDs da tabela.

Passo 2. Escolher as cargas parciais

Selecione **Gasteiger–Marsili** (rápido, usado por padrão) ou **MMFF94** e clique em **Aplicar cargas**. O MMFF94 usa os tipos atômicos e os incrementos de carga de ligação do campo de força do Merck (Halgren, 1996); a implementação foi comparada com o RDKit em 47 moléculas e reproduz as cargas do RDKit com diferença máxima inferior a 10^{-4} e. Íons fora do caso nitro (carboxilato, amônio…) não têm parametrização implementada e recebem Gasteiger, com um aviso ao lado do botão. "Do arquivo" lê o campo `Charges` do SDF (um valor por átomo, na ordem do arquivo).

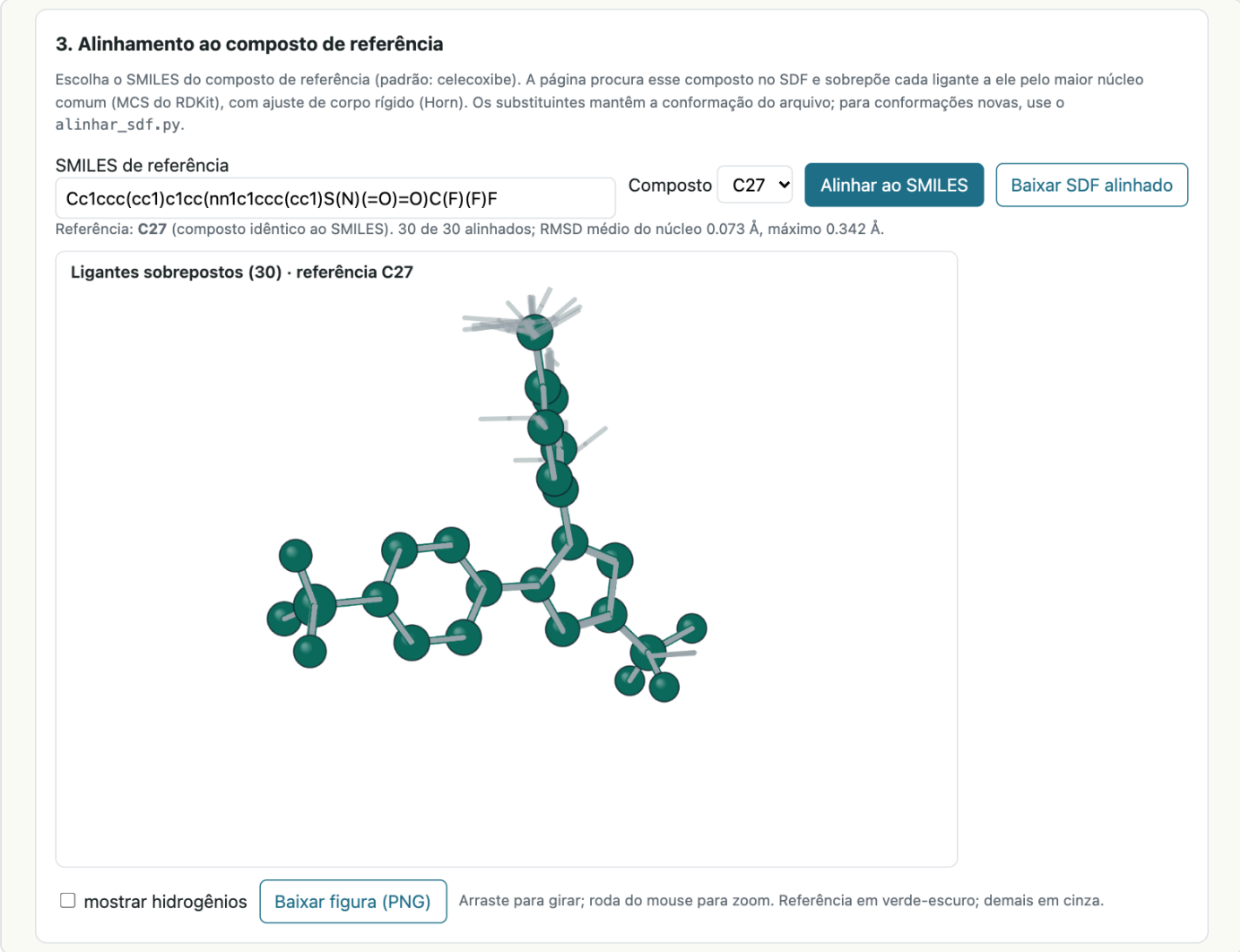


Cartão 2: método de cargas (MMFF94 aplicado).

As cargas alimentam o campo eletrostático do CoMFA e o índice eletrostático do CoMSIA; o HQSAR não as usa.

Passo 3. Alinhar ao composto de referência

O campo “SMILES de referência” vem preenchido com o celecoxibe. Clique em **Alinhar ao SMILES**: a página procura no SDF o composto idêntico ao SMILES (ou, se não houver, o que o contém como subestrutura, ou o mais parecido) e sobrepõe cada ligante a ele. O núcleo comum é encontrado pelo MCS do RDKit (anéis completos, elementos e ordem de ligação), e a superposição é de corpo rígido por quaternions (Horn). A mensagem informa a referência e o RMSD médio e máximo do núcleo; a tabela do passo 4 traz o RMSD de cada composto, e a figura mostra todos sobrepostos (referência em verde-escuro). **Baixar SDF alinhado** salva as coordenadas novas junto com o campo Set.



Cartão 3: SMILES do celecoxibe, composto de referência encontrado (C27) e sobreposição dos 30 ligantes.

> O alinhamento move cada molécula como um corpo rígido; os substituintes mantêm a conformação que tinham no arquivo. Em CoMFA/CoMSIA a conformação e o alinhamento dominam o resultado: para conformações novas por composto, use o script `alinhar_sdf.py` (RDKit, MMFF94, núcleo fixo) e carregue o SDF resultante. Trocar o SMILES (por exemplo para outro composto-modelo) refaz o alinhamento.

Passo 4. Separar treino e teste

Escolha a **divisão**: campo `Set` do SDF; aleatória estratificada por faixas de 1 unidade log (4:1, como no artigo); ou representativa (agrupamento hierárquico com ligação completa e distância euclidiana sobre fragmentos topológicos, e amostragem por faixa de atividade dentro de cada grupo). A semente torna a divisão reproduzível. Na tabela, cada composto pode ser movido para o treino ou para o teste à mão.



Cartão 4: divisão treino/teste e tabela dos ligantes (Y, conjunto, nº de átomos, soma das cargas, RMSD).

> Com poucos compostos o teste externo é instável; no artigo foram 60 de treino e 15 de teste. Aqui o padrão tem 25 e 5.

Passo 5. HQSAR

Na aba **HQSAR** escolha as *distinções de fragmento* (A átomos, B ligações, C conectividade, H hidrogênios, Ch quiralidade, DA doador/aceitador de H), os *comprimentos de holograma* (os 12 do artigo, de 53 a 401) e o intervalo de tamanho dos fragmentos (padrão 4 a 7 átomos). Com "variar o tamanho" marcado, os 5 melhores modelos são refeitos com intervalos de 2, 3 e 4 átomos, como no artigo. Clique em **Buscar modelos HQSAR**.



Parâmetros da busca HQSAR.

A tabela mostra os 12 melhores modelos por q² (leave-one-out no treino), com r², SEE e, se houver teste, r²pred, Q²(F2) e Q²(F3). Selecione o modelo desejado.

