

UNIVERSIDADE ESTADUAL DE GOIÁS
UNIDADE UNIVERSITÁRIA DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
MESTRADO EM CIÊNCIAS MOLECULARES

**PREVISÃO DA SOLUBILIDADE DE FÁRMACOS ORGÂNICOS
EM ÁGUA, BASEADO NA RELAÇÃO QUANTITATIVA ENTRE
ESTRUTURA MOLECULAR E PROPRIEDADES *IN SILICO*.**

Edilson Pinheiro Peixoto

ANÁPOLIS – GO
Junho/2010

UNIVERSIDADE ESTADUAL DE GOIÁS
UNIDADE UNIVERSITÁRIA DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
MESTRADO EM CIÊNCIAS MOLECULARES

PREVISÃO DA SOLUBILIDADE DE FÁRMACOS ORGÂNICOS EM ÁGUA,
BASEADO NA RELAÇÃO QUANTITATIVA ENTRE ESTRUTURA
MOLECULAR E PROPRIEDADES *IN SILICO*.

Dissertação apresentada à banca do
Programa de Mestrado em Ciências
Moleculares da Universidade Estadual de
Goiás como parte dos requisitos
necessários para obtenção do título de
Mestre em Ciências Moleculares.

Autor: Edilson Pinheiro Peixoto

Orientador: Prof. Dr. Ademir J. Camargo

ANÁPOLIS – GO
Junho/2010

Dedicatória

Ao meu filho Pedro pelas colaborações fantásticas e à minha mulher, Do Ó, pela assessoria impecável fora da dissertação e disciplinas, a vida.

AGRADECIMENTOS

Aos meus pais, in memória. Meu pai pela dedicação e visualização econômica da sociedade para a família e minha mãe pela visualização da educação como segurança aos filhos.

Aos meus professores, todos, da tenra idade à Universidade, figura social que representa forte pilar na construção de uma sociedade legal.

Ao amigo e orientador, Prof. Dr. Ademir João Camargo. O estímulo, a paciência nos passos de trabalho, a visão de apontar caminhos de esclarecimento frente ao desconhecido.

À Universidade Estadual de Goiás em especial à Unidade Universitária de Ciências Exatas e Tecnológicas, professores, técnicos, alunos, onde tenho ensinado, aprendido e convivido agradavelmente nos últimos sete anos.

RESUMO

A solubilidade de uma substância, potencialmente fármaco, em água é extremamente importante para o trabalho de desenvolvimento dessa como medicamento. Neste trabalho foram calculadas propriedades topológicas, geométricas e eletrônicas *in silico* de 322 moléculas, estruturalmente diferentes, a partir de parâmetros de suas estruturas espaciais moleculares (3D). As propriedades calculadas foram: área, volume, potencial de ionização, “HOMO”, “LUMO”, momento dipolar, polarizabilidade (alfa), usando o modelo semi-empírico RM1, e número de doadores e aceptores de pontes de hidrogênio por molécula, obtidos da literatura. Os dados do grupo de 322 moléculas foram subdivididos em dois sub-grupos, um com 242 para o desenvolvimento de um modelo preditivo de solubilidade, e outro com 80 moléculas para validação do mesmo. Os dados das propriedades foram correlacionadas com a solubilidade experimental destas mesmas moléculas e a eles aplicado regressão linear multivariada em relação a mesma solubilidade, chegando-se a uma equação pela qual se possa calcular uma previsão de solubilidade a partir dos dados das propriedades *in silico* apenas. Esta equação geral, abrangendo todos os dados do grupo modelo, resultou em $(r)=0,716$ e $(R^2)=51,3\%$, com $S=1,19463$, referente a $\log S$, valores superiores aos encontrados na literatura para modelos de equação única. O $\log S$ calculado para as 80 moléculas de validação demonstrou o potencial preditivo do modelo, resultando em uma correlação de $(r)=0,751$ e erro relativo médio de 4,5% em relação ao $\log S$ experimental.

Palavras-chave: solubilidade, previsão solubilidade em água, QSPR, propriedades moleculares *in silico*.

ABSTRACT

The solubility of a substance, with potential to become a commercial drug, in water is of extreme importance to the development of the drug itself. In this study the *in silico* topological, geometric and electronic properties of 322 molecules, structurally different from each other, were calculated using their 3D spatial structure. Among the properties calculated we considered: molecule area, volume, ionization potential, "HOMO", "LUMO", dipolar moment and polarizability (α), using a semi-empirical model (RM1), and also the number of hydrogen bond donors and acceptors, obtained from the literature. The data was divided into 2 groups, one with 242 molecules, for the development of a predictive model for molecular solubility, and another, with 80 molecules for the model validation. The model data was tested for correlation with the logarithm of the solubility, and a multi-linear regression model was developed, so to achieve an equation capable of predicting the solubility with only *in silico* properties. This equation resulted in the parameters $(r)=0,716$ and $(R^2)=51,3\%$, with $S=1,19463$, referring to $\log S$, relatively larger than those found in literature. The calculated $\log S$ for the 80 validation molecules showed the predictive potential of the model, resulting in a correlation of $(r)=0,751$, and relative mean error of 4,5%, when compared to the experimental $\log S$.

Keywords: solubility, water solubility prediction, QSPR, molecular properties *in silico*.

SUMÁRIO

Dedicatória.....	iv
AGRADECIMENTOS.....	v
RESUMO	vi
ABSTRACT	vii
SUMÁRIO.....	viii
LISTA DE ABREVIATÓES E ACRÔNIMOS.....	x
LISTA DE FIGURAS	xi
LISTA DE TABELAS	xii
Capítulo 1.....	1
1) Introdução.....	1
Capítulo 2.....	4
2) Materiais e Métodos.....	4
2.1) <i>Seleção de substâncias orgânicas, e dados de solubilidade</i>	4
2.2) <i>Estruturas moleculares – fontes e definições</i>	5
2.3) <i>Modelagem molecular</i>	5
2.4) <i>Complemento à modelagem molecular – pontes de hidrogênio</i>	6
2.5) <i>Organização e Análise de dados</i>	6
2.6) <i>Desenvolvimento do Modelo</i>	7
Capítulo 3.....	9
3) Resultados e discussão	9
3.1) <i>Dados</i>	9
3.2) <i>Análise dos dados obtidos das propriedades x logS</i>	20
Capítulo 4.....	34
4) Equação geral e validação:.....	34

4.1 – <i>Equação geral</i>	34
4.2 – <i>Validação</i>	36
Capítulo 5.....	42
<i>Considerações finais</i>	42
5) Referências bibliográficas	43

LISTA DE ABREVIações E ACRÔNIMOS

pH – potencial Hidrogeniônico;

QSPR – “Quantitative Structure Property Relationship”;

QSAR – “Quantitative Structure Activity Relationship”;

HOMO – “High Orbital Molecular Occupied”;

LUMO – “Lowest Unoccupied Molecular Orbital”;

EXCEL – Programa Computacional - Planilha eletrônica da Microsoft 2007;

LISTA DE FIGURAS

Figura 1- Fluxograma de delineamento do trabalho.....	8
Figura 2 – Demonstração do relacionamento entre logS x PM onde a “linha central” representa a regressão linear ponderada dos pontos.....	21
Figura 3 – Demonstração do relacionamento entre logS x $1/PM*1000$ onde a “linha central” representa a regressão linear ponderada dos pontos.	22
Figura 4 – Demonstração do relacionamento de logS x AREA onde a “linha central” representa a regressão linear ponderada dos pontos.....	23
Figura 5- Demonstração do relacionamento de logS x $1/AREA*1000$ onde a “linha central”, representa a regressão ponderada dos pontos.	23
Figura 6- Demonstração do relacionamento de logS x VOL onde a “linha central” representa a regressão linear ponderada dos pontos.....	24
Figura 7 – Demonstração do relacionamento entre logS x $1/VOL*1000$ onde a “linha central” representa a regressão linear ponderada dos pontos.	25
Figura 8 – Demonstração do relacionamento entre logS x PI onde a “linha central” representa a regressão linear ponderada dos pontos.....	26
Figura 9- Demonstração do relacionamento entre logS x $\log PI*10$ com “linha central” representando a regressão linear ponderada.....	26
Figura 10 – Demonstração do relacionamento logS x HOMO onde a “linha central” representa a regressão linear ponderada dos pontos.....	27
Figura 11 – Demonstração do relacionamento entre logS x $\log HOMO^{**2}$ onde a “linha central” representa a regressão linear ponderada dos pontos.	28
Figura 12 – Demonstração do relacionamento entre logS x Polarizabilidade (alfa) onde a “linha central” representa a regressão linear ponderada dos pontos.	28
Figura 13 – Demonstração do relacionamento entre logS x $1/Alfa*100$ onde a “linha central” representa a regressão linear dos pontos.....	29
Figura 14 – Representação do relacionamento entre logS x MD onde a “linha central” representa a regressão linear ponderada dos pontos.....	30
Figura 15 – Representação do relacionamento entre logS x Hdr onde a “linha central” representa uma regressão linear ponderada dos pontos.	31
Figura 16 – Representação do relacionamento entre logS x HAc onde a “linha central” representa a regressão linear ponderada dos pontos.	32
Figura 17 – O gráfico apresenta a relação logS x logScalc onde a “linha central” representa a regressão linear ponderada dos dados.....	35

LISTA DE TABELAS

Tabela 1- Conjunto de dados usados para geração do modelo de solubilidade das substâncias orgânicas em água.....	10
Tabela 2 – Tabela de dados inerentes às substâncias selecionadas para validação do modelo desenvolvido.	36
Tabela 3: Teste de hipótese, t-student pareado, para validação do modelo. Mostra não haver indícios de diferença estatisticamente significativa entre o logS calculado e o logS experimental.....	41

Capítulo 1

1) Introdução

A solubilidade de uma substância em dado solvente é a quantidade que este é capaz de dissolvê-la, a determinadas condições de temperatura, pressão e pH, tomando-se como limite o equilíbrio na saturação [1]. Este conceito é aplicado a todos estados da matéria, mas no nosso caso foram trabalhados somente os estados sólido (solutos ou substâncias) e líquido (solvente ou água).

O ciclo de vida biológica humana de um fármaco, da sua absorção à eliminação, envolve sistemas aquosos, e não há mecanismo de absorção, distribuição, ação, metabolismo ou eliminação que não seja em meio aquoso e que não ocorra a nível molecular [1]. Uma forma de disponibilizar um fármaco para sua absorção, no caso de sólidos, é dissolvê-lo, independente da forma farmacêutica e via de administração. Caso não ocorra solubilização do fármaco em uma forma farmacêutica é necessário prévia dissolução no organismo para que ele seja absorvido. Isso ressalta a grande importância do conhecimento da solubilidade de um fármaco em água.

Conforme Yalkowsky e Valvani [2], a solubilidade aquosa de uma substância orgânica é governada por três fatores principais: a) a entropia de mistura; b) a diferença entre as interações adesivas fármaco-água (DW) e a soma das interações coesivas fármaco-fármaco (DD) e água-água (WW); e c) as interações adicionais associadas à energia de cristal dos fármacos cristalinos, que é designada por DD e não é aplicável a líquidos. Baseado nestes princípios eles desenvolveram equações que associadas a dados experimentais de substâncias já existentes (entropia de fusão, coeficientes de partição e solubilidade aquosa das substâncias classificadas por grupos funcionais) e com tratamento estatístico final dos dados tornou-se possível a previsão de solubilidade de substâncias orgânicas em água [2].

A solubilidade de dada substância, não iônica, em dado solvente, é melhor quando há semelhança entre as forças de interação DW^1 e DD.

¹ DW se solvente água.

A solubilidade de compostos orgânicos é normalmente representada por logS, onde S é a concentração do composto em mol/L. Na prática, cerca de 85% dos fármacos tem valores de logS entre -1 e -5 e virtualmente não tem valores abaixo de -6, [6] e [14]. Valores acima de -1 estão associados com moléculas bastante polares tais como açúcares ou pequenos peptídeos, têm baixa permeabilidade de membrana na ausência de transporte ativo mas não são problemáticos. Empiricamente, é aparente que o alvo de logS entre -1 a -5 para a maioria dos fármacos reflita a relação entre a polaridade necessária para razoável solubilidade aquosa e a hidrofobicidade necessária para a passagem aceitável de membrana [6].

São encontrados na literatura vários métodos desenvolvidos para previsão de solubilidade teórica a partir de estrutura molecular e propriedades físico-químicas das moléculas entre os quais, o trabalho de Yalkowsky e Valvani, citado acima, contribuição de grupo funcional [3], [4] e [5] e “QSPR” (“Quantitative Structure Property Relationship”) [11], [12], [13], [14], [15] e [16].

O atual trabalho não aprofunda a revisão dos dois primeiros métodos e sim toma a cronologia dos fatos, observa principalmente os últimos trabalhos de revisão do assunto [6] e [7] e conclui que o último, “QSPR”, está se evidenciando como o melhor e mais prático método para avaliar a solubilidade em questão, embora entenda que o trabalho de Yalkowsky e colaboradores [2] use dados mais intrinsecamente envolvidos com o conceito de solubilização, e que contribuição de grupo funcional pareça dizer mais detalhes da molécula. O trabalho, portanto, pretende contribuir com mais um método onde se cria a possibilidade de previsão de solubilidade de dado fármaco orgânico e não salino, em água, a partir de propriedades *in silico*², obtidas diretamente de parâmetros topológicos, geométricos e eletrônicos de sua estrutura espacial molecular (3D).

A idéia da relação quantitativa estrutura-propriedade (ou estrutura atividade), respectivamente “QSPR/QSAR”, foi introduzida por Hans e colaboradores em 1963 e primeiro aplicado à análise da importância da lipofilicidade para potência biológica [8]-[10].

² Em computador

Interpretando o trabalho de Yalkowsky e colaboradores [2], é evidente a dependência de solubilidade das forças intermoleculares. A ruptura das ligações intermoleculares de um soluto para interagir com um solvente depende dessa possibilidade quantitativa de interação. Essa compreensão é dinâmica porém a forma de trabalho usada em “QSPR” relaciona propriedades “estáticas” que se relacionam com solubilidade. Foram escolhidas propriedades como: área, volume, “HOMO”, “LUMO”, momento bipolar, polarizabilidade (alfa) e potencial de ionização inerente a cada estrutura molecular de cada substância.

Na revisão bibliográfica deste trabalho não foram localizados outros que destacassem conceitualmente justificativas específicas da relação propriedades/solubilidade, então premissas tradicionais do que é estudado em relação a solubilidade, de que necessariamente um soluto ocupa um espaço no solvente (área e volume), os orbitais externos das moléculas têm influência entre as ligações soluto-solvente (“HOMO” e “LUMO”), que medidas de momento dipolar e polarizabilidade são inerentes a resultante das “cargas” das moléculas e potencial de ionização, embora estejamos tratando de moléculas virtualmente não ionizáveis, podem possuir um mínimo de ionização, foram decisivas para suas escolhas.

Um caminho de interpretar o potencial de inter-relações entre moléculas quanto às pontes de hidrogênio foi pesquisado.

Os dados estatisticamente trabalhados é que mostram a relação matemática das propriedades em estudo com as solubilidades conhecidas das substâncias.

Foram usados para obtenção das propriedades dois “softwares” reconhecidos na literatura científica, HyperChem 8.0 e MOPAC 2009. Nos dois foi usado o RM1, como método semi-empírico, para definição dos parâmetros geométricos e eletrônicos do HyperChem 8.0 e para cálculos das propriedades desenvolvidos pelo MOPAC. Nelson H. Morgon e Kaline Coutinho citam que a maioria dos grandes avanços obtidos de métodos semi-empíricos foi incorporada nessa série de programas intitulados MOPAC [25] sendo que a revisão de programa utilizada foi a 14^a. O método RM1, literalmente Recife Método 1, desenvolvido pelo Grupo de Química Teórica do Departamento de Química Fundamental, da Universidade Federal de Pernambuco, na Cidade de Recife, Brasil, [25] foi incorporado no MOPAC 2007 e ao HyperChem 8.0, permanecendo até hoje.