208 modelos avaliados em 0.9 s. Selecione o modelo para ver validação e mapas.

	Distinção	Fragmentos	HL	Comp.	q² (treino)	r² (ajuste)	SEE	r²pred (teste)	Q²(F2)	Q²(F3)
◉	A/C/H	5–7	61	6	0.70	0.98	0.18	0.80	0.71	0.81
○	A/C/H	6–7	61	6	0.69	0.99	0.16	0.86	0.80	0.87
○	A/B/C	5–6	59	6	0.66	0.98	0.21	0.02	−0.41	0.07
○	A/C/H	4–7	61	6	0.65	0.98	0.19	0.86	0.79	0.86
○	A/C	6–7	83	4	0.63	0.95	0.30	0.44	0.19	0.47
○	A/B/C/H	5–7	59	6	0.62	0.98	0.19	0.13	−0.26	0.18
○	A/B/C/H	4–7	59	6	0.60	0.98	0.21	−0.33	−0.92	−0.26
○	A/C	4–7	83	4	0.60	0.94	0.31	−0.06	−0.54	−0.01
○	A/B/C/H	4–6	59	6	0.59	0.97	0.25	−0.68	−1.42	−0.59
○	A/B/C	4–6	59	5	0.59	0.97	0.25	0.05	−0.37	0.10
○	A/B/C/H	5–8	59	4	0.58	0.94	0.33	−0.34	−0.93	−0.27
○	A/B/C/H	3–6	59	4	0.56	0.93	0.35	−0.32	−0.91	−0.25

Resultado da busca: modelos ordenados por q².

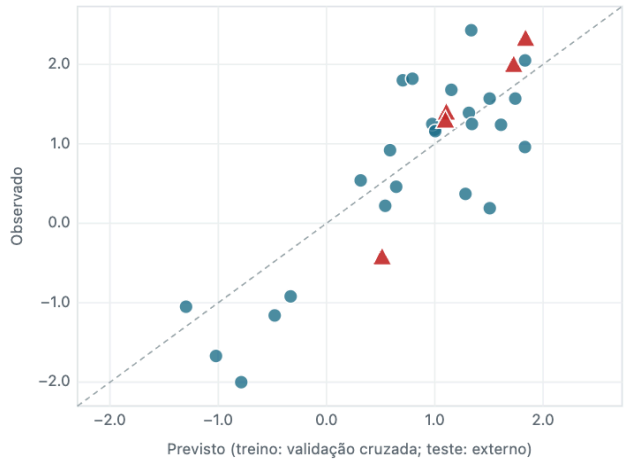
O modelo selecionado exibe os indicadores internos (q², r², SEP, SEE, nº de componentes, F) e a **validação externa** com os critérios da literatura e selos “sim/não”.

Modelo selecionado: A/C/H, fragmentos 5–7, HL 61

0.70	0.98	0.65	0.18	6	165.9	25
q² (validação cruzada)	r² (ajuste)	SEP	SEE	componentes PLS	F	compostos de treino

Validação externa (teste, n = 5)	r²pred	Q²(F2)	Q²(F3)	RMSEP	MAE	rm²	rm²′	Δrm²	CCC
valor	0.80	0.71	0.81	0.51	0.44	0.55	0.23	0.32	0.77
critério	> 0,6–0,7	> 0,7	> 0,7	menor é melhor	≤ 0,1·faixa	> 0,5	> 0,5	< 0,2	> 0,85
atende?	sim	sim	sim	–	bom	não (média 0.39)	–	não	não

Modelo selecionado: validação interna e externa (r²pred, Q²(F2), Q²(F3), RMSEP, MAE, rm², Δrm², CCC).



Observado × previsto: treino por validação cruzada (círculos) e teste (triângulos).

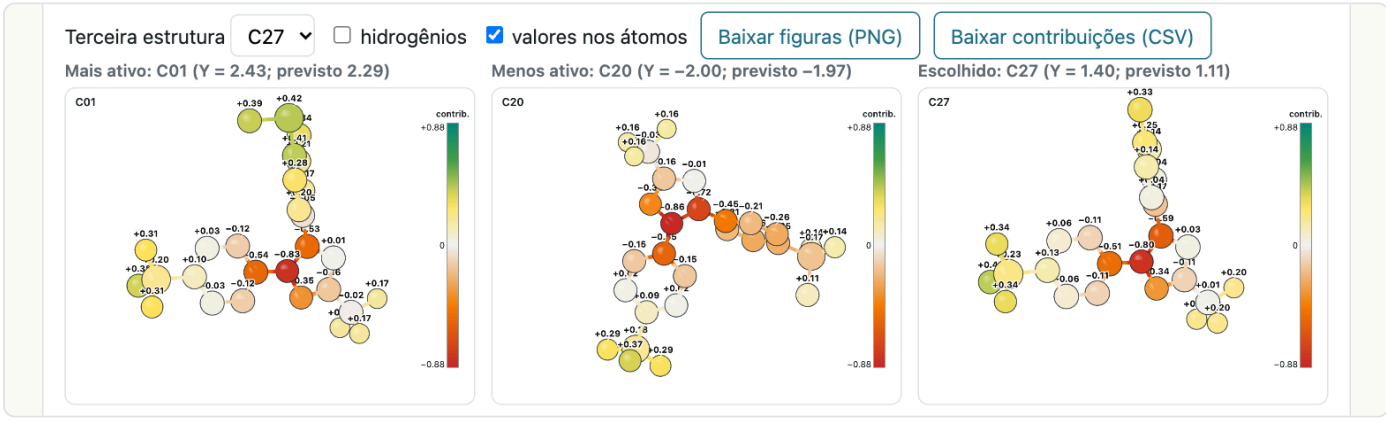
Leave-N-out (5 a 50 grupos, em triplicata) mede a estabilidade do q² diante de mudanças na composição do treino. **y-scrambling** (80 execuções) embaralha Y e refaz o modelo; o intercepto Q²′ deve ficar abaixo de 0,05 e a inclinação abaixo de 1,2 (Clark e Fox, 2004).



Leave-N-out e y-scrambling do modelo selecionado.

A figura 3D das contribuições

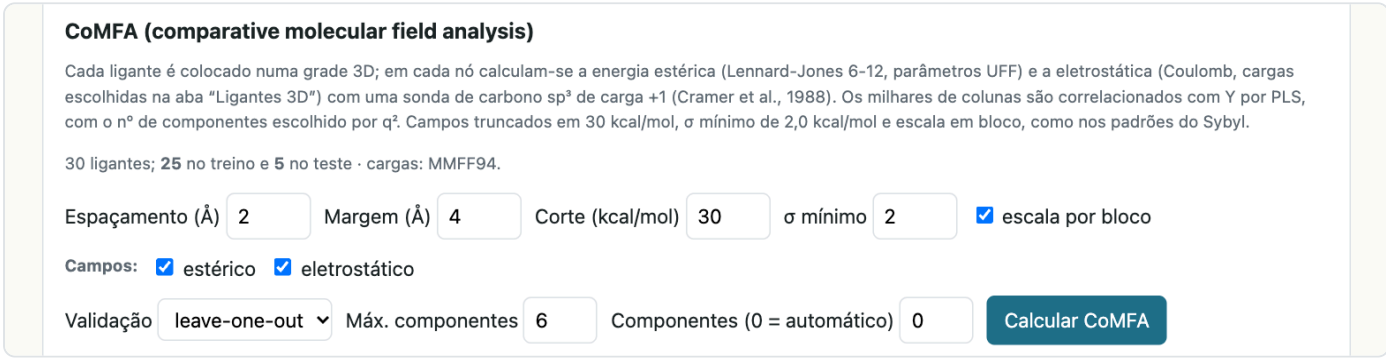
Para o composto mais ativo, o menos ativo e um terceiro à escolha, cada átomo é colorido pela soma dos coeficientes dos fragmentos a que pertence (cada coeficiente é dividido entre os átomos do fragmento). Verde, verde-azulado e amarelo indicam contribuição positiva para a atividade; laranja e vermelho, negativa; a escala é comum às três estruturas. Os valores são relativos à média do treino, e a soma por molécula é igual ao previsto menos a média de Y. Arraste para girar e use a roda do mouse para o zoom; "valores nos átomos" liga os números; **Baixar figuras (PNG)** salva as três imagens e **Baixar contribuições (CSV)** salva a tabela por átomo.



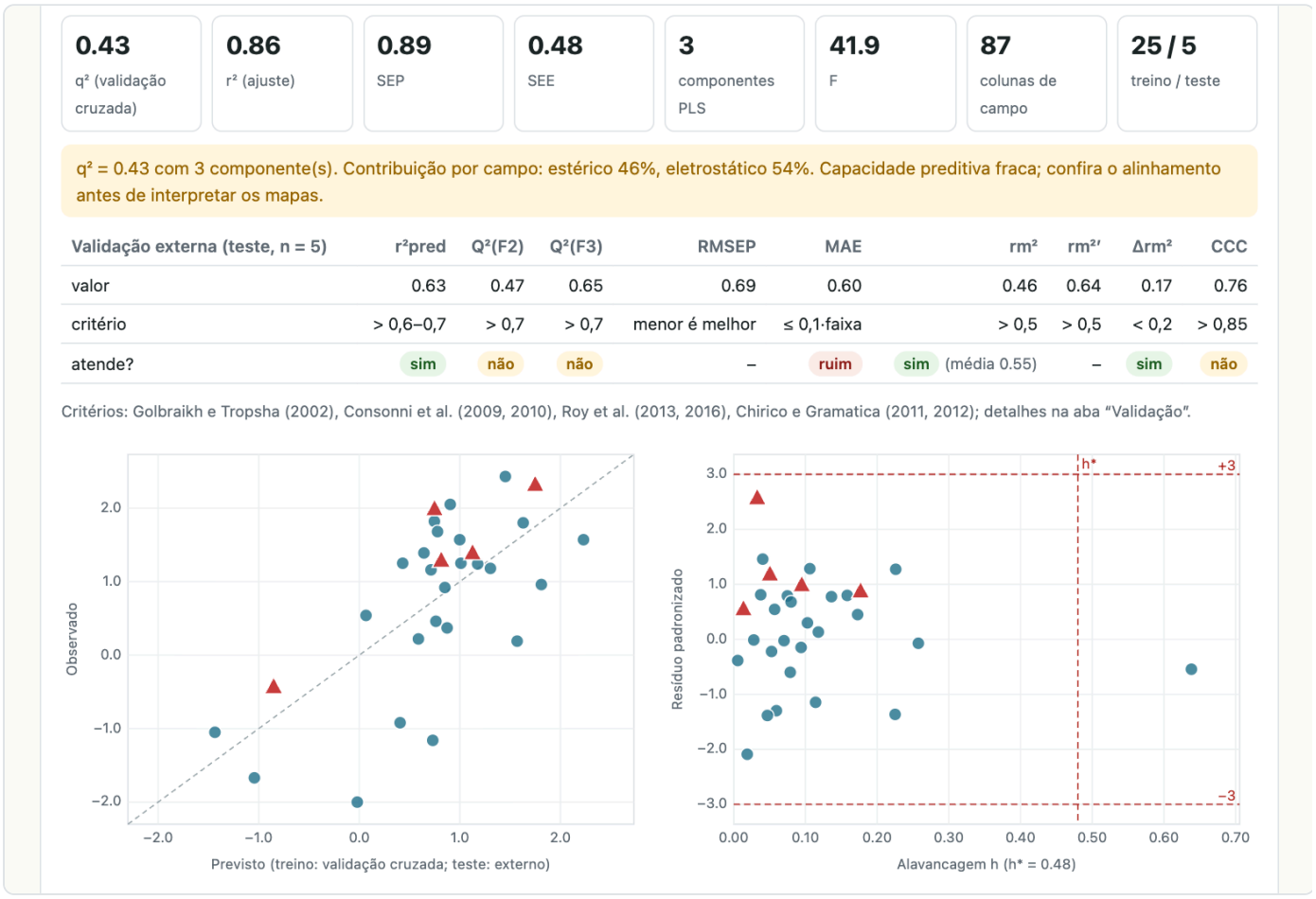
Contribuições atômicas do HQSAR em 3D: mais ativo (C01), menos ativo (C20) e composto escolhido (C27).

Passo 6. CoMFA

Defina espaçamento da grade (2 Å), margem (4 Å), corte de energia (30 kcal/mol), σ mínimo (2,0 kcal/mol), os campos (estérico e eletrostático), a validação cruzada e o número de componentes (0 = escolhe pelo melhor q² com parcimônia). **Calcular CoMFA** monta a grade, calcula Lennard-Jones (UFF) e Coulomb (cargas do passo 2) com sonda C sp³ de carga +1, e ajusta o PLS no treino.



Parâmetros do CoMFA.



Resultados do CoMFA: indicadores, validação externa, observado × previsto e gráfico de Williams (domínio de aplicabilidade).

No gráfico de Williams, compostos com alavancagem h acima de h* = 3(A+1)/n ou resíduo padronizado fora de ±3 estão fora do domínio de aplicabilidade do modelo; previsões para eles devem ser tratadas com cautela. **Leave-N-out** e o **embaralhamento progressivo de Y** (80 execuções, versão simplificada de Clark e Fox) aparecem abaixo.

Leave-N-out (3 repetições por número de grupos)

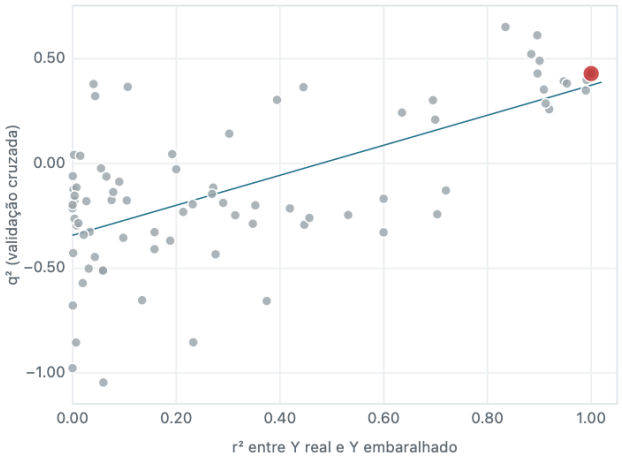
Grupos	5	10
q² médio	0.41	0.39
DP	0.13	0.10

q² médio mínimo: 0.39.