Capítulo 2

2) Materiais e Métodos

Definido o foco do trabalho, foram escolhidas as substâncias como compostos orgânicos e suas solubilidades e os programas de computação para serem utilizados. A sequência de trabalho foi desenvolvida: primeiro o desenho das estruturas moleculares, segundo os cálculos das propriedades das moléculas a partir das estruturas moleculares (o que corresponde à modelagem *in silico*) e terceiro as interpretações e cálculos estatísticos a partir dos dados das propriedades e sua relação com a solubilidade em água.

Em termos de materiais em si, foi intencionalmente previsto o uso de computador de pequeno porte tipo “notebook” com bom desempenho, também para que quem necessite fazer cálculos futuros tenha facilidade para tal. Todo trabalho foi desenvolvido em Notebook Dell Inspiron 1545-501, Processador Intel® Core 2 Duo T6400, 4GB de memória “RAM” e Windows Vista® Home Premium, como sistema operacional.

2.1) Seleção de substâncias orgânicas, e dados de solubilidade

Com base em diversos trabalhos verificados na literatura foi escolhido um conjunto de 322 moléculas [11] sendo nossa atenção voltada para a amplitude dos grupos de substâncias, quanto a solubilidade, quanto à diversidade de ação farmacológica e quanto a segurança apresentada nas fontes de dados de solubilidades anotadas – “Analytical Profile of Drug Substances”, “the Merck Index” e mais 17 trabalhos científicos devidamente referenciados.

As 322 moléculas com os seus dados foram subdivididas em 242 para o desenvolvimento do modelo (Tabela 1) e 80 (Tabela 2), aleatoriamente escolhidas, para validação do modelo.

O conceito de solubilidade é simples, porém a sua execução prática não é facilmente desenvolvida com precisão. Um dos métodos descrito na literatura traz que, definidas as condições de temperatura, pressão e pH deve ser adicionado o soluto no solvente sob agitação até saturação e determinado a quantidade de soluto no filtrado por uma método analítico confiável. Repetida a agitação por mais tempo o

fim da definição da solubilidade se dá quando dois resultados sucessivos são iguais [21]. Questões envolvendo estabilidade da substância em solução aquosa, recristalização em caso de resfriamento, método de análise e fontes de matérias primas são pontos que podem trazer variações de resultados. Há discussões na literatura somente a esse respeito no sentido de entender a propagação de erros estatísticos oriundos destas variações, [20] e [21].

Não foi localizado na literatura o desenvolvimento em um único laboratório de testes de solubilidade de uma quantidade grande de substâncias de modo que também outros autores têm buscado na literatura científica dados de solubilidade confiáveis para desenvolver seus trabalhos [14] e [15].

A Tabela I, à frente, mostra os compostos e dados de solubilidade obtidos e mais.

2.2) Estruturas moleculares – fontes e definições

O programa escolhido para definição espacial das estruturas moleculares foi o HyperChem 8.0. Todas as moléculas foram cuidadosamente desenhadas e conferidas a partir do trabalho de Rytting e colaboradores [11] como fonte primária e confirmação de estrutura planar pela base de dados de estruturas ChemIDplus da “United States National Library of Medicine” [22]. A otimização geométrica foi feita usando sempre os mesmos parâmetros no HyperChem, ou seja, RM1 como método semi-empírico e Polak-Ribiere, em vácuo, como algoritmo de otimização.

Foi desenvolvido um trabalho minucioso e longo onde a premissa foi a inexistência de erros devido a possibilidade de propagação para os dados finais.

A não exposição das estruturas moleculares planares se deu porque elas só teriam função visual de ilustração uma vez que os dados usados para os cálculos das propriedades foram dos parâmetros das estruturas tridimensionais (3D), além de tornarem o trabalho bastante extenso.

2.3) Modelagem molecular

O MOPAC 2009, tendo por traz o já citado HyperChem, foi escolhido como programa de modelagem molecular também por sua leveza, precisão e fácil

comunicação com os dados dos parâmetros moleculares espaciais gerados pelo HyperChem, para o cálculo de propriedades *in silico*.

Foram usadas as palavras chaves RM1, PRECISE, POLAR, VECTORS e XYZ no MOPAC como forma de comandos para o cálculo das propriedades. A todas moléculas foram aplicadas as mesmas palavras chaves.

As propriedades de interesse calculadas, ou seja, Área, Volume, Potencial de Ionização, HOMO, LUMO, Momento dipolar e Polarizabilidade (alfa) foram coletadas pelo MolecOrg³ e transferidas para o Excel.

Todos os dados das propriedades obtidas estão na Tabela 1.

2.4) Complemento à modelagem molecular – pontes de hidrogênio

A obtenção dos dados quanto ao número de Hidrogênios doadores de pontes de hidrogênio e átomos eletronegativos (O, N e F) receptores de pontes de hidrogênio nas moléculas foi a partir do PubMed , mesmo site das estruturas moleculares [21] e mesma base de dados, ChemIDplus. Vale esclarecer que estes dados não se originam de contagem visual dos aceptores ou doadores mas a partir de programa tendo um método semi-empírico como base para checar os orbitais passíveis de contribuir como tais e numéricamente estão, de modo oficial, publicados no site em questão com devida referência [24].

2.5) Organização e Análise de dados

Todos os dados foram organizados em planilha EXCEL– Tabela I e posteriormente transferidos para o Minitab®.

A análise dos dados foi feita integralmente pelo Minitab 15, acadêmico, programa estatístico com redução do seu porte de base de dados mas com manutenção de todos os seus recursos de cálculos profissionais suficientes para as análise. As duas versões, profissional e acadêmico, são de grande uso em universidades e empresas. Este programa se inclui no propósito do trabalho que é fazê-lo com “facilidade” computacional – tudo desenvolvido em Notebook, sem

³ Programa em FORTRAN por Pedro da Silva Peixoto para coleta de dados do MOPAC .

nenhuma perda de qualidade- bem como, do mesmo modo, para futuros cálculos para moléculas novas.

No Minitab também foram trabalhadas as variáveis (propriedades) em função da equação final do trabalho.

2.6) Desenvolvimento do Modelo

Os dados obtidos da literatura inerente às moléculas, como peso molecular e solubilidade, foram transferidos para planilha EXCEL e encima dessa planilha foram transferidos todos os dados das propriedades obtidas pelo MOPAC tais como, área, volume, potencial de ionização, HOMO, LUMO, momento dipolar e polarizabilidade (alfa). Os dados do número de H (hidrogênios) como doadores de pontes de hidrogênio e número de átomos eletronegativos, como aceptores de pontes de hidrogênio foram inseridos, da fonte, diretamente para a planilha EXCEL.

As análises estatísticas dos dados foram desenvolvidas no Minitab. Alguns princípios foram seguidos para o desenvolvimento do modelo:

- a) escalonamento e linearização dos dados – todas as variáveis foram transformadas em números positivos entre 0 e 50 com o fim de reduzir erros nas contribuições de cada conjunto de variáveis para os cálculos, sendo checado passo a passo a redução de erro via coeficiente de correlação de Pearson entre a variável “alterada” e logS. Todas as variáveis “alteradas”, tiveram melhora na correlação (r);
- b) foram plotados gráficos de cada propriedade *versus* logS para visualização da dispersão dos dados e da linha que representa a regressão ponderada (Figuras 2 a 16);
- c) assim trabalhados os dados, foram calculadas as correlações (r) de todas as variáveis entre si e analisadas uma a uma para melhor compreensão do inter-relacionamento dos dados.
- d) a partir dos gráficos foram pensadas novas curvas de regressão no sentido de linearização dos dados, gerando uma transformação ao que pudesse melhor exprimir a regressão linear de dados para obtenção de uma melhor equação final;

- e) foram feitas as regressões [33] e analisadas as equações resultantes⁴, verificando-se os coeficientes de determinação (R^2), que representam o índice de acerto (R^2) da equação proposta, e os p-valores correspondentes para cada variável, indicando a viabilidade de se manter ou retirar a variável da equação, ambos parâmetros calculados pelo Minitab®.
- f) foram tiradas conclusões quanto a eficiência da equação do item e) (acima) com o cálculo do logS calculado e visto seu (R^2) e (r) .
- g) foi desenvolvida a validação da equação com novos dados.

Sintetizando:

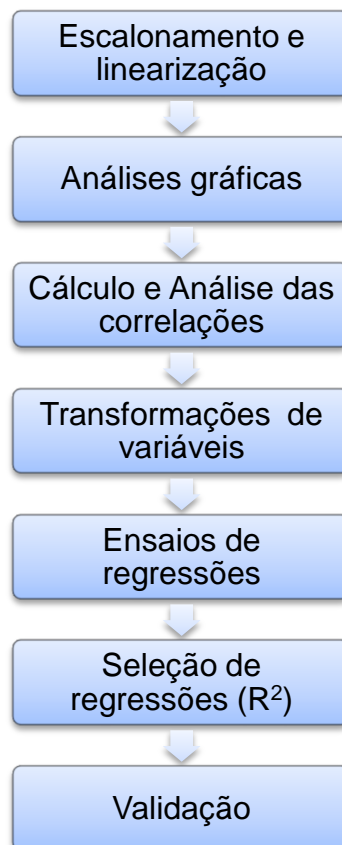


Figura 1- Fluxograma de delineamento do trabalho.

⁴ Trata-se de regressões múltiplas, pelo método dos mínimos quadrados, onde a variável independente (logS) é igualada à combinação dos dados das propriedades (variáveis dependentes) . Cálculos desenvolvidos no Minitab.

Capítulo 3

3) Resultados e discussão

3.1) *Dados*

Abaixo, Tabela 1 onde foi inserido o conjunto de dados da solubilidade experimental como logS [11], o logS calculado a partir da equação conclusiva do trabalho, o residual inerente à diferença ($\log S - \text{LogScalc}$), pesos moleculares [11], as propriedades das moléculas calculadas pelo MOPAC (área, volume, potencial de ionização, “HOMO”, “LUMO”, momento dipolar e polarizabilidade (alfa)) e os números correspondentes a quantidade de Hidrogênios doadores de pontes de hidrogênios e átomos aceptores de pontes de hidrogênios. Para facilitar a compreensão em relação à solubilidade, o logS foi colocado em ordem crescente da menor solubilidade para a maior.

Não foi prioridade do trabalho fazer considerações sobre grupos funcionais e suas influências sobre as solubilidade. As propriedades calculadas, todas referem-se as moléculas como um todo e não partes, no entanto é evidente na tabela que compostos com características aromáticas, predominantemente, como Perilene, Naftaceno, Criseno, etc possuem baixíssima solubilidade o que por si só sairiam fora da faixa usual para fármacos. A grande maioria das moléculas da Tabela 1 possuem logS entre -6 e -1 como previsto na literatura [6] e citado na introdução como a faixa de solubilidade em que se encaixam a grande maioria dos fármacos. Esta é mais uma justificativa da amplitude das moléculas aqui trabalhadas.

Dados como quantidade de doadores e aceptores de pontes de hidrogênios parecem estranhos uma vez que são quantidades expressas em números inteiros, sem unidades específicas. Não podemos abstrair a impossibilidade de se relacionarem com polaridade ou polarizabilidade. Certamente definir via modelagem molecular a energia previsível desses orbitais seria mais compatível com as outras medidas e possivelmente dariam mais acerto aos cálculos. A literatura pesquisada só mostra esta forma de quantificar potenciais pontes de hidrogênios [24].

O caminho QSPR coloca os dados, uma vez gerados, isoladamente, e os trabalha estatisticamente como puros e vai buscar todas as possíveis correlações com o dado experimental logS que representa mais a natureza.

Tabela 1- Conjunto de dados usados para geração do modelo de solubilidade das substâncias orgânicas em água.

Nome Substância	logS	logScalc	Residual	PM	AREA	VOL	PI	HOMO	LUMO	MD	Alfa	Hdr	HAc
Perilene	-8,800	-6,2022	-2,5978	252,31	260,46	289,25	7,7811	-7,781	-0,869	0,001	29,4266	0	0
Naftaceno	-8,690	-5,9044	-2,7856	228,29	348,95	400,20	7,7307	-7,731	-0,980	0,041	44,6975	0	0
Criseno	-8,060	-5,6561	-2,4039	228,29	250,09	268,77	8,2695	-8,269	-0,395	0,001	25,5323	0	0
3_4-Benzopireno	-7,820	-6,2244	-1,5956	252,31	263,39	291,54	7,8384	-7,838	-0,825	0,037	30,3356	0	0
Hexaclorobenzeno	-7,760	-5,2066	-2,5534	284,78	215,61	232,60	9,9127	-9,913	-1,189	0,000	14,2747	0	0
Benzantraceno	-7,210	-5,7752	-1,4348	228,30	253,09	272,12	8,1036	-8,104	-0,552	0,032	26,3855	0	0
Etomidato	-6,735	-2,8176	-3,9174	244,29	271,88	300,66	9,5145	-9,515	0,066	3,054	19,6109	0	3
Antraceno	-6,380	-5,0053	-1,3747	178,23	210,62	217,85	8,0366	-8,037	-0,578	0,000	20,0637	0	0
Pireno	-6,180	-5,3607	-0,8193	202,26	220,40	237,79	8,0247	-8,025	-0,627	0,000	22,5581	0	0
Lovastatina	-6,000	-2,7580	-3,2420	404,55	409,51	514,15	8,7836	-8,784	0,563	5,827	30,9467	1	5
Tioridazina	-5,824	-4,7562	-1,0678	370,57	372,51	455,53	7,8927	-7,893	0,151	2,125	33,1457	0	2
Clofazimina	-5,800	-5,2123	-0,5877	473,40	452,41	550,57	7,8227	-7,823	-0,733	1,496	50,5356	1	4
Ádico Fólico	-5,441	-1,8469	-3,5941	441,40	446,51	501,78	8,6519	-8,652	-1,060	9,968	35,5315	6	11
Fenbufeno	-5,301	-4,0221	-1,2789	254,28	299,60	321,85	9,1121	-9,112	-0,883	3,511	23,9698	1	3
Equilin	-5,282	-3,6512	-1,6308	268,35	278,76	326,51	8,7250	-8,725	0,495	3,128	22,1025	1	2
Clorpirifos	-5,244	-3,5882	-1,6558	350,58	309,18	349,96	9,7813	-9,781	-1,422	4,762	21,5127	0	4
Fludioxonil	-5,210	-1,9465	-3,2635	248,19	241,65	260,04	9,3932	-9,393	-0,150	7,260	17,5897	1	5
Dantron	-5,187	-4,2341	-0,9529	240,21	232,32	248,99	9,3153	-9,315	-1,458	0,815	20,9102	2	4
Fenantreno	-5,150	-4,5300	-0,6200	178,23	207,57	217,64	8,4949	-8,495	-0,155	0,021	18,7887	0	0
Diclofenaco	-5,097	-4,4583	-0,6387	296,15	256,60	276,05	8,0711	-8,071	0,240	1,218	20,6754	2	3
Fenpiclonil	-5,074	-4,2285	-0,8455	237,09	226,74	241,31	8,9034	-8,903	-0,373	7,702	17,9130	1	1
Sulindac	-5,000	-4,1032	-0,8968	356,41	348,55	407,57	8,9867	-8,987	-0,806	4,799	31,4255	1	4
Estradiol	-4,845	-3,7914	-1,0536	272,39	286,32	338,57	8,5817	-8,582	0,672	2,612	21,9317	2	2
Indometacina	-4,775	-4,2614	-0,5136	357,79	345,63	399,73	8,5054	-8,505	-0,659	2,632	29,9755	1	4
Stanolone	-4,743	-2,8584	-1,8846	290,44	299,06	373,84	10,1406	-10,141	1,202	3,983	21,4302	1	2
Fenoxicarb	-4,719	-3,8438	-0,8752	301,34	345,49	366,10	8,4584	-8,458	-0,897	5,638	25,6209	1	4

Terfenadina	-4,674	-4,8313	0,1573	471,68	510,62	613,47	8,9422	-8,942	0,479	3,518	40,5928	2	3
Noretisterona	-4,630	-3,9926	-0,6374	298,42	264,40	309,22	9,8360	-9,836	0,355	4,325	23,4066	1	2
Ácido Flufenâmico	-4,623	-2,9386	-1,6844	281,23	273,96	292,93	8,4523	-8,452	-0,624	4,939	22,7484	2	6
G-BHC-(Lindano)	-4,600	-3,6882	-0,9118	290,83	212,20	252,35	10,2413	-10,241	-1,271	3,889	12,2776	0	0
Acenafteno	-4,590	-3,5803	-1,0097	154,21	187,25	192,55	8,3673	-8,367	0,048	0,613	15,1211	0	0
Ácido Iopanóico	-4,580	-5,2201	0,6401	570,93	287,22	345,16	8,5054	-8,505	-0,125	0,884	22,8769	2	3
Glafenina	-4,571	-3,7859	-0,7851	372,81	370,19	416,03	8,7797	-8,780	-0,747	3,738	31,0927	3	6
Diflunisal	-4,479	-3,3703	-1,1087	250,20	244,64	257,36	8,7868	-8,787	-0,625	1,916	20,0184	2	5
5-Etil-5-nonilbarbiturato	-4,462	-3,4976	-0,9644	282,38	335,82	369,23	10,9808	-10,981	-0,078	1,591	20,4691	2	3
Amitriptilina	-4,456	-4,4748	0,0188	277,41	316,63	373,03	8,5677	-8,568	0,170	1,395	26,1087	0	1
1_3_5-Triclorobenzeno	-4,440	-3,5192	-0,9208	181,45	176,80	175,27	9,9826	-9,983	-0,517	0,000	10,5465	0	0
Haloperidol	-4,429	-3,9655	-0,4635	375,87	391,55	443,06	9,2754	-9,275	-0,425	3,785	28,9702	1	4
Progesterona	-4,420	-3,4543	-0,9657	314,47	321,18	405,63	9,9115	-9,912	0,344	3,050	24,2785	0	2
Iprodiona	-4,405	-4,0404	-0,3646	330,17	325,73	352,39	9,6647	-9,665	-0,449	6,145	23,6847	1	3
Benperidol	-4,280	-4,0930	-0,1870	381,45	388,29	449,82	8,4137	-8,414	-0,371	4,385	31,7844	1	4
Prometazina	-4,260	-4,2008	-0,0592	284,42	299,20	351,63	7,5633	-7,563	0,263	1,801	26,4882	0	2
Fenitoina	-4,226	-3,9805	-0,2455	252,27	259,16	292,04	9,5961	-9,596	0,070	3,265	20,9609	2	2
5_5-Difenilbarbiturato	-4,196	-3,9327	-0,2633	280,28	271,15	313,06	9,6319	-9,632	-0,109	1,954	22,8723	2	3
Spirolactona	-4,173	-3,1818	-0,9912	416,57	378,97	499,34	9,9636	-9,964	-0,030	5,076	31,2199	0	4
Simvastatina	-4,145	-2,7413	-1,4037	418,57	374,84	515,26	8,7682	-8,768	0,549	3,300	30,2753	1	5
Triazolam	-4,090	-4,1222	0,0322	343,21	315,66	366,81	9,7768	-9,777	-0,961	5,962	27,5778	0	3
1_4-Dibromobenzeno	-4,070	-4,7064	0,6364	235,91	168,80	167,01	9,5693	-9,569	-0,107	0,000	10,8928	0	0
Prasterona	-4,064	-3,1515	-0,9125	288,43	300,12	371,51	9,4094	-9,409	1,176	2,178	21,7583	1	2
Metiltestosterona	-3,990	-3,4122	-0,5778	302,46	307,00	387,87	9,8433	-9,843	0,407	4,314	23,3081	1	2
Estrona	-3,955	-3,5653	-0,3897	270,37	283,97	333,32	8,7304	-8,730	0,532	3,407	21,8217	1	2
Trianterene	-3,946	-2,9836	-0,9624	253,27	264,62	278,23	8,2998	-8,300	-0,490	2,651	26,2865	3	7
Fenclofos	-3,905	-3,8853	-0,0197	321,55	276,80	307,14	9,6613	-9,661	-1,651	6,105	19,0625	0	3
Fenclofenaco	-3,854	-3,9640	0,1100	297,14	252,35	270,04	8,9328	-8,933	0,224	2,333	19,3651	1	3

Pentazocin	-3,803	-3,9351	0,1321	285,43	311,17	373,89	8,5164	-8,516	0,636	0,806	24,6402	1	2
Indapamida	-3,792	-3,8116	0,0196	365,84	343,18	395,42	8,9735	-8,973	-1,190	5,547	27,3138	2	5
Metilclotiazida	-3,778	-1,5857	-2,1923	360,23	283,21	332,42	10,3334	-10,333	-1,181	11,903	19,9801	2	7
Betametasona	-3,770	-2,3502	-1,4198	392,47	340,21	453,85	9,9263	-9,926	0,754	2,036	26,1619	3	6
1_2_3-Triclorobenzeno	-3,760	-3,1397	-0,6203	181,45	170,43	172,24	9,7566	-9,757	-0,417	2,627	10,3501	0	0
Diuron	-3,760	-4,1233	0,3633	233,10	260,09	273,88	9,4030	-9,403	-0,734	0,848	15,4454	1	1
Diazepam	-3,754	-4,3527	0,5987	284,74	284,40	322,79	9,1730	-9,173	-0,545	3,074	24,5641	0	2
Flurbiprofeno	-3,740	-3,4788	-0,2612	244,26	266,29	286,56	9,0135	-9,014	-0,085	2,683	20,3018	1	3
Talidomida	-3,699	-3,1819	-0,5171	258,23	261,82	293,09	9,0430	-9,043	-1,038	5,736	21,1942	1	4
Triancinolona	-3,693	-2,4039	-1,2891	394,44	337,75	442,51	10,0360	-10,036	-0,179	4,100	26,7627	4	7
Naftaleno	-3,610	-2,8621	-0,7479	128,17	164,35	163,06	8,5953	-8,595	-0,022	0,000	12,7653	0	0
Lorazepam	-3,604	-4,4590	0,8550	321,16	287,32	333,59	9,0736	-9,074	-0,533	2,993	24,5429	2	3
Dexametasona	-3,590	-2,2671	-1,3229	392,47	348,78	452,83	10,1751	-10,175	0,765	3,547	25,9105	3	6
Guanina	-3,577	-2,1828	-1,3942	151,13	164,73	158,76	8,6941	-8,694	-0,016	2,563	11,9022	3	4
Bumetanida	-3,562	-2,5213	-1,0407	364,42	348,36	412,33	9,4296	-9,430	-1,025	8,407	27,3139	3	7
5-t-Butil-5-(3-metilbut-2-enil)barbiturato	-3,551	-2,9139	-0,6371	252,31	264,34	313,79	10,1638	-10,164	-0,011	1,850	18,5668	2	3
Sulfadiazina	-3,529	-1,9405	-1,5885	250,27	253,11	264,50	9,3981	-9,398	-0,326	6,500	19,2538	1	6
Melfalan	-3,485	-3,0996	-0,3854	305,20	319,46	363,86	9,6069	-9,607	0,039	3,499	21,1329	2	4
Nitrofurantoina	-3,477	-2,7352	-0,7418	238,16	244,46	242,32	9,5384	-9,538	-1,292	6,861	20,2416	1	6
Isoproturona	-3,469	-3,5434	0,0744	206,29	259,05	271,55	8,6334	-8,633	0,584	3,230	17,9511	1	1
Clortalidona	-3,451	-3,6114	0,1604	338,76	304,01	341,02	9,3913	-9,391	-0,926	7,766	24,4203	3	5
Fludrocortisona	-3,434	-2,6315	-0,8025	380,46	340,18	443,21	10,0697	-10,070	0,163	1,614	25,9173	3	6
Ibuprofeno	-3,420	-2,6625	-0,7575	206,28	253,08	275,63	9,2408	-9,241	0,436	1,806	16,3101	1	2
Ácido úrico	-3,402	-3,3297	-0,0723	168,11	169,64	165,06	8,9767	-8,977	-0,401	2,725	11,9895	4	3
Zileuton	-3,373	-4,0631	0,6901	236,29	246,58	268,58	8,6040	-8,604	-0,198	4,075	19,4911	2	2
Ácido nalidixico	-3,366	-2,4480	-0,9180	232,24	230,72	243,27	9,1224	-9,122	-0,594	4,987	17,9388	1	5
Carbamazepina	-3,294	-4,1285	0,8345	236,27	251,94	283,51	8,7235	-8,723	0,197	3,991	21,1289	1	1
Griseofulvina	-3,246	-2,6580	-0,5880	352,77	324,14	379,41	9,2623	-9,262	-0,578	3,080	26,7224	0	6

Corticosterona	-3,240	-2,9278	-0,3122	346,47	335,63	425,44	9,8711	-9,871	0,400	5,383	25,3808	2	4
5-Etil-5-heptilbarbiturato	-3,218	-3,1008	-0,1172	254,33	297,84	325,91	11,0918	-11,092	-0,078	1,590	18,1895	2	3
Metoclopramida	-3,176	-3,4448	0,2688	299,80	332,49	366,98	8,6102	-8,610	0,220	5,967	23,4898	3	4
Cetoprofeno	-3,155	-3,5456	0,3906	254,28	281,85	304,23	9,6906	-9,691	-0,158	1,851	20,7146	1	3
Butambeno	-3,131	-2,5513	-0,5797	193,25	244,57	245,80	8,5350	-8,535	0,215	4,008	16,4274	1	3
Alclofenaco	-3,125	-3,1390	0,0140	226,66	232,91	237,77	9,5307	-9,531	-0,501	4,641	15,7079	1	3
Butilparabeno	-3,101	-2,7064	-0,3946	194,23	240,50	241,66	9,3567	-9,357	-0,154	1,297	15,4568	1	3
Hydrocortisona	-3,100	-2,6612	-0,4388	362,47	339,48	445,83	9,8658	-9,866	0,403	5,422	26,4535	3	5
Dapsona	-3,094	-2,9053	-0,1887	248,30	255,23	272,60	9,0417	-9,042	0,029	7,815	20,2578	2	4
Ciclopentano-spirobarbiturato	-3,060	-2,1755	-0,8845	196,21	201,44	219,14	10,9965	-10,996	-0,088	1,712	13,6595	2	3
Hexetal	-3,049	-2,9139	-0,1351	240,30	276,91	301,13	11,1479	-11,148	-0,081	1,643	17,1283	2	3
Khellin	-3,017	-2,7720	-0,2450	260,25	263,13	284,11	8,5952	-8,595	-0,344	2,548	21,7778	0	5
Heptabarbital	-3,000	-2,9600	-0,0400	250,30	252,39	299,40	9,7860	-9,786	-0,009	1,923	18,3326	2	3
Disulfiram	-2,995	-5,5514	2,5564	296,52	300,91	356,06	8,6026	-8,603	-1,192	5,557	27,8556	0	0
Cicloheptano-spirobarbiturato	-2,982	-2,5219	-0,4601	224,26	226,38	259,41	10,8284	-10,828	-0,083	1,828	15,8438	2	3
Oxamniquine	-2,965	-3,0681	0,1031	279,34	303,35	338,69	8,5240	-8,524	-0,565	6,554	22,8571	3	5
Brinzolamida	-2,950	-2,1970	-0,7530	383,50	356,17	404,22	9,5977	-9,598	-1,706	5,919	23,2861	2	8
Metaqualona	-2,921	-4,1083	1,1873	250,30	267,67	301,46	8,8389	-8,839	-0,275	1,936	24,0587	0	2
Fenoltaleina	-2,900	-4,1362	1,2362	318,33	314,41	358,02	8,9295	-8,930	-0,418	2,104	27,1268	2	4
Praziquantel	-2,893	-3,7943	0,9013	312,41	321,88	376,30	9,4920	-9,492	0,136	5,991	25,7160	0	2
Trimetoprima	-2,861	-1,9423	-0,9187	290,32	312,56	341,62	8,5821	-8,582	0,432	4,096	24,7498	2	7
Chlorzoxazona	-2,831	-3,0485	0,2175	169,57	173,62	170,07	9,3240	-9,324	-0,396	2,487	12,0548	1	2
Diclorprop	-2,827	-1,3913	-1,4357	235,07	233,05	264,20	9,9587	-9,959	0,524	6,086	13,4865	1	3
Quinidina	-2,812	-3,3258	0,5138	324,42	333,99	403,15	8,6799	-8,680	-0,143	4,950	27,6080	1	4
Heroína	-2,798	-2,3297	-0,4683	369,42	345,19	426,56	8,7593	-8,759	0,313	1,782	28,1269	0	6
Quinina	-2,790	-3,4865	0,6965	324,42	338,85	405,64	8,5505	-8,550	-0,135	3,526	27,7996	1	4

Ácido diatrizóico	-2,788	-5,3084	2,5204	613,90	313,41	369,77	8,8170	-8,817	-0,729	3,421	26,3560	3	4
Reposal	-2,773	-3,0989	0,3259	262,31	258,67	310,76	9,7391	-9,739	-0,043	1,869	18,9317	2	3
5_5-Di-i-propilbarbiturato	-2,766	-1,9643	-0,8017	212,25	218,37	254,82	11,0658	-11,066	-0,026	1,772	14,6122	2	3
Tibamato	-2,739	-2,1392	-0,5998	274,36	336,84	362,34	9,9875	-9,988	1,173	5,032	19,1963	2	4
6-Hidroxipteridina	-2,714	-1,5536	-1,1604	148,12	161,65	156,94	9,9968	-9,997	-1,265	2,868	12,0777	1	5
Pteridina-7-tiol	-2,706	-2,4106	-0,2954	164,18	175,87	174,61	9,7200	-9,720	-1,337	3,181	14,2000	0	4
Sulfametoxazol	-2,705	-1,7806	-0,9244	253,28	255,08	270,71	9,5537	-9,554	-0,547	8,797	17,9922	2	6
Hidroclorotiazida	-2,689	-1,6293	-1,0597	297,74	243,94	267,21	10,3425	-10,343	-1,213	11,070	16,7457	3	7
Tebaina	-2,658	-3,2990	0,6410	311,38	308,63	360,28	8,4614	-8,461	0,310	1,489	26,4849	0	4
Pteridina-4-tiol	-2,646	-2,3058	-0,3402	164,18	173,58	173,67	9,5908	-9,591	-1,378	3,510	13,8876	0	4
Primidona	-2,640	-2,9156	0,2756	218,25	228,64	259,33	9,6000	-9,600	0,302	4,173	16,6703	2	2
Diosgenina	-2,618	-3,5891	0,9711	414,63	386,70	499,40	9,2057	-9,206	1,452	2,268	30,1321	1	3
Etil-4-aminobenzoato (Benzocaina)	-2,616	-2,0665	-0,5495	165,19	204,74	201,40	8,5342	-8,534	0,216	4,021	14,0161	1	3
5-Metil-5-(3-metilbut-2-enil)barbiturato	-2,602	-2,4089	-0,1931	210,23	243,85	261,67	9,4312	-9,431	-0,055	4,700	15,5039	2	3
5-i-Propyl-5-(3-methylbut-2-enyl)barbiturato	-2,593	-2,5877	-0,0053	238,29	276,26	315,62	9,5243	-9,524	-0,087	4,534	17,4507	2	3
Mitomicina-C	-2,564	-2,1092	-0,4548	334,33	289,64	351,48	8,7855	-8,785	-0,961	3,154	23,6124	3	8
Teobromina	-2,557	-2,3454	-0,2116	180,17	193,14	198,09	9,0144	-9,014	-0,207	4,177	13,8872	1	3
Propilparabeno	-2,557	-2,3048	-0,2522	180,20	220,62	219,65	9,3580	-9,358	-0,153	3,066	14,2596	1	3
Lomefloxacina	-2,533	-2,0144	-0,5186	351,35	337,99	397,63	8,6961	-8,696	-0,606	5,767	29,1307	2	8
Carbofuran	-2,500	-2,9185	0,4185	221,26	253,97	265,99	8,7442	-8,744	0,390	2,177	17,5728	1	3
Acetazolamida	-2,489	-1,9527	-0,5363	222,25	210,04	209,83	10,0272	-10,027	-1,739	8,815	13,9032	2	6
Amobarbital	-2,470	-2,4364	-0,0336	226,27	237,93	262,97	11,1945	-11,194	-0,084	1,425	14,6646	2	3
Prostaglandina-E2	-2,470	-2,9884	0,5184	352,47	397,24	458,53	9,4121	-9,412	0,961	1,824	25,3772	3	5
Isocarboxazida	-2,461	-2,3947	-0,0663	231,25	258,19	279,26	9,7205	-9,720	0,068	6,577	18,8093	2	4