Embaralhamento progressivo de Y (80 execuções)

0.65	−0.11	−0.34	0.72
maior q² com Y embaralhado	q² médio embaralhado	intercepto Q²′ ($r^2_{yy'} = 0$)	inclinação $dQ^2/dr^2_{yy'}$

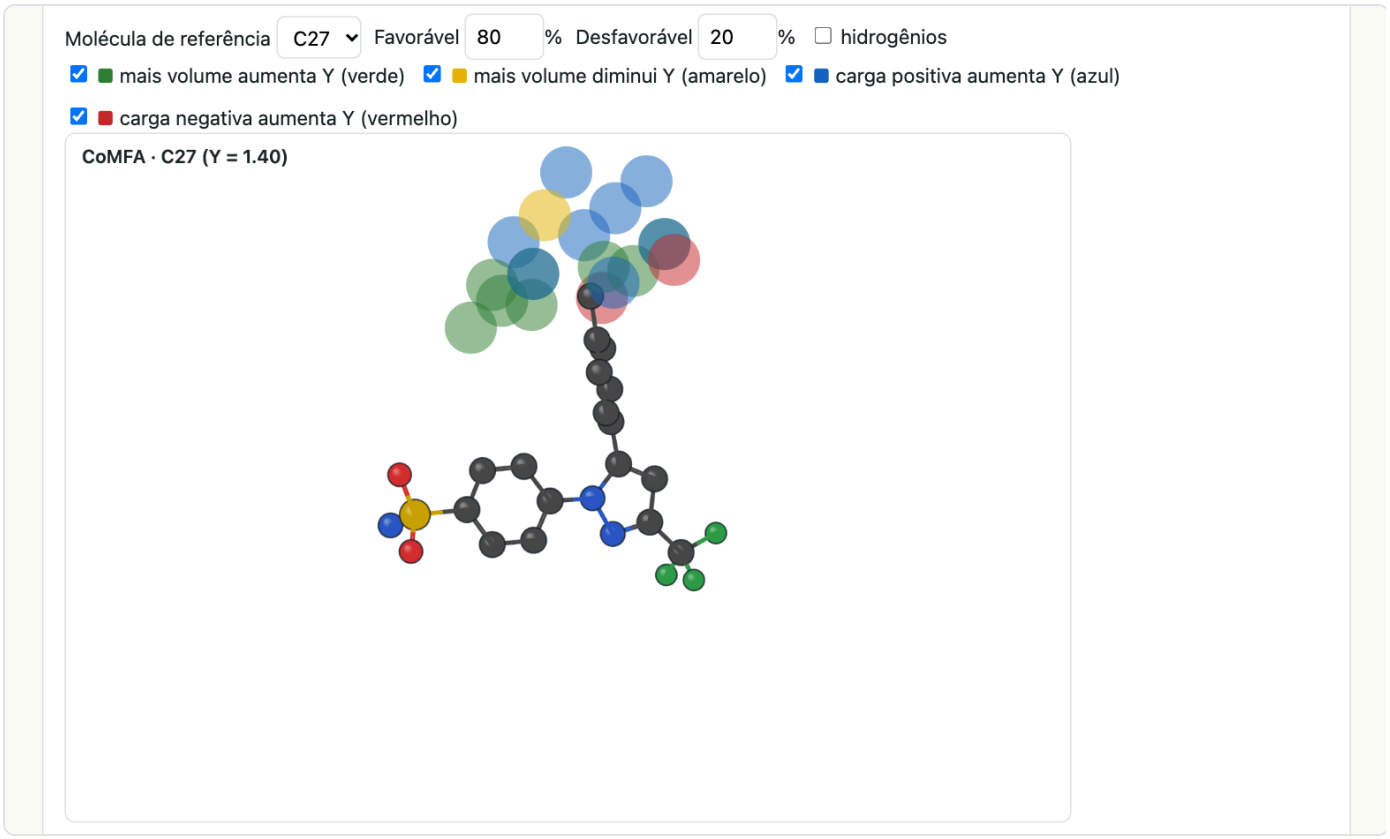
Sem sinal de correlação ao acaso: Q²′ < 0,05 e inclinação < 1,2 (critérios de Clark e Fox, 2004).



Leave-N-out e embaralhamento progressivo do CoMFA.

Mapas de contorno 3D

Como na Figura 5 do artigo, os contornos favoráveis envolvem 80% e os desfavoráveis 20% da contribuição acumulada (coeficiente × desvio-padrão) de cada campo; os dois valores são ajustáveis. **Estérico:** verde = mais volume aumenta a atividade; amarelo = mais volume a diminui. **Eletrostático:** azul = carga positiva aumenta; vermelho = carga negativa aumenta. Escolha a molécula de referência, ligue ou desligue cada cor, gire e dê zoom. **Baixar figura (PNG)**, **PDB para o ChimeraX** e **CSV de coeficientes** permitem levar o mapa para outros programas.



Mapa de contornos do CoMFA ao redor do celecoxibe (C27).

Passo 7. CoMSIA

O CoMSIA usa índices de similaridade gaussianos (α = 0,3) para cinco campos: estérico (volume de van der Waals), eletrostático (carga parcial), hidrofóbico (tabela atômica simplificada tipo Viswanadhan), doador e aceitador de ligação de hidrogênio. Clique em **Buscar melhor combinação (31)** para testar todas as combinações e marcar a de maior q²; depois, **Calcular CoMSIA**.

CoMSIA (comparative molecular similarity indices analysis)

Em vez de energias, CoMSIA usa índices de similaridade com função gaussiana (Klebe et al., 1994, α = 0,3), sem singularidades na superfície molecular: estérico (volume de van der Waals), eletrostático (carga parcial), hidrofóbico (parametrização atômica simplificada tipo Viswanadhan), doador e aceitador de ligação de hidrogênio. O botão de busca testa as 31 combinações de campos pelo q² (como no artigo) e escolhe a melhor.

30 ligantes; **25** no treino e **5** no teste · cargas: MMFF94.

Espaçamento (Å) 2 Margem (Å) 4 α (gaussiana) 0,3 σ mínimo 0,05 ☑ escala por bloco

Campos: ☐ estérico ☑ eletrostático ☐ hidrofóbico ☑ doador de H ☑ aceitador de H [Buscar melhor combinação (31)]

Melhor combinação: **eletrostático + doador de H + aceitador de H** (q² = 0.64). Marcada acima; clique em "Calcular CoMSIA".

Combinação	1º	2º	3º	4º	5º	6º	7º	8º
campos	EDA	EHA	EHDA	EA	SEA	SEDA	SEHA	SEH
q² (LOO)	0.64	0.62	0.60	0.59	0.59	0.59	0.58	0.57
comp.	2	2	2	2	2	2	2	3

S estérico, E eletrostático, H hidrofóbico, D doador, A aceitador.

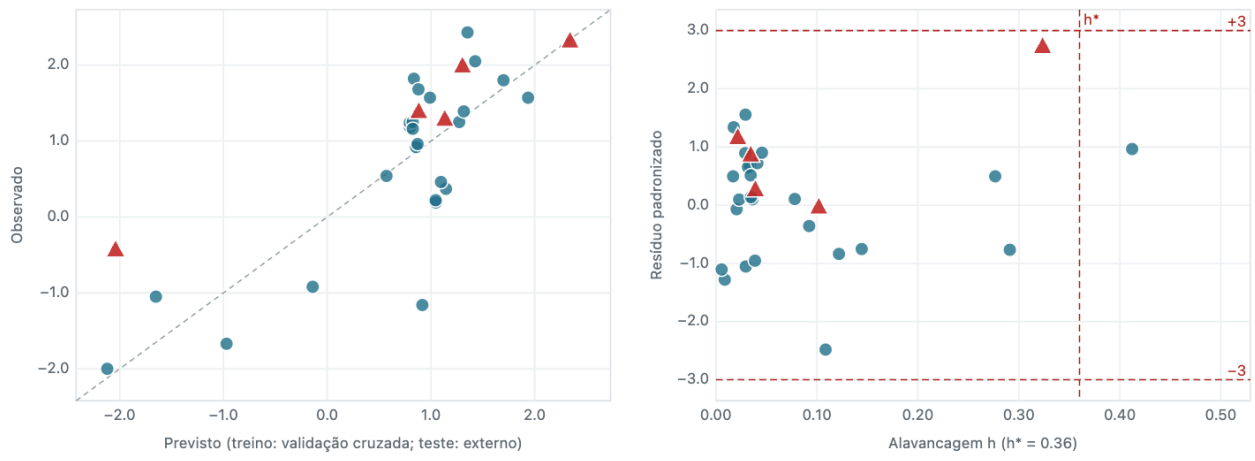
Busca da melhor combinação de campos do CoMSIA (31 combinações, ordenadas pelo q²).