Vinbarbital	-2,458	-2,7643	0,3063	224,26	244,12	271,24	9,8878	-9,888	-0,040	1,720	16,3799	2	3
Alopurinol	-2,453	-1,4397	-1,0133	136,11	150,49	143,93	9,7460	-9,746	-0,523	0,682	10,1494	2	4
Pentobarbital	-2,410	-2,4618	0,0518	226,27	245,92	277,56	11,0919	-11,092	-0,050	1,730	15,8439	2	3
Heptobarbital	-2,380	-3,0278	0,6478	218,21	222,63	243,77	9,8191	-9,819	-0,130	1,696	16,4307	2	3
Fenacetina	-2,371	-2,8379	0,4669	179,22	224,08	222,20	8,2161	-8,216	0,528	4,289	15,6496	1	2
Fenobarbital	-2,366	-3,2184	0,8524	232,24	235,73	262,34	9,7364	-9,736	-0,090	1,730	17,7645	2	3
Ciclobutano- spirobarbiturato	-2,349	-1,8609	-0,4881	182,18	190,34	202,00	11,1563	-11,156	-0,111	1,583	12,3625	2	3
Etilparabeno	-2,346	-2,0327	-0,3133	166,18	200,46	197,24	9,3577	-9,358	-0,153	3,081	13,0470	1	3
5-Alil-5-fenilbarbiturato	-2,346	-3,4345	1,0885	244,25	249,99	280,73	9,7795	-9,779	-0,123	1,703	19,2419	2	3
6-Aminopteridina	-2,343	-1,4958	-0,8472	147,14	165,28	161,12	9,1844	-9,184	-0,968	5,018	13,1138	1	5
5-Etil-5- pentilbarbiturato	-2,340	-2,5896	0,2496	226,27	257,74	282,23	11,1654	-11,165	-0,077	1,591	15,9039	2	3
Secobarbital	-2,333	-2,7500	0,4170	238,29	249,96	290,88	10,3546	-10,355	-0,075	1,780	17,2653	2	3
Iodamida	-2,321	-5,1370	2,8160	627,94	332,01	393,48	9,0932	-9,093	-0,949	6,640	27,7000	3	4
4-Aminopteridina	-2,313	-1,3443	-0,9687	147,14	163,98	159,97	9,4195	-9,419	-0,840	4,311	12,6065	1	5
7-Aminopteridina	-2,313	-1,4307	-0,8823	147,14	165,77	161,50	9,4893	-9,489	-0,946	4,601	12,9438	1	5
Hipoxantina	-2,280	-1,9177	-0,3623	136,11	149,75	143,14	9,2974	-9,297	-0,306	2,207	10,3679	2	3
Sulfametazina	-2,268	-2,4210	0,1530	278,33	289,60	307,80	9,3403	-9,340	-0,265	5,957	21,9465	2	6
Busulfan	-2,267	-0,8563	-1,4107	246,30	258,75	255,33	11,6457	-11,646	-0,382	0,002	12,3286	0	6
Cocaina	-2,260	-2,4351	0,1751	303,36	316,66	363,82	9,2927	-9,293	-0,105	2,950	22,9626	0	5
5-Etil-5-(3-metilbut- 2-enil)barbiturato	-2,253	-2,6562	0,4032	224,26	247,23	277,85	10,2298	-10,230	-0,055	1,661	16,4504	2	3
Propiltiouracil	-2,185	-3,5208	1,3358	170,23	203,84	204,56	9,1957	-9,196	-0,616	5,786	14,4625	2	1
Chlordiazepoxido	-2,176	-3,9281	1,7521	299,76	301,33	339,24	8,4756	-8,476	-0,447	2,878	27,6619	1	4
2-Naftol	-2,159	-3,1837	1,0247	144,17	174,50	173,76	8,5198	-8,520	-0,115	0,937	13,6700	1	1
Tianfenicol	-2,154	-3,1934	1,0394	356,22	321,63	378,23	10,3640	-10,364	-0,974	5,343	21,0585	3	5
Probarbital	-2,153	-1,8495	-0,3035	198,22	209,08	234,00	11,1342	-11,134	-0,073	1,827	13,5741	2	3
Atropina	-2,124	-2,6103	0,4863	289,37	297,29	359,35	9,4814	-9,481	0,353	2,853	22,3158	1	4

Cloranfenicol	-2,111	-3,2833	1,1723	323,13	306,52	353,82	10,3213	-10,321	-1,102	4,040	20,9576	3	5
Tetroxoprim	-2,101	-1,8907	-0,2103	334,38	381,44	422,74	8,7904	-8,790	-0,255	4,989	28,6863	2	8
Canfora	-2,086	-0,1894	-1,8966	152,24	184,00	207,13	9,9701	-9,970	1,269	2,809	11,1416	0	1
Ceftazidima	-2,038	-2,2801	0,2421	546,57	490,75	580,79	8,3222	-8,322	-0,441	4,164	43,5840	3	10
Propoxur	-2,020	-1,1179	-0,9021	209,24	239,76	261,63	9,1987	-9,199	0,317	1,787	16,2619	0	5
Talbutal	-2,016	-2,4845	0,4685	224,26	238,07	272,01	10,3012	-10,301	-0,057	1,892	15,8748	2	3
Ácido p-Aminosalicilico	-1,963	-1,5338	-0,4292	153,14	172,02	168,65	8,6737	-8,674	0,181	5,439	11,8212	3	4
2-Hidroxipteridina	-1,947	-1,6057	-0,3413	148,12	161,57	156,46	10,0301	-10,030	-1,256	2,812	12,1449	1	5
Hiosciamina	-1,910	-2,6914	0,7814	289,37	303,03	361,52	9,3839	-9,384	0,266	2,475	21,9615	1	4
Salicilamida	-1,836	-1,7301	-0,1059	137,14	160,93	156,88	9,1880	-9,188	-0,069	3,902	10,4977	2	2
L-DOPA	-1,818	-1,9103	0,0923	198,19	218,41	227,73	8,6784	-8,678	0,473	1,592	14,1878	4	5
7-Butilteofilina	-1,805	-2,9100	1,1050	236,27	265,08	284,79	8,8916	-8,892	-0,103	3,811	18,8574	0	3
Ácido salicílico	-1,804	-1,3356	-0,4684	138,12	158,10	153,48	9,3250	-9,325	-0,230	3,132	10,1463	2	3
Alobarbital	-1,796	-2,4830	0,6870	208,22	225,46	249,33	10,2843	-10,284	-0,071	1,446	15,3028	2	3
Aspirina	-1,733	-1,3974	-0,3356	180,16	205,85	221,36	10,1553	-10,155	-0,965	5,561	14,1289	1	4
Adenosina	-1,728	-1,7480	0,0200	267,24	258,49	287,49	9,0633	-9,063	-0,276	1,686	19,3035	4	8
Sacarina	-1,725	-2,4472	0,7222	183,18	177,90	180,14	10,7437	-10,744	-1,387	5,086	11,9002	1	3
Azintamida	-1,716	-3,5432	1,8272	259,75	270,38	292,88	9,1801	-9,180	-0,891	4,205	20,1218	0	3
Aprobarbital	-1,710	-2,2092	0,4992	210,23	221,60	251,42	10,3190	-10,319	-0,056	1,820	14,7525	2	3
Metil-p-hidroxibenzoato	-1,705	-1,6220	-0,0830	152,15	189,97	186,74	9,5797	-9,580	-0,139	2,429	12,3570	1	3
Disopiramida	-1,701	-3,6447	1,9437	339,48	355,14	445,97	9,1421	-9,142	0,336	3,289	28,3778	1	3
Tropicamida	-1,698	-3,5877	1,8897	284,36	276,56	345,19	9,1497	-9,150	-1,080	3,877	22,7065	1	3
Baclofeno	-1,696	-2,7165	1,0205	213,66	228,85	243,38	9,4722	-9,472	0,021	1,990	14,8402	2	3
Butobarbitone (Butetal)	-1,686	-2,3025	0,6165	212,25	237,29	259,62	11,1646	-11,165	-0,077	1,561	14,7620	2	3
Ciclopropano- spirobarbiturato	-1,655	-1,5594	-0,0956	168,15	179,49	185,74	11,2091	-11,209	-0,161	1,542	11,2852	2	3
Aztreonam	-1,639	-1,3156	-0,3234	435,43	400,06	464,48	9,1319	-9,132	-0,587	7,362	29,2105	3	10
5-Etil-5-alilbarbiturato	-1,614	-1,9961	0,3821	196,21	209,87	232,32	10,4608	-10,461	-0,066	1,505	13,6666	2	3

Cimetidina	-1,613	-2,4096	0,7966	252,34	277,53	286,47	8,3123	-8,312	0,558	8,361	19,4264	3	5
7-Isobutil-8-metilteofilina	-1,599	-2,8748	1,2758	250,30	271,87	306,53	8,7682	-8,768	-0,035	4,311	20,1893	0	3
Ácido Benzóico	-1,555	-0,7446	-0,8104	122,12	169,15	166,02	10,0475	-10,048	-0,766	3,497	10,4223	1	2
Pteridina-7-metil-tiol	-1,551	-2,5942	1,0432	178,21	192,41	194,07	9,3489	-9,349	-1,231	3,952	15,8287	0	4
Ampicillina	-1,539	-3,3069	1,7679	349,40	330,71	402,40	9,6048	-9,605	0,216	1,611	25,9901	3	5
Timina	-1,499	-1,3573	-0,1417	126,11	148,95	140,82	9,5696	-9,570	-0,125	4,172	9,3604	2	2
Uracil	-1,493	-0,7834	-0,7096	112,09	131,76	120,29	9,9158	-9,916	-0,172	4,242	7,9245	2	2
5-Etil-5-propilbarbiturato	-1,491	-1,9578	0,4668	198,22	217,54	238,27	11,1672	-11,167	-0,080	1,592	13,6473	2	3
5-Etil-barbiturato	-1,427	-1,2359	-0,1911	156,14	173,43	174,57	11,2727	-11,273	-0,175	1,353	10,3882	2	3
5_5-Dietilbarbiturato	-1,410	-1,7189	0,3089	184,19	198,18	214,12	11,1316	-11,132	-0,090	1,649	12,6453	2	3
Acetanilida	-1,398	-1,8375	0,4395	135,17	172,28	170,93	9,0975	-9,097	0,236	3,470	11,5953	1	1
Ácido 4-Aminobenzoico	-1,368	-1,4807	0,1127	137,14	164,70	159,30	8,6075	-8,608	0,123	4,430	11,2668	2	3
Sulfanilamida	-1,361	-1,1959	-0,1651	172,20	181,80	181,40	9,3677	-9,368	-0,282	8,453	11,4527	2	4
Teofilina	-1,335	-2,2753	0,9403	180,17	209,21	219,03	8,9307	-8,931	-0,126	3,644	15,2488	1	3
Sulfamerazina	-1,218	-2,2993	1,0813	264,30	272,40	285,41	9,3691	-9,369	-0,295	6,580	20,6322	2	6
Cefradina	-1,215	-2,9508	1,7358	349,40	332,19	409,87	9,2220	-9,222	0,408	4,821	26,1252	3	5
1-Propilteobromina	-1,207	-2,7032	1,4962	222,25	245,99	264,11	8,8917	-8,892	-0,098	3,495	17,5952	0	3
5-Metil-5-etilbarbiturato	-1,162	-1,2770	0,1150	170,17	184,36	196,60	11,2104	-11,210	-0,096	1,333	11,3838	2	3
5-Metil-5-alilbarbiturato	-1,160	-1,7252	0,5652	182,18	197,39	214,17	10,5237	-10,524	-0,082	1,342	12,6386	2	3
6-Metoxipteridina	-1,139	-1,5992	0,4602	162,15	181,37	177,57	9,8642	-9,864	-1,167	3,579	13,6362	0	5
5-Metil-barbiturato	-1,126	-0,9860	-0,1400	142,11	156,94	152,70	11,3257	-11,326	-0,185	1,021	9,2804	2	3
6-Clorpteridina	-1,124	-2,3388	1,2148	166,57	171,27	169,29	10,3466	-10,347	-1,537	1,167	12,5466	0	4
4-Metoxipteridina	-1,112	-1,3314	0,2194	162,15	177,39	176,98	10,0405	-10,040	-1,132	4,337	13,3823	0	5
2-Metoxipteridina	-1,112	-1,7168	0,6048	162,15	181,24	177,82	9,8878	-9,888	-1,147	2,201	13,7633	0	5
Zidovudina	-1,029	-2,1588	1,1298	267,24	269,08	299,72	9,5468	-9,547	-0,335	3,503	20,0299	2	6
4-Dimetilaminopteridina	-1,021	-2,1094	1,0884	175,19	197,68	201,85	8,9052	-8,905	-0,703	4,993	16,0478	0	4
Nadolol	-1,008	-3,1620	2,1540	309,40	344,52	399,01	9,3389	-9,339	0,340	1,792	23,3757	4	5

Metocarbamol	-0,985	-1,8123	0,8273	241,24	261,71	285,20	9,3718	-9,372	0,120	5,380	17,2412	2	5
Flucitosina	-0,959	-1,2003	0,2413	129,09	142,92	133,75	8,9893	-8,989	-0,356	6,171	9,1323	2	3
Benzamida	-0,953	-1,1687	0,2157	121,14	174,65	171,26	9,7621	-9,762	-0,279	2,710	10,8619	1	1
Colchicina	-0,944	-2,6544	1,7104	399,44	400,12	498,00	9,2752	-9,275	-0,614	8,541	34,3418	1	6
Minociclina	-0,944	-2,0208	1,0768	457,48	419,05	516,69	8,9222	-8,922	-0,644	6,483	35,8080	5	10
Didanosina	-0,937	-2,1903	1,2533	236,23	240,84	261,21	9,0666	-9,067	-0,379	6,178	17,6084	2	5
Pirazinamida	-0,914	-0,6976	-0,2164	123,11	147,63	139,20	10,4132	-10,413	-0,899	3,339	9,0732	1	3
7-Metoxipteridina	-0,910	-1,5100	0,6000	162,15	180,82	178,14	10,0652	-10,065	-1,221	4,328	13,7082	0	5
7-Clorpteridina	-0,876	-2,2391	1,3631	166,57	171,15	169,17	10,3694	-10,369	-1,523	2,359	12,5543	0	4
7-Metilpteridina	-0,854	-1,4886	0,6346	146,15	170,89	167,88	10,0119	-10,012	-1,143	3,086	12,5341	0	4
7-Etilteofilina	-0,757	-2,5139	1,7569	208,22	225,36	240,17	8,8983	-8,898	-0,107	3,778	16,4610	0	3
1-Etilteobromina	-0,719	-2,4966	1,7776	208,22	226,35	241,75	8,8906	-8,891	-0,097	3,515	16,4160	0	3
Propranolol	-0,714	-4,0217	3,3077	259,35	310,26	335,30	8,3991	-8,399	-0,047	1,121	22,7776	2	3
2-Clorpteridina	-0,699	-2,2055	1,5065	166,57	171,56	168,20	10,3679	-10,368	-1,513	4,056	12,6568	0	4
Guaifenesina	-0,598	-2,0499	1,4519	198,22	230,24	239,17	8,7888	-8,789	0,415	1,491	14,8761	2	4
Etambutol	-0,565	-1,1537	0,5887	204,31	272,69	287,06	8,9958	-8,996	2,548	2,202	14,9290	4	4
Pentoxifilina	-0,558	-2,8715	2,3135	278,31	294,13	338,14	9,2779	-9,278	-0,748	3,526	21,2848	0	4
Metiprilon	-0,382	-1,2399	0,8579	183,25	208,16	236,13	10,0083	-10,008	0,747	2,715	12,9331	1	2
Cefamandol	-0,143	-2,0817	1,9387	462,50	415,74	500,65	9,3467	-9,347	-0,071	10,038	34,7246	3	8
Penicilamina	-0,128	0,1225	-0,2505	149,21	169,61	181,55	9,5159	-9,516	0,707	0,787	9,5189	2	3
2-Metilpteridina	-0,094	-1,6814	1,5874	146,15	170,94	167,06	9,9722	-9,972	-1,125	1,898	12,6774	0	4
7-Dimetilaminopteridina	-0,021	-2,2736	2,2526	175,19	201,54	203,71	9,2754	-9,275	-0,925	4,497	16,2794	0	4
Pteridina	0,021	-1,1685	1,1895	132,12	151,42	146,07	10,0793	-10,079	-1,232	2,667	10,9605	0	4
Dimorfolamina	0,098	-2,7754	2,8734	398,54	413,87	526,24	9,0463	-9,046	0,942	4,968	29,6360	0	4
Ácido ascórbico	0,277	-0,6064	0,8834	176,13	180,64	183,34	9,6710	-9,671	-0,380	3,093	10,4733	4	6
2-Dimetilaminopteridina	0,360	-2,6291	2,9891	175,19	202,13	204,21	8,7520	-8,752	-0,752	1,265	16,8831	0	4
Antipirina	0,480	-2,7727	3,2527	188,23	220,88	232,72	8,6294	-8,629	0,041	4,550	16,9645	0	2
Proxifilina	0,623	-2,4910	3,1140	238,25	253,00	276,22	8,8260	-8,826	-0,036	4,233	18,1960	1	4

Sorbitol	1,148	0,9548	0,1932	182,17	199,25	205,92	10,3807	-10,381	2,374	6,048	10,0747	6	6
Prolina	1,149	1,6415	-0,4925	115,13	147,48	141,27	9,7147	-9,715	0,934	2,840	7,3839	2	3

Notas 1 – legenda para o cabeçalho da Tabela 1

logS- logaritmo da solubilidade das substâncias em água (S em mol/L), obtido da literatura [11];

logScalc – logS calculado a partir da equação geral obtida no item 4.1 (S em mol/L);

Residual= logS-logScalc

PM- Peso molecular (em g/mol), obtido da literatura [11];

VOL* – volume das moléculas (A^3); AREA * – área das moléculas (A^2); PI* - potencial de ionização (EV);

HOMO* – “High occupied molecular orbital” (EV); LUMO* – “Lowest unoccupied molecular orbital” (EV);

MD* – momento dipolar (debye); ALFA* – polarizabilidade (alfa) (a 0,5 EV; em $ANG.^3$)

Hdr – Número de Hidrogênios doadores de pontes de Hidrogênio em cada molécula.

HAcrr- Número de átomos aceptores de pontes de hidrogênio em cada molécula.

*(Todos os dados com cálculos desenvolvidos pelo MOPAC a partir dos parâmetros das moléculas desenhadas no HyperChem. Estes mais Hdr e HAcrr foram os dados utilizados para o cálculo da equação geral).

3.2) Análise dos dados obtidos das propriedades x logS.

3.2.1 – Entendendo a correlação de dados:

A solubilidade está expressa aqui como o logaritmo⁵ da solubilidade, logS, por melhor se correlacionar com todas as propriedades obtidas do que utilizando simplesmente a própria solubilidade S(g/L). Isso é observável se introduzida uma nova coluna da Tabela 1, calculada como 10**logS/PM (g/L), ou seja solubilidade como S(g/L) e ser feito o relacionamento por correlação de Pearson com todas as propriedades. Tomando como exemplo a relação com Polarizabilidade alfa, obtem-se um índice (r) de -0,164 e com Momento dipolar com um (r) de 0,022, índices muito inferiores aos encontrados para logS x Polarizabilidade alfa e logS x momento dipolar, com valores de (r) respectivamente 0,458 e -0,310.

O coeficiente de correlação de Pearson pode ser entendido da seguinte maneira:

se temos series de n medidas de X e Y escritos como x_i e y_i onde $i = 1, 2, \dots, n$, o coeficiente de correlação r entre X e Y de uma amostra é dado por:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)s_x s_y},$$

onde \bar{x} and \bar{y} são médias de X e Y e s_x e s_y representam o desvio padrão de X e Y [31]. Ela é uma medida do grau de relação linear entre as variáveis, onde entende-se que zero indica ausência de relação, e 1 relação linear perfeita [32], [33].

A forma usada para observar a correlação visual de dados entre as diversas propriedades e logS está representada abaixo em 15 figuras e de cada uma calculado o coeficiente de correlação de Pearson. A interpretação destas aponta caminho de como trabalhar a regressão múltipla para obtenção de uma equação final que calcule logS satisfatoriamente.

⁵ A base de logaritmo considerada foi a base 10

A Figura 2, abaixo, mostra a correlação entre dados de logS x PM e possui uma linha representando a regressão linear ponderada⁶ e que, pelo seu modo de desenho, exprime uma curva que se assemelha a uma função $f(x)=1/x$. A relação tem um coeficiente de correlação de Pearson (r) de -0,310. Uma interpretação matemática da curva sugere que se trabalhado o seu inverso a linha de regressão ponderada passa a ter uma aproximação maior de uma reta e se torna ascendente. Isto considerado, na equação final não altera os dados.

A Figura 3, demonstrando o relacionamento entre logS x $1/PM*1000$ é a

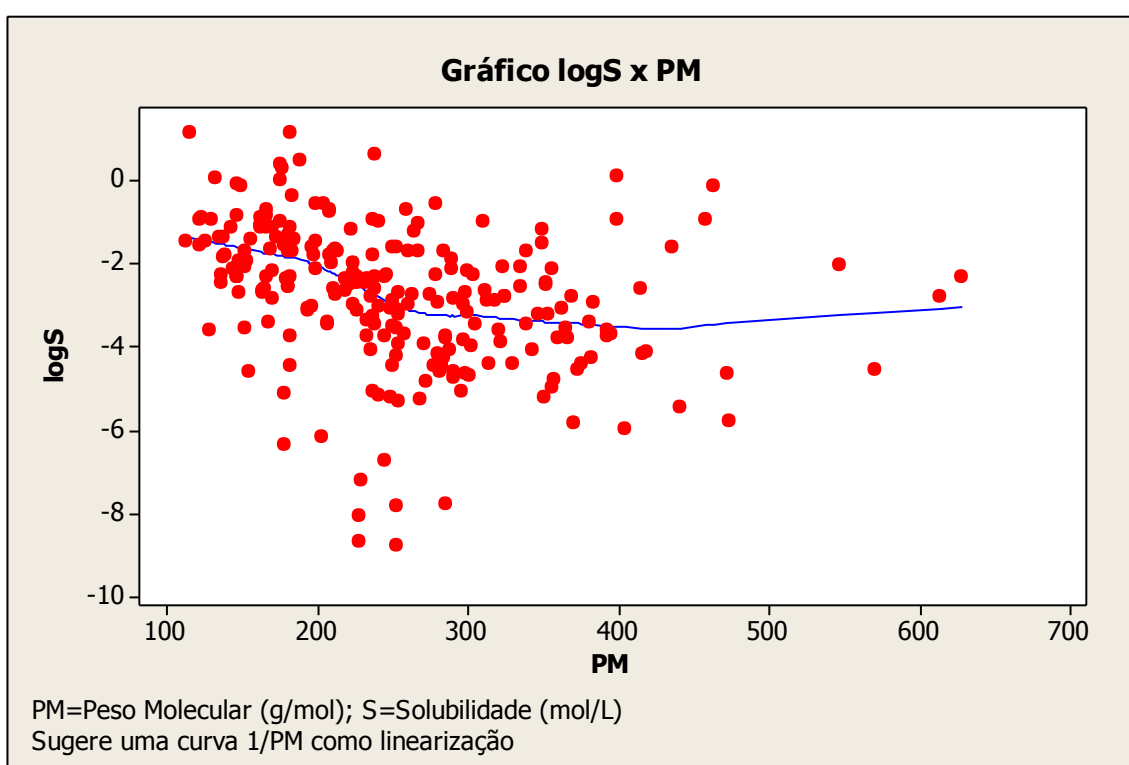


Figura 2 – Demonstração do relacionamento entre logS x PM onde a “linha central” representa a regressão linear ponderada dos pontos.

a expressão do raciocínio em relação a inversão de logS x PM. A multiplicação de $1/PM$ por mil foi aplicada em função do escalonamento dos dados. O coeficiente de correlação de Pearson (r) é de 0,392, cerca de 8% superior ao da Figura 2 o que explica a possibilidade de influência positiva na precisão da equação final.

⁶ Recurso do Minitab®, em inglês, “smoother line” que está em todos os gráficos representando uma regressão linear ponderada, ou seja, uma regressão feita por partes.

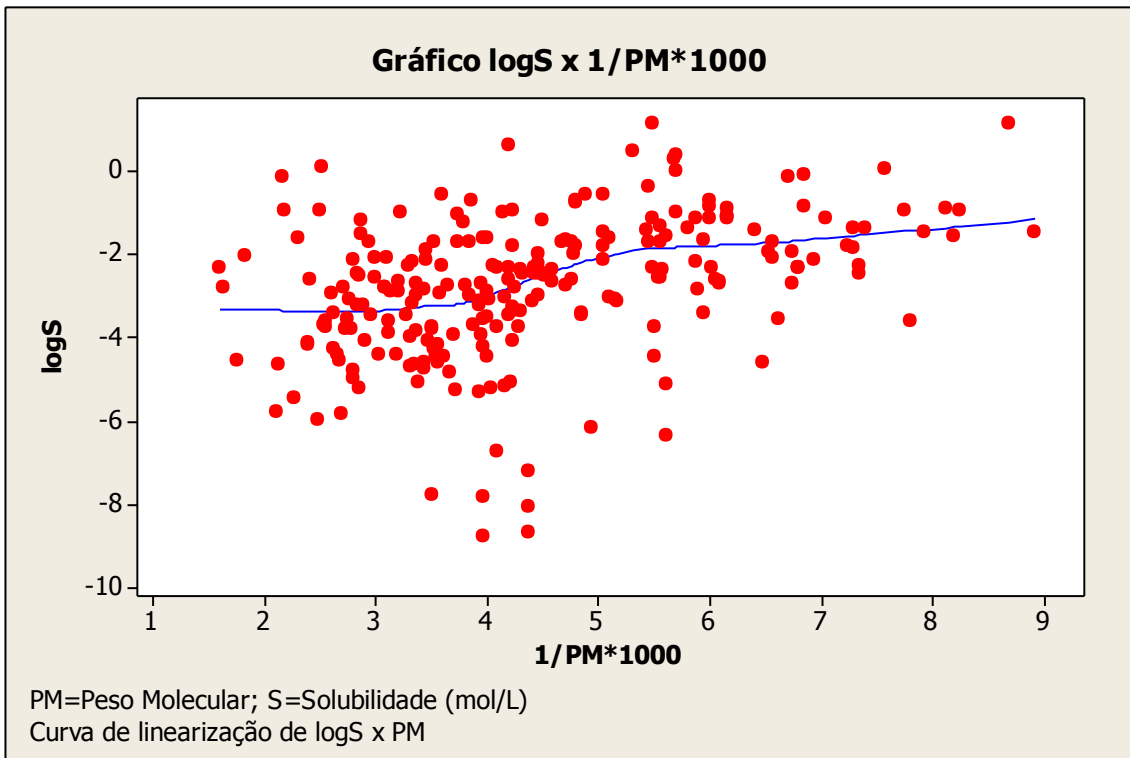


Figura 3 – Demonstração do relacionamento entre $\log S \times 1/PM \times 1000$ onde a “linha central” representa a regressão linear ponderada dos pontos.

A Figura 4 representa $\log S \times \text{AREA}$, do mesmo modo que a Figura 2 possui uma linha de regressão ponderada que aparenta uma equação $f(x)=1/x$, em menor acentuação. A sua inversão deve também somar para linearização dos dados. A Figura 5 representa exatamente essa inversão. Os coeficientes de correlação de Pearson são: Figura 4, 0,321 e Figura 5, 0,363, representando uma melhora de 4,2% em linearização.

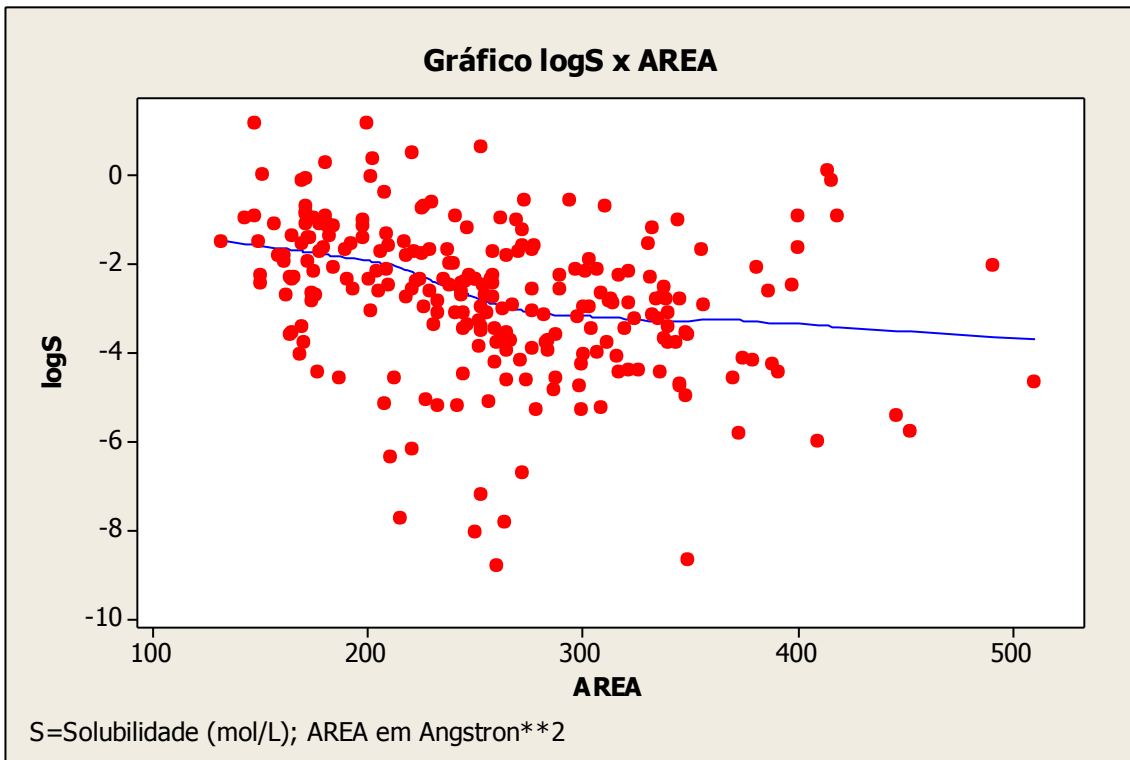


Figura 4 – Demonstração do relacionamento de logS x AREA onde a “linha central” representa a regressão linear ponderada dos pontos.