0.64	0.78	0.71	0.59	2	38.4	51	25 / 5
q² (validação cruzada)	r² (ajuste)	SEP	SEE	componentes PLS	F	colunas de campo	treino / teste

q² = 0.64 com 2 componente(s). Contribuição por campo: eletrostático 57%, doador de H 14%, aceitador de H 29%. Robustez interna aceitável (q² ≥ 0,5); confirme com o teste externo e com o embaralhamento de Y.

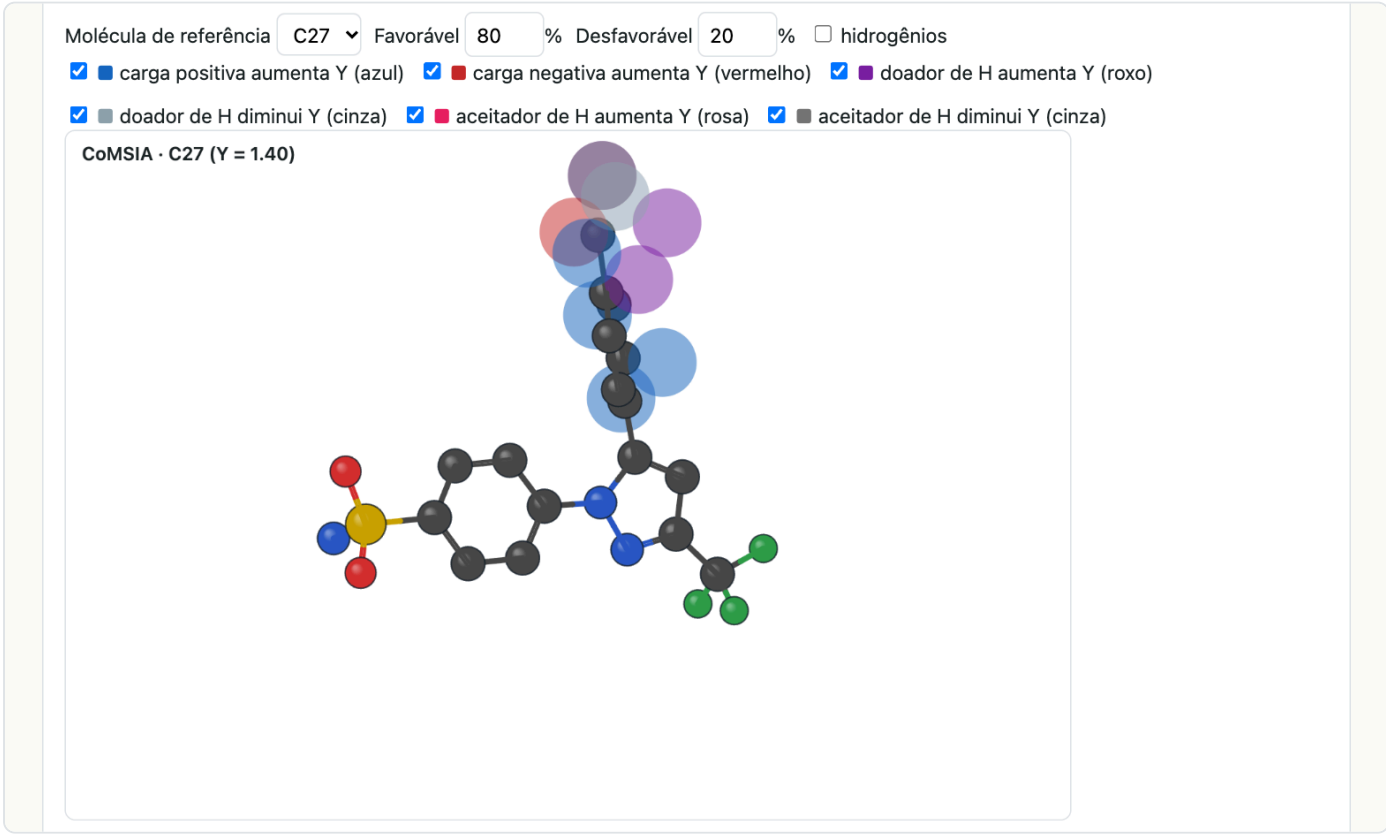
Validação externa (teste, n = 5)	r²pred	Q²(F2)	Q²(F3)	RMSEP	MAE	rm²	rm²′	Δrm²	CCC
valor	0.48	0.25	0.50	0.83	0.60	0.18	0.47	0.29	0.80
critério	> 0,6–0,7	> 0,7	> 0,7	menor é melhor	≤ 0,1·faixa	> 0,5	> 0,5	< 0,2	> 0,85
atende?	não	não	não	–	ruim	não (média 0.33)	–	não	não

Critérios: Golbraikh e Tropsha (2002), Consonni et al. (2009, 2010), Roy et al. (2013, 2016), Chirico e Gramatica (2011, 2012); detalhes na aba "Validação".



Resultados do CoMSIA: indicadores, validação externa e gráficos.

Cores dos contornos: estérico verde/amarelo; eletrostático azul/vermelho; hidrofóbico laranja (favorável) e cinza-claro (desfavorável); doador roxo e cinza; aceitador rosa e cinza (favorável e desfavorável, respectivamente).



Mapa de contornos do CoMSIA.

Passo 8. Validação consolidada

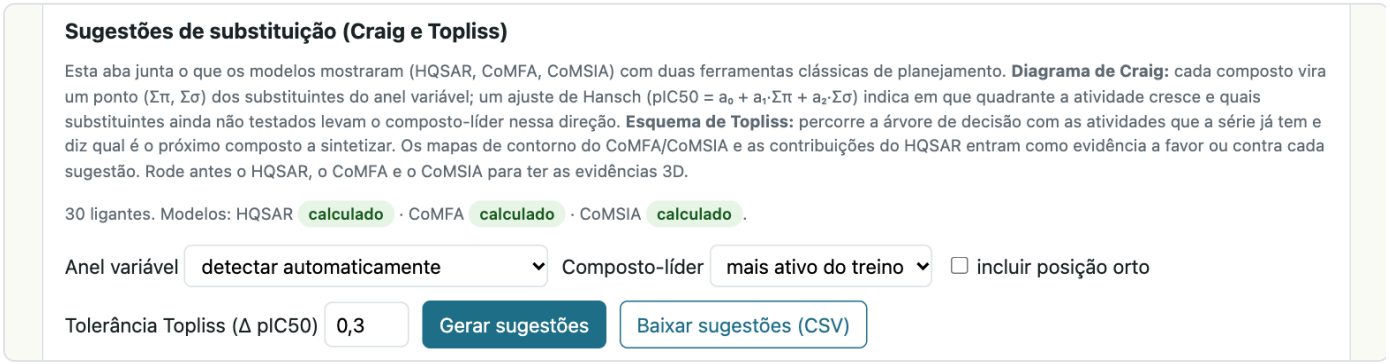
A aba **Validação** reúne os modelos calculados na sessão em uma tabela no formato da Tabela 4 do artigo, com selos verdes quando o valor atende ao critério da literatura e amarelos quando não atende. Abaixo, os gráficos observado × previsto do teste (Figura 3 do artigo) e um glossário com a definição, o critério e a referência de cada métrica. **Baixar tabela (CSV)** exporta tudo.