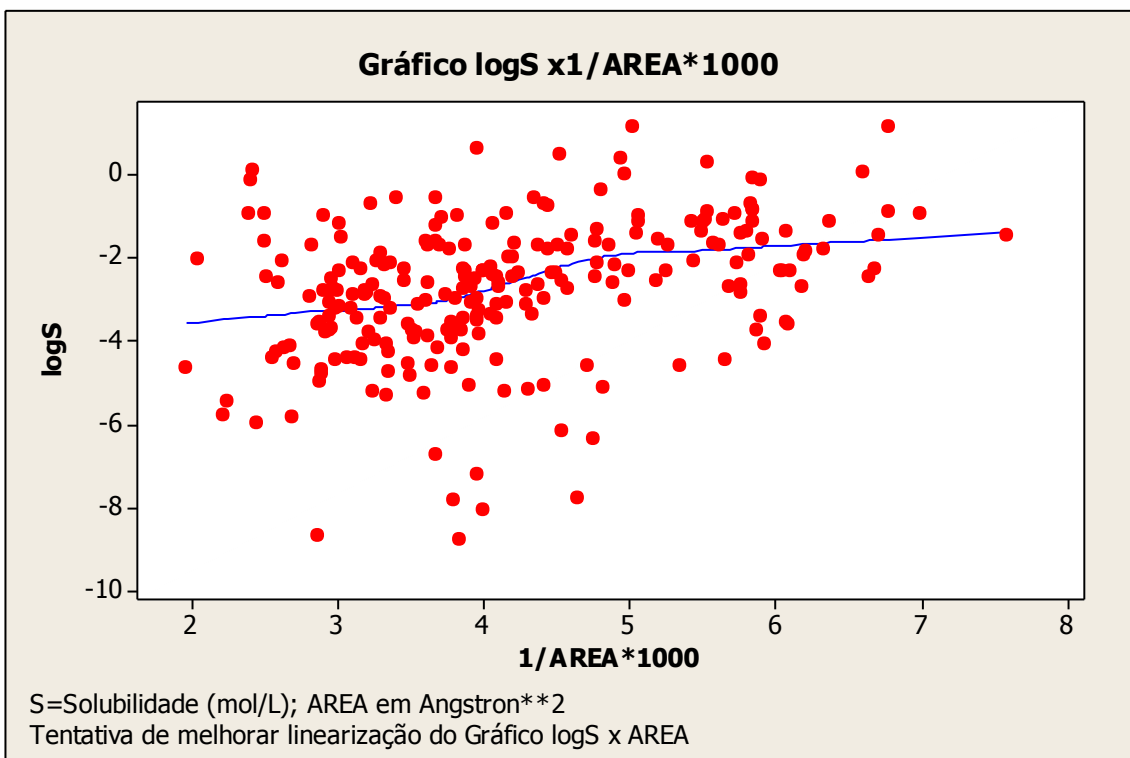


Figura 5- Demonstração do relacionamento de logS x 1/AREA*1000 onde a “linha central”, representa a regressão ponderada dos pontos.

A Figura 6 representa $\log S$ x Volume com $(r) = 0,313$. A Figura 7 tem coeficiente de correlação de Pearson de 0,362, 4,9% superior ao anterior.

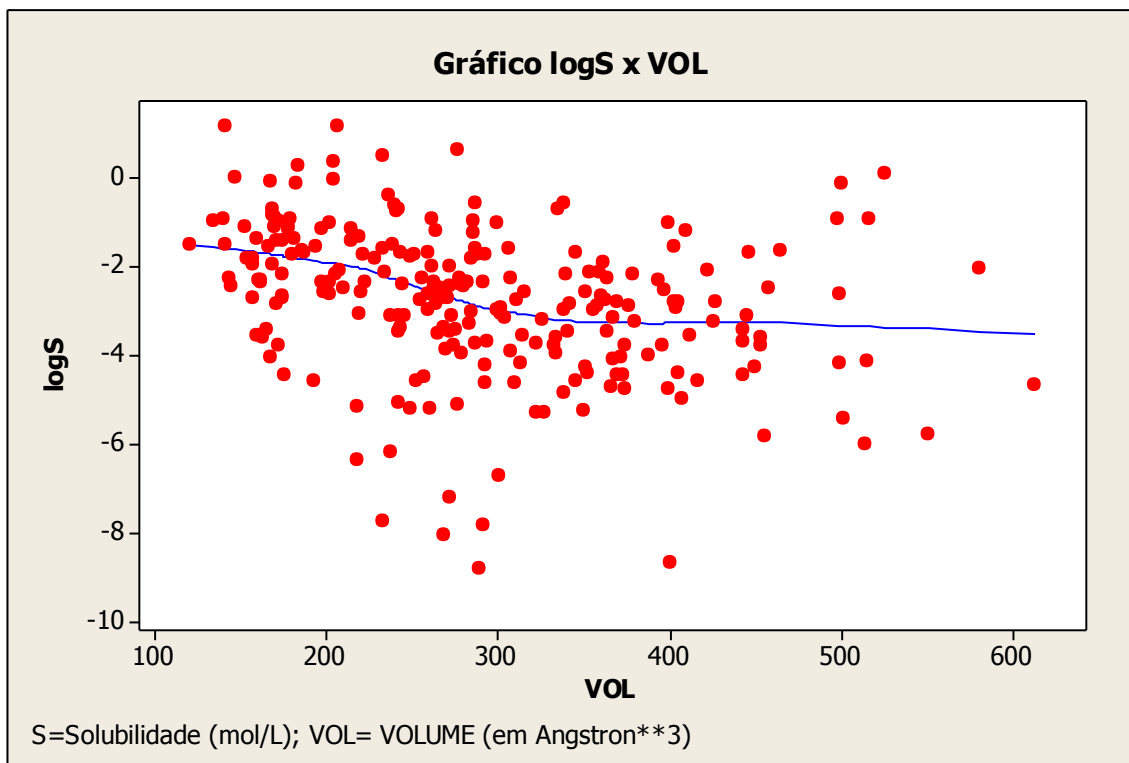


Figura 6- Demonstração do relacionamento de $\log S$ x VOL onde a “linha central” representa a regressão linear ponderada dos pontos.

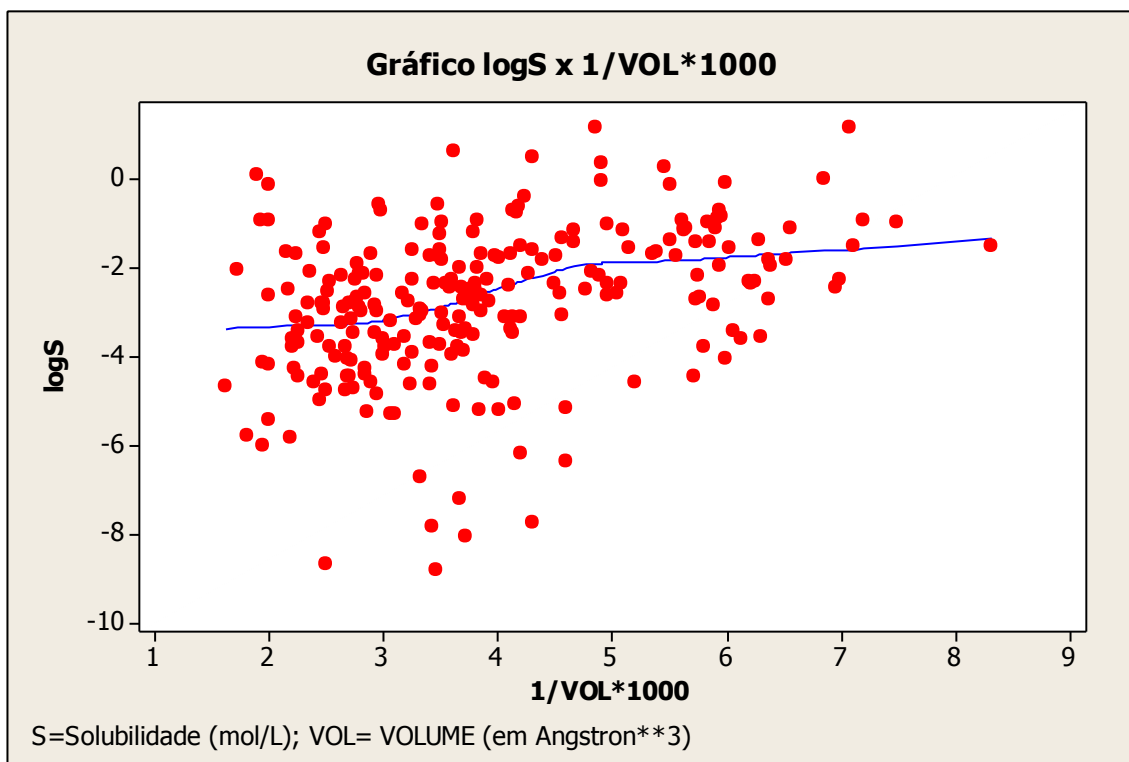


Figura 7 – Demonstração do relacionamento entre logS x 1/VOL*1000 onde a “linha central” representa a regressão linear ponderada dos pontos.

A Figura 8, logS x PI, $(r)=0,336$ com linha de regressão ponderada que aparenta-se a uma curva logarítmica ($f(x)=\log x$), originando portanto um novo gráfico, Figura 9. Esta representa logS x logPI*10 onde o coeficiente de correlação de Pearson é 0,351, mudando pouco em relação à Figura 8, 1,5%. Estes dados não foram usados na equação final devido a sua baixa contribuição para aumento de precisão da regressão. Quimicamente poderia ser interpretado de que as moléculas selecionadas para o ensaio tenham baixa ionização pois estamos trabalhando com moléculas virtualmente não ionizáveis.

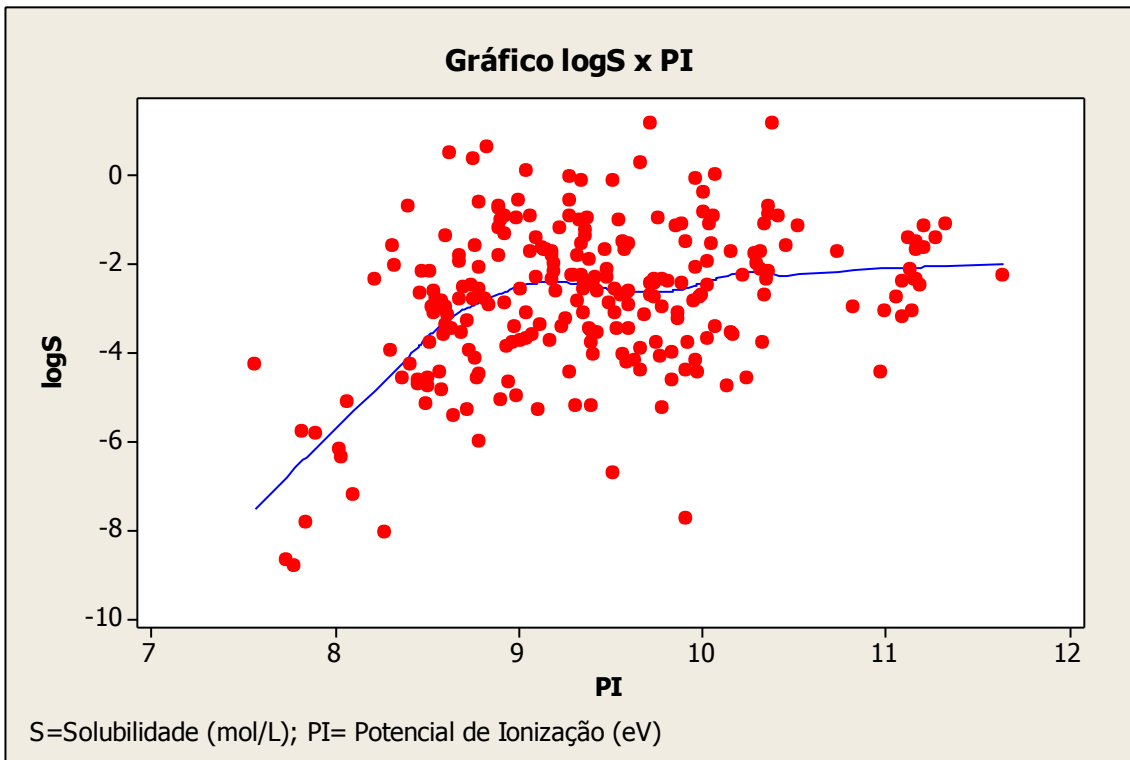


Figura 8 – Demonstração do relacionamento entre logS x PI onde a “linha central” representa a regressão linear ponderada dos pontos.

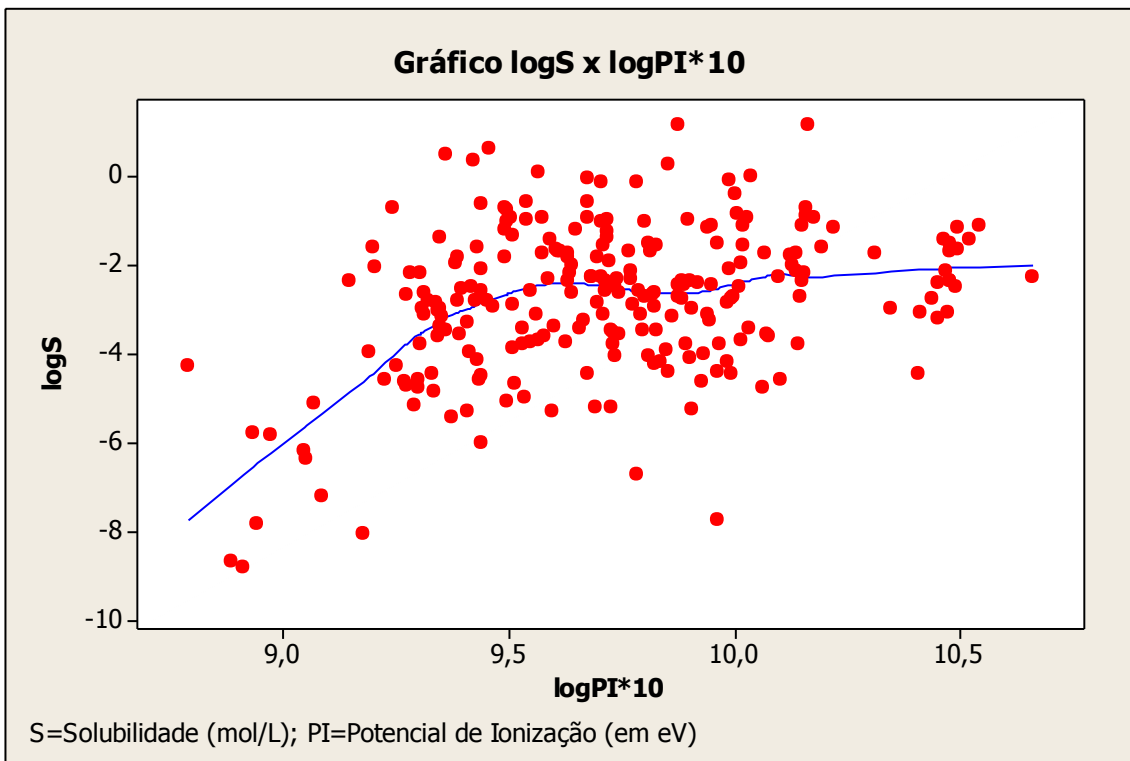


Figura 9- Demonstração do relacionamento entre logS x logPI*10 com “linha central” representando a regressão linear ponderada.