Tabela de validação dos modelos e glossário de métricas.

Passo 9. Sugestões de substituição (aba “Sugestões (Craig e Topliss)”)

Depois de rodar HQSAR, CoMFA e CoMSIA, abra a aba e clique em **Gerar sugestões**. A página mapeia cada ligante no composto de referência (MCS do RDKit), identifica o **anel variável** (o aromático de 6 átomos com mais padrões de substituição; pode ser trocado na lista), lê os substituintes por posição (orto, meta, para) e reconhece os grupos pela estrutura (Cl, CH_3, OCH_3, SCH_3, CF_3, NO_2, SO_2NH_2 etc.). As constantes π, σm, σp e MR são as de Hansch, Leo e Taft, as mesmas da aba QSAR clássico. Ajuste a **tolerância do Topliss** (Δ de pIC50 que separa “igual” de “mais/menos ativo”) e o **composto-líder** (padrão: o mais ativo do treino). Posições orto ficam de fora por padrão, porque o ajuste não modela o efeito estérico sobre o ângulo entre os anéis; marque “incluir posição orto” para vê-las.



Controles: anel variável, composto-líder, posição orto e tolerância do Topliss.

1. Série por posição. Para cada composto, os substituintes do anel, Σπ, Σσ, ΣMR, o pIC50, o conjunto (treino/teste) e o previsto pelo ajuste de Hansch. Substituintes que não constam na tabela aparecem marcados (“?”) e ficam fora do ajuste. Quando compostos com o mesmo padrão de substituição têm atividades diferentes (por exemplo CF_3 × CHF_2 no pirazol), a página avisa que a variação vem de outra parte da molécula.

1. Série por posição (anel 1; 29 compostos)

Composto	Substituintes	Σπ	Σσ	ΣMR	pIC50	Conjunto	Previsto (Hansch)
C01 líder	3-CH3, 4-SCH3	1.17	−0.07	1.95	2.43	treino	1.46
C29	4-N(CH3)2	0.18	−0.83	1.55	2.33	teste	1.24
C14	4-SCH3	0.61	0.00	1.38	2.05	treino	0.89
C26	4-Cl	0.71	0.23	0.60	2.00	teste	0.78
C22	4-OCH3	−0.02	−0.27	0.79	1.82	treino	0.56
C16	4-NHCH3	−0.47	−0.70	1.03	1.80	treino	0.54
C24	3-Cl, 4-OCH3, 5-Cl	1.40	0.47	1.99	1.68	treino	1.19

Série por posição: substituintes, Σπ, Σσ, ΣMR, pIC50, conjunto e previsto.

2. Diagrama de Craig e ajuste de Hansch. Ajuste de $pIC50 = a_0 + a_1 \cdot \Sigma\pi + a_2 \cdot \Sigma\sigma$ no treino, com r^2, q^2 (LOO) e s, e a indicação do **quadrante** em que a atividade cresce (por exemplo, π+ e σ−: lipofílicos e doadores de elétrons). O gráfico mostra cada composto como ponto (Σπ, Σσ) colorido pela atividade, o líder em destaque, as sugestões numeradas em laranja e uma reta com a direção de maior ganho. Com q^2 baixo (< 0,3) a página avisa que a relação com π e σ é fraca e que o resultado é só uma orientação.

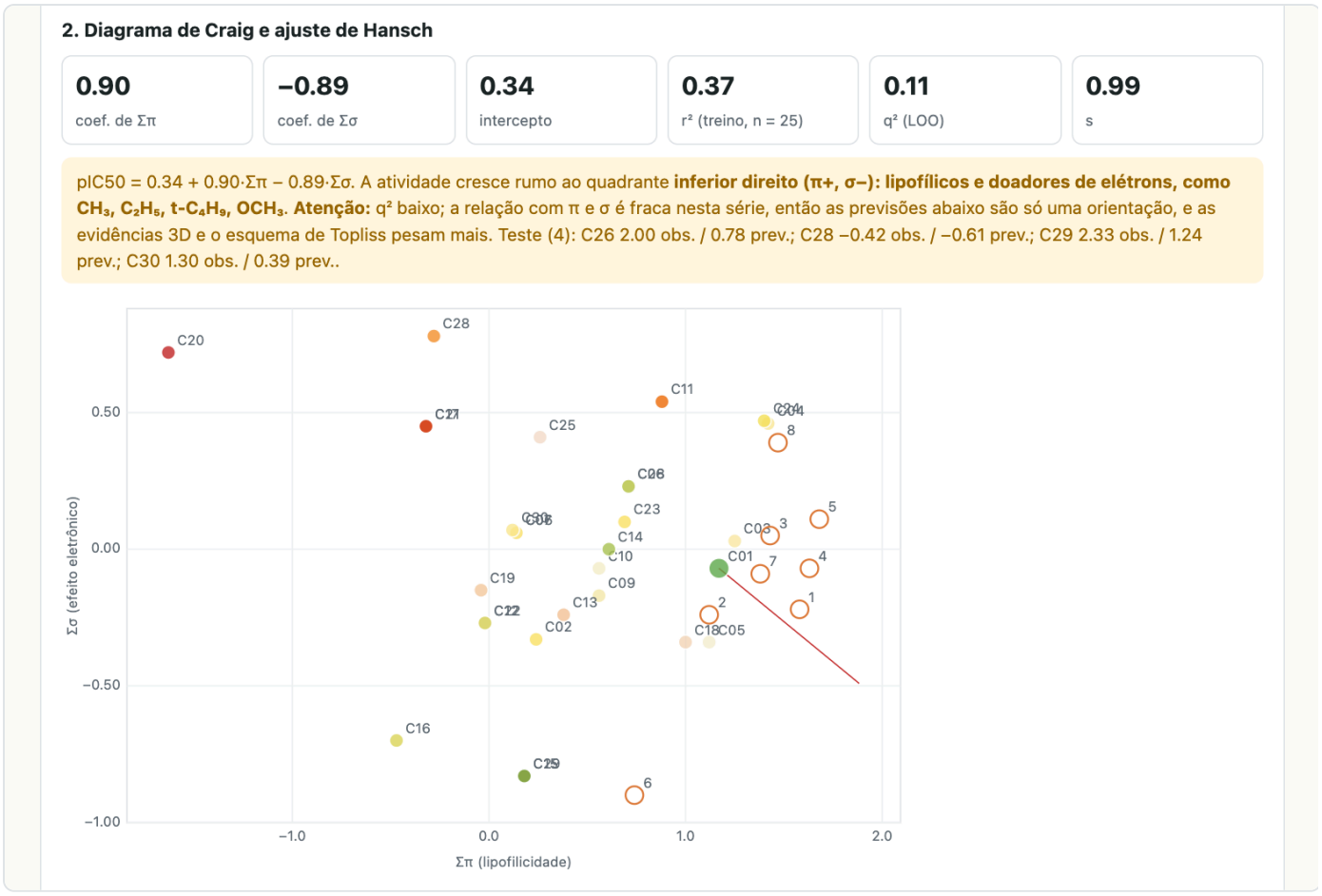


Diagrama de Craig da série: pontos coloridos pelo pIC50, sugestões numeradas e direção de ganho a partir do líder.