A Figura 10, representando logS x HOMO, apresenta todos os seus dados negativos, com o coeficiente de correlação de Pearson de 0,336. Neste caso foi desenvolvida a função $f(x)=\log(\text{HOMO})^2$ possibilitando um caminho de linearização. A Figura 11 representa esta transformação e possui $(r)=0,351$. Esta curva, embora aqui apresente um acréscimo de 1,5% em (r) , representa mais no ajuste da equação final, sendo nesta considerada.

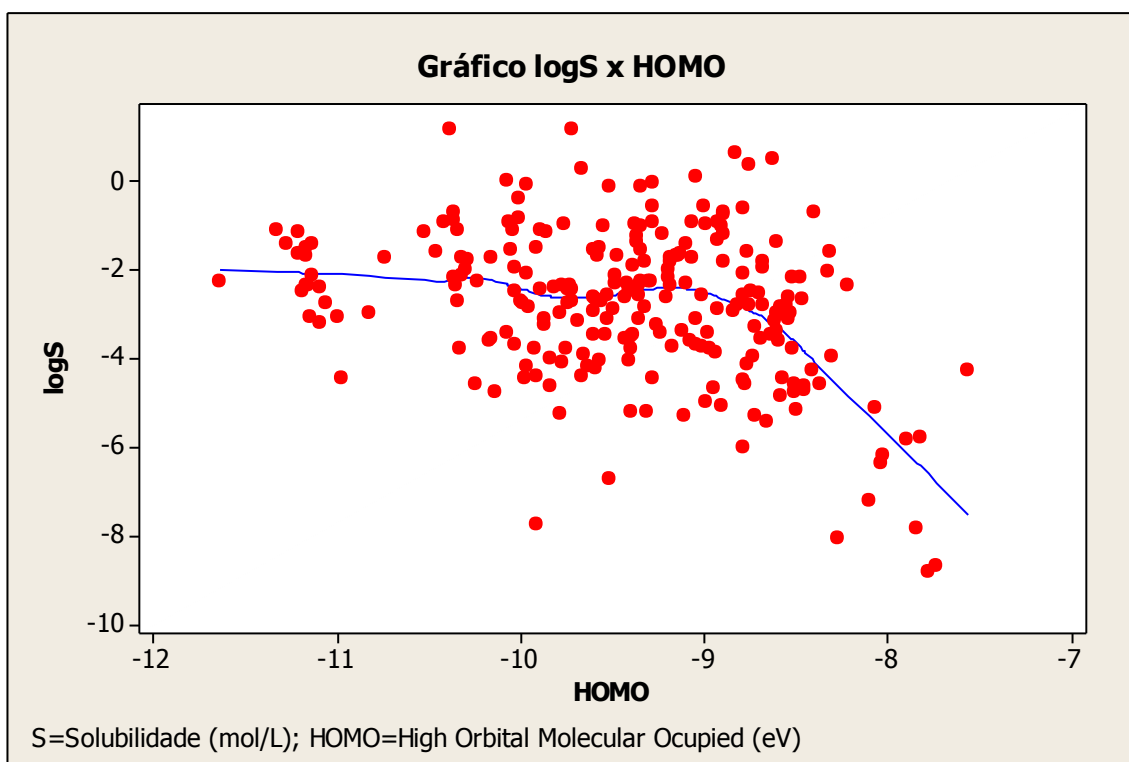


Figura 10 – Demonstração do relacionamento logS x HOMO onde a “linha central” representa a regressão linear ponderada dos pontos.

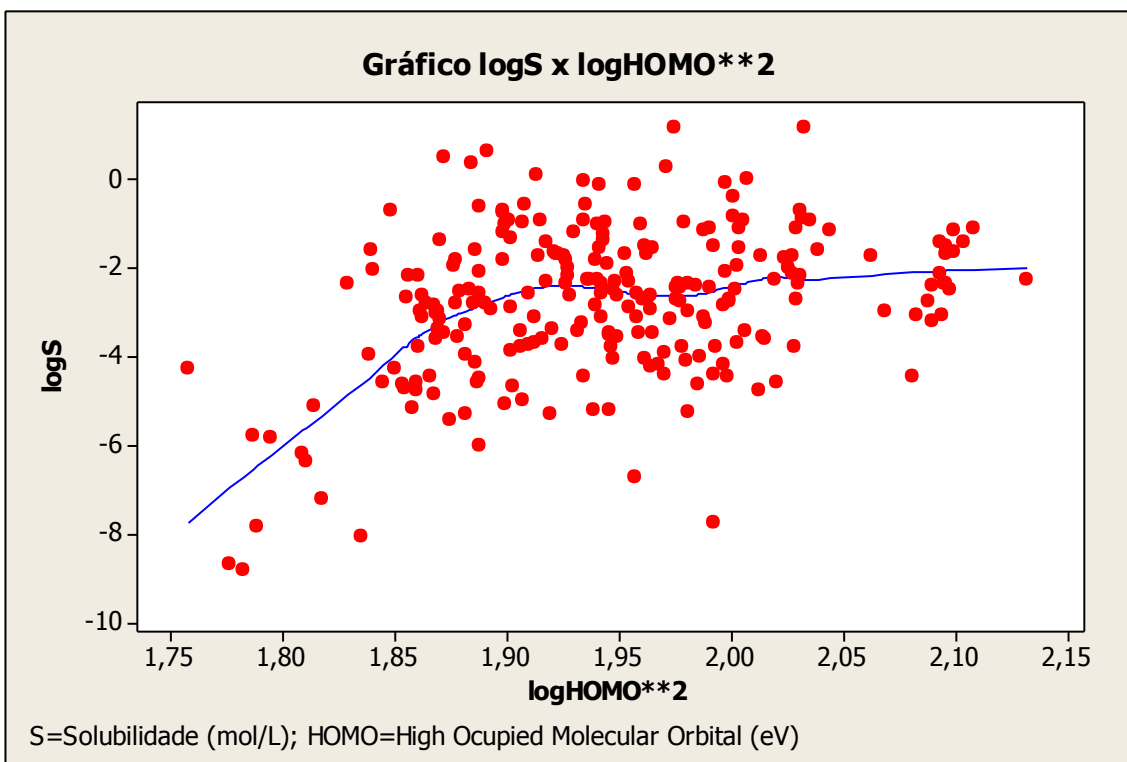


Figura 11 – Demonstração do relacionamento entre logS x logHOMO2 onde a “linha central” representa a regressão linear ponderada dos pontos.**

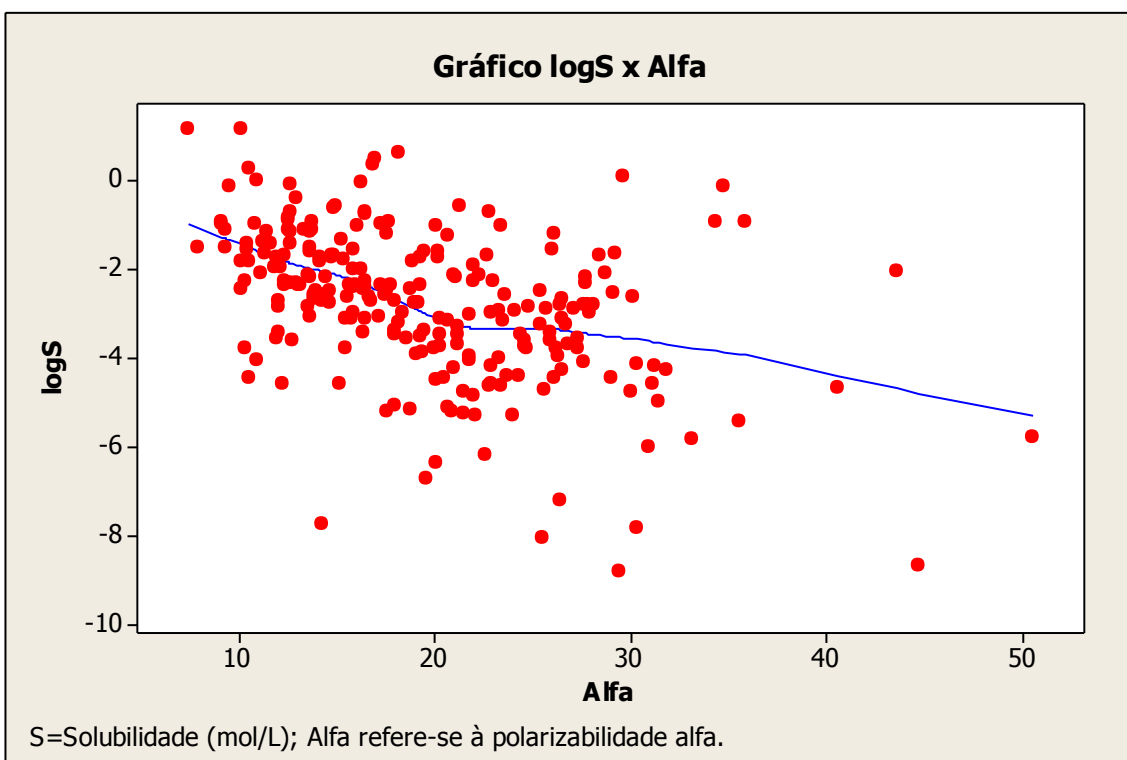


Figura 12 – Demonstração do relacionamento entre logS x Polarizabilidade (alfa) onde a “linha central” representa a regressão linear ponderada dos pontos.

A Figura 12, $\log S$ x polarizabilidade (α), acima, melhor coeficiente de correlação de Pearson até então verificado nas propriedades observadas diretamente, 0,458, representa em sua linha da regressão ponderada uma leve curva representando uma função $f(x)=1/x$. Assim a Figura 13, $\log S$ x $1/\alpha \cdot 100$, representa a transformação, tendo um (r) de 0,469, como previsto apresentando pouca correção mas de forte contribuição no cálculo final da equação.

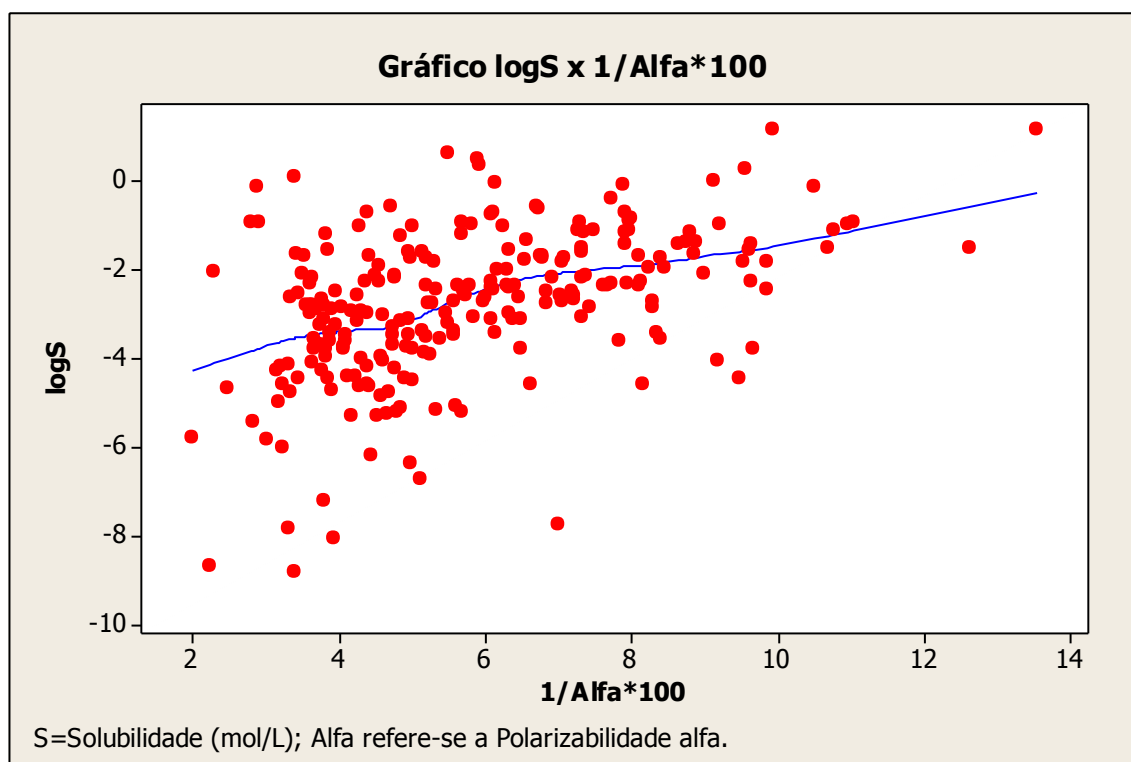


Figura 13 – Demonstração do relacionamento entre $\log S$ x $1/\text{Alfa} \cdot 100$ onde a “linha central” representa a regressão linear dos pontos.

A Figura 14, correspondente a relação entre $\log S$ x Momento dipolar, tem curva de regressão ponderada que apresenta-se como duas partes lineares (retilíneas) assim significando que não seria possível transformá-la em nova função única. A primeira parte da curva corresponde a compostos de polaridades próximas de zero o que não corresponde à solubilidades (em log) sempre muito baixas. A literatura [26] tradicionalmente ainda se refere à polaridade ou momento dipolar, se relacionando muito bem com solubilidade. Este trabalho mostra através da Figura 14 que isso não é verdade. O problema maior estaria nas moléculas de momento dipolar zero (0)?. Do mesmo modo

na parte linear mais longa da curva da Figura 14, deveríamos ter maior aproximação dos pontos da linha de regressão ponderada.

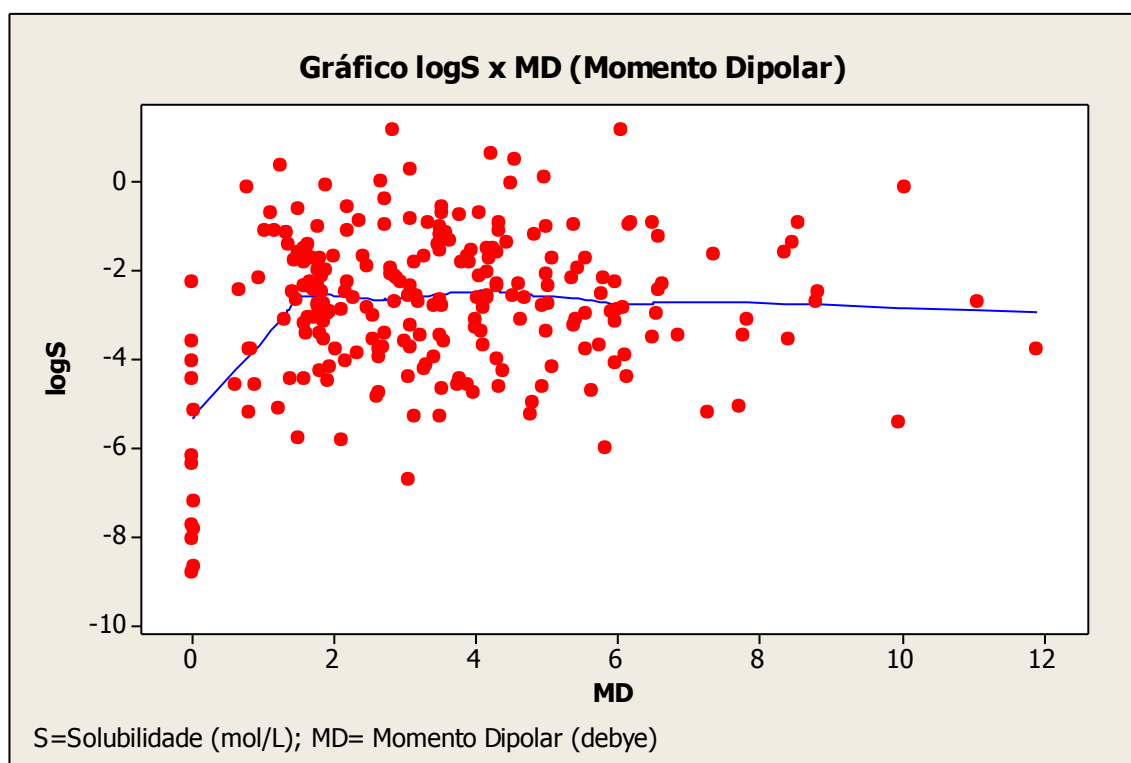


Figura 14 – Representação do relacionamento entre logS x MD onde a “linha central” representa a regressão linear ponderada dos pontos.