3. Substituições sugeridas. Para cada posição (para e meta, e orto se marcada) a página troca o substituinte do líder por cada candidato da tabela de Hansch-Leo-Taft, calcula Σπ e Σσ novos e o pIC50 previsto (líder + ganho do ajuste); candidatos já presentes na série são excluídos e os que saem da faixa de π ou σ medida recebem o selo "extrapolação". A coluna **Evidência dos modelos 3D** cruza a troca com os contornos do CoMFA (CF) e do CoMSIA (CS) perto da posição no líder: ✓ significa que o candidato vai no sentido que o contorno pede (mais volume na região verde do estérico, maior π na região hidrofóbica favorável, retirador de elétrons na região azul etc.); ✕, que vai no sentido oposto. Passe o mouse sobre o selo para ler o motivo. O ranking soma o ganho de Hansch e um pequeno peso dessas evidências.

3. Substituições sugeridas a partir do líder C01 (3-CH3, 4-SCH3; pIC50 2.43)

#	Troca	Σπ	Σσ	ΣMR	Δπ	Δσ	pIC50 previsto	Evidência dos modelos 3D
1	**para (4): SCH3 → C2H5**	1.58	−0.22	1.59	+0.41	−0.15	**2.93** (+0.50)	CF estérico ✕ · CF eletrostático ✓ · CS eletrostático ✓ · CS doador de H ✓ · CS aceitador de H ✓
2	**para (4): SCH3 → CH3**	1.12	−0.24	1.13	−0.05	−0.17	**2.54** (+0.11)	CF estérico ✕ · CF eletrostático ✓ · CS eletrostático ✓ · CS doador de H ✓ · CS aceitador de H ✓
3	**meta (3): CH3 → CH=CH2**	1.43	0.05	2.48	+0.26	+0.12	**2.56** (+0.13)	CF estérico ✓ · CF eletrostático ✓ · CS eletrostático ✓ · CS doador de H ✕ · CS aceitador de H ✓
4	**meta (3): CH3 → C2H5**	1.63	−0.07	2.41	+0.46	+0.00	**2.85** (+0.42)	CF estérico ✓ · CF eletrostático ✕ · CS eletrostático ✕ · CS doador de H ✕ · CS aceitador de H ✓
5	**para (4): SCH3 → I**	1.68	0.11	1.96	+0.51	+0.18	**2.73** (+0.30)	CF estérico ✕ · CF eletrostático ✕ · CS eletrostático ✕ · CS doador de H ✓ · CS aceitador de H ✓
6	**para (4): SCH3 → N(CH3)2**	0.74	−0.90	2.12	−0.43	−0.83	**2.78** (+0.35)	CF estérico ✓ · CF eletrostático ✓ · CS eletrostático ✓ · CS doador de H ✕ · CS aceitador de H ✕
7	**para (4): SCH3 → CH=CH2**	1.38	−0.09	1.66	+0.21	−0.02	**2.64** (+0.21)	CF estérico ✕ · CF eletrostático ✕ · CS eletrostático ✕ · CS doador de H ✓ · CS aceitador de H ✓
8	**meta (3): CH3 → Br**	1.47	0.39	2.27	+0.30	+0.46	**2.29** (−0.14)	CF estérico ✓ · CF eletrostático ✓ · CS eletrostático ✓ · CS doador de H ✕ · CS aceitador de H ✓
9	**meta (3): CH3 → SCH3**	1.22	0.15	2.76	+0.05	+0.22	**2.28** (−0.15)	CF estérico ✓ · CF eletrostático ✓ · CS eletrostático ✓ · CS doador de H ✕ · CS aceitador de H ✓
10	**para (4): SCH3 → OC2H5**	0.94	−0.31	1.81	−0.23	−0.24	**2.44** (+0.01)	CF estérico ✕ · CF eletrostático ✓ · CS eletrostático ✓ · CS doador de H ✓ · CS aceitador de H ✕

Ordenação: ganho previsto pelo ajuste de Hansch somado a um pequeno peso das evidências 3D (cada campo cujo contorno vizinho é atendido ou contrariado pelo candidato conta +1 ou −1, ponderado pela importância do contorno; CF = CoMFA, CS = CoMSIA, ✓ = o candidato vai no sentido que o contorno pede, ✕ = no sentido oposto; passe o mouse para ver o motivo). Os candidatos vêm da tabela de substituintes (Hansch-Leo-Taft); combinações já presentes na série foram excluídas. "Extrapolação" = Σπ ou Σσ fora da faixa medida (± margem). São hipóteses a testar, não previsões garantidas.

4. O que os modelos dizem em cada posição do líder

Substituições sugeridas a partir do líder, com Σπ, Σσ, ΣMR, Δπ, Δσ, pIC50 previsto e evidência dos modelos 3D.

4. O que os modelos dizem em cada posição. Mostra o substituinte atual do líder em cada posição e a contribuição do HQSAR para o grupo: negativa indica um bom ponto de troca; positiva, que convém manter ou trocar por grupo de propriedades parecidas; posição livre (H) é candidata a receber um substituinte.

5. Esquema de Topliss. A página percorre a árvore de decisão (aquela da seção 13 da aba QSAR clássico) com as atividades da série: parte do composto parente (H), compara o 4-Cl com ele e classifica como mais ativo (M), igual (E) ou menos ativo (L), seguindo o ramo correspondente até o primeiro composto da árvore que ainda não foi sintetizado. Se a série não tem o composto H, como no material padrão, o primeiro passo não pode ser classificado: a página mostra o que fazer (medir o H) e lista os nós da árvore que a série já cobre, com a diferença de atividade em relação ao 4-Cl.

4. O que os modelos dizem em cada posição do líder

Posição	Substituinte atual	Contribuição HQSAR	Leitura
2 (orto)	H	–	posição livre (H): candidata a ganhar um substituinte
3 (meta)	CH3	0.34	contribuição positiva: manter ou trocar por grupo de propriedades parecidas
4 (para)	SCH3	0.81	contribuição positiva: manter ou trocar por grupo de propriedades parecidas
5 (meta)	H	–	posição livre (H): candidata a ganhar um substituinte
6 (orto)	H	–	posição livre (H): candidata a ganhar um substituinte

5. Esquema de Topliss com as atividades da série

Nó do esquema	Composto(s)	pIC50	Δ frente ao 4-Cl
4-Cl	C26 (2.00)	2.00	–
4-OCH_3	C22 (1.82)	1.82	−0.18
4-$N(CH_3)_2$	C29 (2.33)	2.33	+0.33
4-CF_3 [Br, I]	C11 (−0.92)	−0.92	−2.92
3-CH_3	C10 (0.96)	0.96	−1.04
2,4-Cl_2	C04 (1.25)	1.25	−0.75
2-Cl; 2-CH_3; 2-OCH_3	C08 (1.25)	1.25	−0.75
4-NO_2	C28 (−0.42)	−0.42	−2.42
4-F	C07 (1.39)	1.39	−0.61

Próximo passo (Topliss): a série não tem o composto parente (anel sem substituinte, H), e o primeiro passo do esquema compara o 4-Cl com o H. Sintetize e meça o H para classificar o 4-Cl como mais ativo (M → 3,4-Cl_2), igual (E → 4-CH_3) ou menos ativo (L → 4-OCH_3). Enquanto isso, a tabela acima mostra os nós do esquema que a série já cobre, com o 4-Cl como referência.

Seções 4 e 5: leitura por posição (HQSAR) e esquema de Topliss com os nós da árvore já presentes na série.

As sugestões são hipóteses para planejar a próxima rodada de síntese, não previsões garantidas. O ajuste de Hansch usa só o anel variável (outras diferenças não entram), os contornos 3D dependem do alinhamento, e com 25 compostos de treino o q² do ajuste costuma ser baixo. Use o resultado em conjunto com a validação dos modelos e com o conhecimento químico da série. **Baixar sugestões (CSV)** exporta a tabela.

4. Como ler os resultados

Etapa	Critério usado no artigo e na literatura
Validação interna	q² (LOO) > 0,6 para seguir à validação externa; q² médio do leave-N-out > 0,8 com DP baixo indica consistência.
Validação externa	r²pred e $Q^2(F2)$, $Q^2(F3)$ > 0,7; RMSEP baixo; MAE ≤ 0,1 × faixa de Y do treino; rm^2 médio > 0,5 e Δrm^2 < 0,2; CCC > 0,85.
Correlação ao acaso	y-scrambling / embaralhamento progressivo: $Q^{2\prime}$ < 0,05 e inclinação < 1,2.
Domínio de aplicabilidade	alavancagem h ≤ h* = 3(A+1)/n e |resíduo padronizado| ≤ 3 (Gramatica, 2007; princípios da OCDE).

Com o material padrão (25 treino, 5 teste) alguns critérios externos não são atendidos em todos os modelos. Isso é esperado: são 5 compostos de teste, e um único composto mal previsto muda r²pred e $Q^2(F2)$ de forma drástica. Para conclusões, use séries maiores (o artigo usou 75 compostos) e confira o resultado com um teste realmente independente.

5. Salvar em PDF e exportar

- **Figuras:** cada visualizador 3D tem botão de PNG (resolução 2×).
- **Tabelas:** previsões, coeficientes, contribuições atômicas e a tabela de validação saem em CSV (separador ";", vírgula decimal).
- **SDF alinhado:** botão na aba Ligantes 3D.
- **Página inteira em PDF:** Imprimir → Salvar como PDF (Ctrl/Cmd + P). Só a aba aberta é impressa; o botão "Imprimir esta página" na aba Guia faz o mesmo.
- **Este guia em PDF:** `guia_SnorMol_QSAR.pdf`, na mesma pasta da página. Para refazê-lo depois de mudanças: `node src/make_guide.js`.

6. Problemas comuns

Mensagem ou sintoma	O que fazer
"SDF em formato V3000 não é suportado"	Salve o SDF como V2000 (Open Babel: `obabel in.sdf -O out.sdf`; RDKit grava V2000 por padrão).
"Moléculas sem hidrogênios explícitos"	Adicione os hidrogênios em 3D antes de alinhar (RDKit: `Chem.AddHs(m, addCoords=True)`).
"núcleo comum muito pequeno" no alinhamento	O composto tem pouco em comum com a referência; escolha outra referência na lista ou outro SMILES.
Aviso de Gasteiger de reserva no MMFF94	A molécula tem íon fora do caso nitro ou grupo sem tipo MMFF implementado; use Gasteiger para o conjunto todo, para manter as cargas comparáveis.
"Grade com mais de 60000 pontos"	Aumente o espaçamento ou diminua a margem.
q^2 baixo no CoMFA/CoMSIA	Revise o alinhamento e as conformações antes de qualquer outra coisa; depois teste σ mínimo, cargas e a combinação de campos.
HQSAR com "limite de fragmentos"	Reduza o tamanho máximo enumerado (padrão 9).

7. Limites e transparência

- O HQSAR usa invariantes de subgrafo conexo com hash em bins; **não é idêntico** ao algoritmo proprietário do Sybyl, e o mapa de contribuições divide o coeficiente de cada fragmento igualmente entre seus átomos.
- O CoMFA usa energias UFF/Coulomb, e não o campo de força Tripos; o CoMSIA usa uma tabela de hidrofobicidade atômica reduzida; o embaralhamento progressivo é uma versão simplificada. Os números podem, portanto, diferir dos do Sybyl.
- O MMFF94 está implementado para C, H, N, O, F, S, Cl, Br e I em moléculas neutras (e nitro). Outros casos recebem Gasteiger.
- Ferramenta de análise exploratória e didática; não substitui revisão estatística nem validação experimental.

8. Referências

1. Ferreira GM, de Magalhães JG, Maltarollo VG, Kronenberger T, Ganesan A, Emery FS, Trossini GHG. QSAR studies on the human sirtuin 2 inhibition by non-covalent 7,5,2-anilinobenzamide derivatives. *J Biomol Struct Dyn* 2019. doi:10.1080/07391102.2019.1574603 (modelo de análise e de validação deste programa).
2. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). *J Am Chem Soc* 1988;110:5959.
3. Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA). *J Med Chem* 1994;37:4130.
4. Tong W, Lowis DR, Perkins R, Chen Y, Welsh WJ, Goddette DW, Heritage TW, Sheehan DM. Evaluation of quantitative structure–activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J Chem Inf Comput Sci* 1998;38:669 (HQSAR).
5. Seel M, Turner DB, Willett P. Effect of parameter variations on the effectiveness of HQSAR analyses. *Quant Struct-Act Relat* 1999;18:245.
6. Halgren TA. Merck molecular force field. I–V. *J Comput Chem* 1996;17:490–641 (MMFF94, cargas parciais).
7. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity. *Tetrahedron* 1980;36:3219.
8. Golbraikh A, Tropsha A. Beware of q². *J Mol Graph Model* 2002;20:269.
9. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 2010;29:476.
10. Schüürmann G, Ebert RU, Chen J, Wang B, Kühne R. External validation and prediction employing the predictive squared correlation coefficient. *J Chem Inf Model* 2008;48:2140.
11. Consonni V, Ballabio D, Todeschini R. Comments on the definition of the Q² parameter for QSAR validation. *J Chem Inf Model* 2009;49:1669; Evaluation of model predictive ability by external validation techniques. *J Chemometrics* 2010;24:194.
12. Chirico N, Gramatica P. Real external predictivity of QSAR models: how to evaluate it? *J Chem Inf Model* 2011;51:2320; 2012;52:2044 (CCC).
13. Roy PP, Roy K. On some aspects of variable selection for PLS regression models (rm²). *QSAR Comb Sci* 2008;27:302; Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN. Some case studies on application of "rm²" metrics. *J Comput Chem* 2013;34:1071.
14. Roy K, Das RN, Ambure P, Aher RB. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemom Intell Lab Syst* 2016;152:18 (MAE).
15. Kunal R, Supratik K, Das RN. *Statistical Methods in QSAR/QSPR*. Springer, 2015.
16. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255.
17. Clark RD, Fox PC. Statistical variation in progressive scrambling. *J Comput Aided Mol Des* 2004;18:563.
18. Gramatica P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 2007;26:694 (princípios da OCDE e domínio de aplicabilidade).
19. Rappé AK et al. UFF. *J Am Chem Soc* 1992;114:10024. RDKit: Landrum G, rdkit.org (percepção de aromaticidade, MCS).
20. Penning TD et al. Synthesis and biological evaluation of the 1,5-diarylpyrazole class of COX-2 inhibitors. *J Med Chem* 1997;40:1347 (material padrão).

© 2026 LiBNM – Laboratório Integrativo de Bioprospecção de Novas Moléculas. Dúvidas e relatos de erro: gmf@usp.br (Glaucio M. Ferreira).