O que é mais provável, por este e outros trabalhos de “QSPR” é que nenhuma propriedade representa sozinha o fenômeno “simples” da solubilidade porque ele é multifatorial.

O índice de correlação de Pearson entre logS e MD é 0,310, visivelmente pequeno o que também pode ser observado pela dispersão dos dados na Figura 14.

As Figuras 15 e 16 representando respectivamente logS x número de doadores de pontes de hidrogênio e logS x número de aceptores de pontes de hidrogênio, com $(r)=0,189$ e $0,297$, também respectivamente. São índices de correlação muito baixos, certamente influenciado pelo tipo de dado, números inteiros. Não dá para desconsiderar estes dados pela sua importância teórica na formulação da idéia de formação intermolecular de pontes de hidrogênio, contribuindo para solubilização. Polarizabilidade ou momento dipolar e as outras medidas de variáveis parecem mais condizente com a realidade das moléculas. Esta questão pode-se ser entendida como permanecendo aberta embora

tenham sido utilizado o melhor método de contagem do seu significado na literatura atual [22].

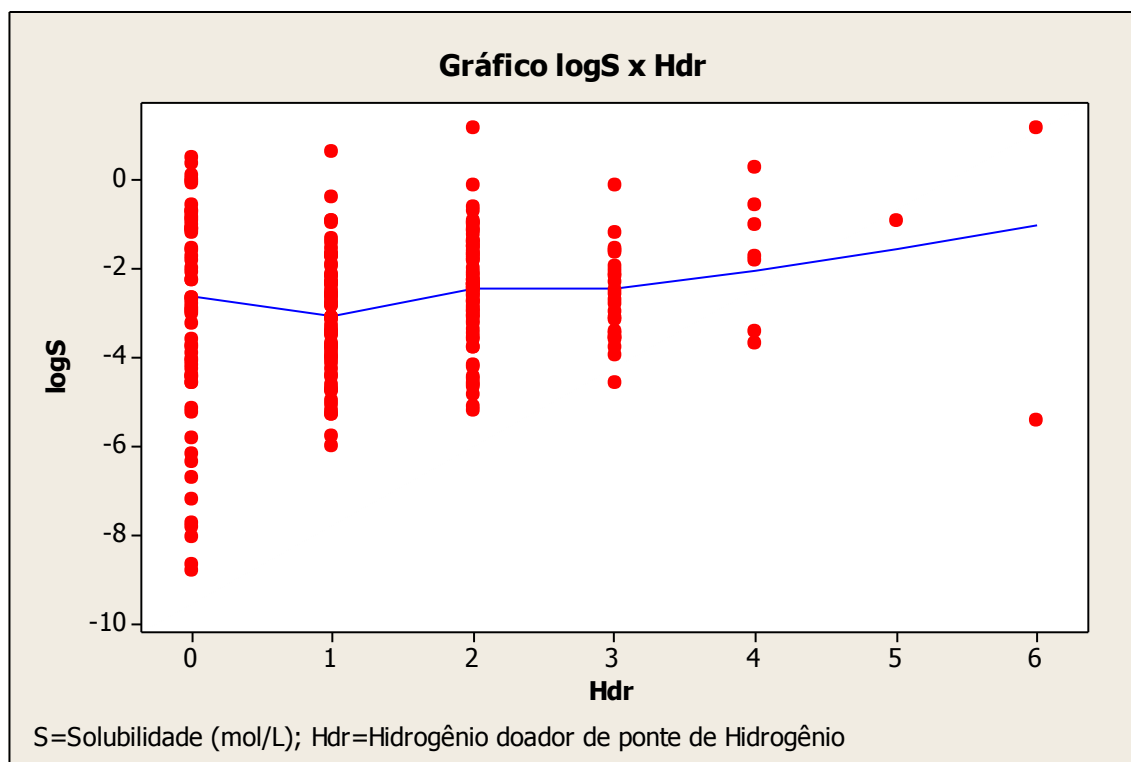


Figura 15 – Representação do relacionamento entre logS x Hdr onde a “linha central” representa uma regressão linear ponderada dos pontos.

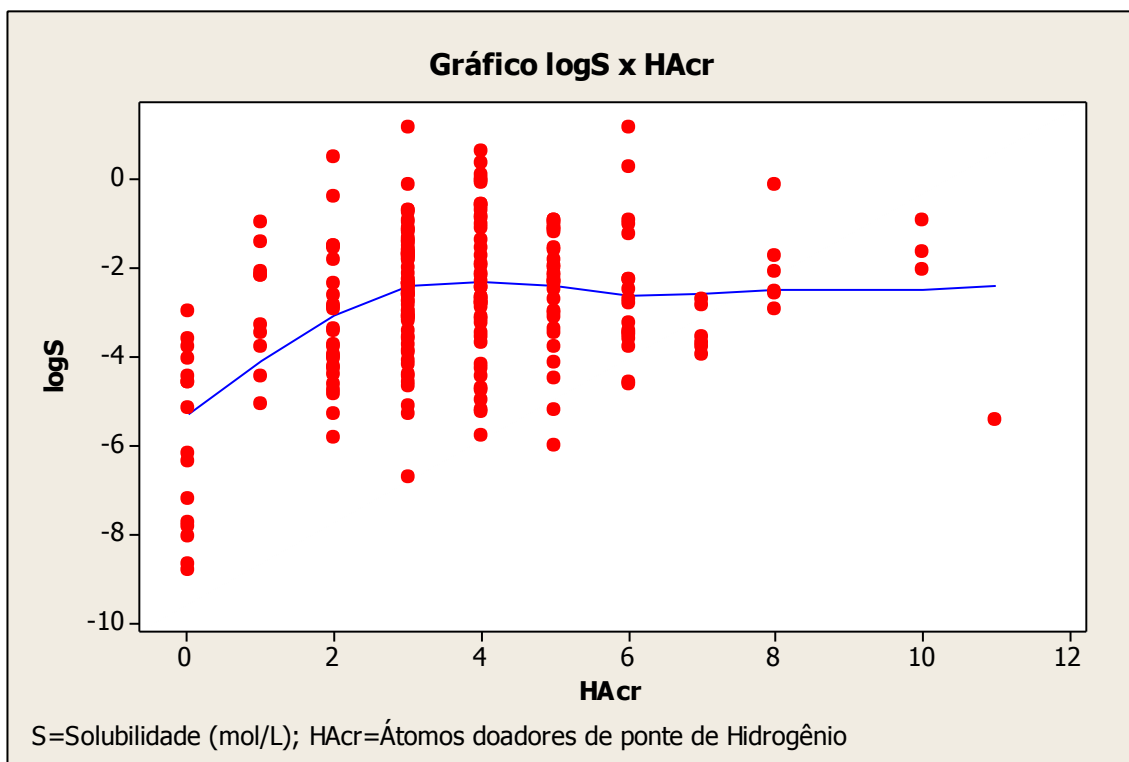


Figura 16 – Representação do relacionamento entre logS x HAcr onde a “linha central” representa a regressão linear ponderada dos pontos.

3.2.2 – O conjunto dos gráficos:

A análise do conjunto dos gráficos traz mais clareza ao que foi anteriormente escrito como desenvolvimento do modelo. O valor do coeficiente de correlação de Pearson (r) expressa um modo de comparação estatística entre todos os dados. Os maiores (r)s com certeza são os índices de como os dados se relacionam melhor num só conjunto, idem quando o relacionamento é com logS. Se forem considerados todos os índices (r)s maiores, para selecionar aqueles que “somariam” para a regressão, onde estaria o ponto de corte? O nível de relacionamento de um único conjunto de dados permanece o mesmo, aumenta ou diminui quando logS é relacionado com dados de outras propriedades?

O conjunto dos gráficos foi determinante para orientar a estruturação da equação final. Existiu tanto bom senso, tentativa e erro, como uso de ferramentas estatísticas e ainda princípios físico-químicos já conhecidos. Assim sendo foram logicamente desprezados todos os gráficos individuais que sofreram linearização e cujo (r), para o novo gráfico, aumentou. Pelos conhecimentos tradicionais da físico-química ou química orgânica [26] a

polaridade das moléculas, aqui expresso como Momento dipolar necessariamente se relaciona com a solubilidade das mesmas. O que fica evidente é a relevância menor, pelo baixo (r). Necessariamente também o número de doadores e aceptores de pontes de hidrogênio complementam o relacionamento da polaridade. É observável na tabela 1, vários compostos cuja momento dipolar é “zero” porém possuem alguma solubilidade em água devido as formações de pontes de hidrogênio.

Polarizabilidade (α), uma medida que traduz a simulação prática de uma molécula, em determinadas condições, colocada em um campo elétrico [29] e aqui considerado um campo de $0,5\text{eV}$, é a propriedade que melhor se relaciona com $\log S$ porém ainda num nível baixo ($0,458$)⁷, ou seja, a exemplo da polaridade, por si só “não representa solubilidade”.

Por fim, tentativa e erro complementou o desenvolvimento da equação final para o conjunto de todas as propriedades envolvidas, ou seja, trabalhando com os dados que representam melhor correlação de Pearson, incluso os dados das propriedades transformadas, em função da melhor regressão múltipla com $\log S$, foram desenvolvidos no Minitab diversas possibilidades de regressão. A cada cálculo de regressão tentada obteve-se uma equação que expressa o caminho de cálculo para obtenção de $\log S$ em função das propriedades selecionadas.

⁷ Coeficiente de Correlação de Pearson é uma medida que varia de -1 a +1 e pode ser entendida também em porcentagem. -1 ou +1 representa 100%. 0,458 representa um relacionamento de 45,8%.

Capítulo 4

4) Equação geral e validação:

4.1 – Equação geral:

A equação de melhor aproximação para expressar logS a partir das propriedades das 242 moléculas (modelo – Tabela 1), foi⁸:

$$\begin{aligned} \log_{10} S = & -10,7 + 1,15 \log_{10}(HOMO)^2 + 0,0814 MD - 0,221Hdr \\ & + 0,508HAc r + 0,753 (1/PM) 1000 + 1,10(1/\text{Área}) 1000 \\ & + 0,510(1/\alpha)100 - 1,77(1/Vol)1000 + 0,418 LUMO . \end{aligned}$$

Nesta regressão tem-se:

⁹S = 1,19463; R² = 51,3%; R² (ajustado) = 49,4%; P-valor¹⁰ = 0,000.

Aplicada a equação para logS (Tabela 1) obtem-se logScalc com um (r) entre os dois de 0,716 log, bastante superior ao encontrado por CHEN ET AL [30] de 0,56 log para o seu Modelo A, equação geral.

O gráfico abaixo representa a relação logS x logScalculado apresentando sua linha de regressão ponderada praticamente reta, indicando que a linha de regressão linear é bastante semelhante.

A dispersão dos pontos trazem a demonstração de um desvio padrão amostral “grande” mas observa-se uma grande aproximação de pontos da linha de regressão em solubilidades “médias” onde estão a grande maioria dos fármacos (-6 a -1), porém aí também estão as maiores dispersões e não nas extremidades.

⁸ Regressão Linear multivariada (ou múltipla) desenvolvida no MiniTab 15 acadêmico.

⁹ S= Desvio Padrão Amostral.

¹⁰ P-valor – probabilidade de erro no cálculo.

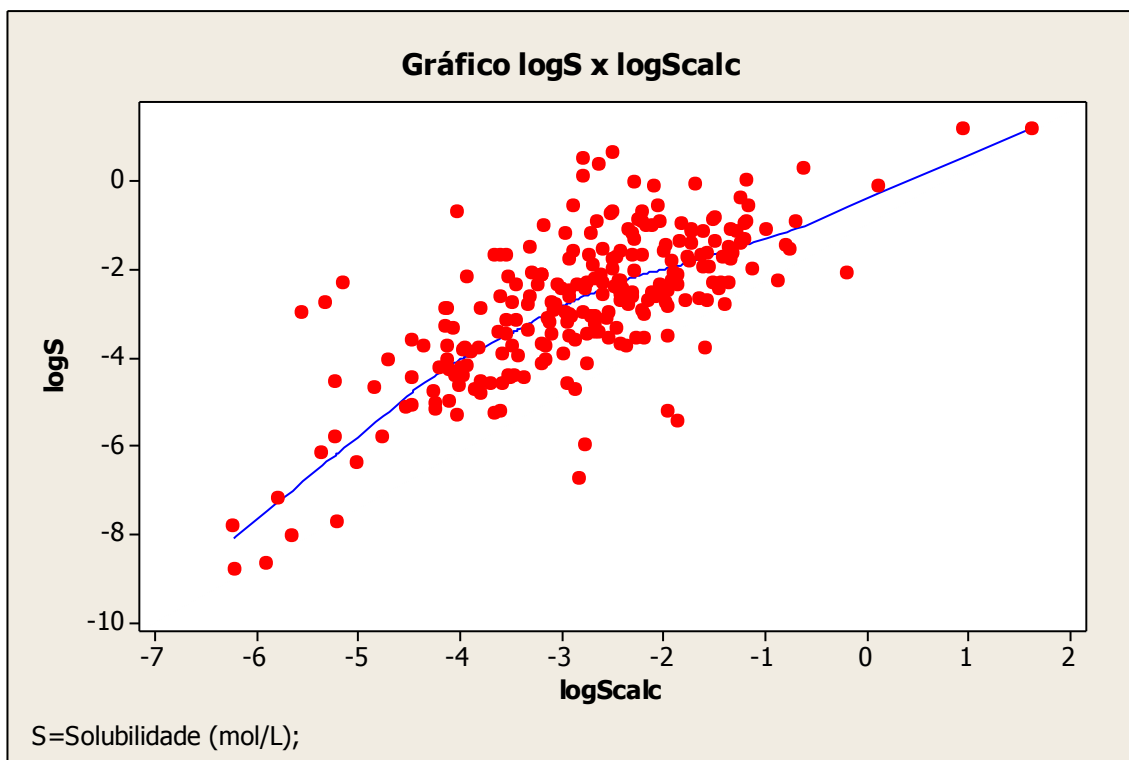


Figura 17 – O gráfico apresenta a relação logS x logScalc onde a “linha central” representa a regressão linear ponderada dos dados.

A equação geral, sem subdivisão do conjunto de 242 moléculas, obtida aqui demonstra uma evolução no trabalho de “QSPR” em relação a essa amplitude dos dados. O trabalho de CHEN ET AL.[30], desenvolvido com as mesmas moléculas, trouxe resultado inferior como já mostrado. No entanto o trabalho deles deu sequência na divisão das suas 267 moléculas de desenvolvimento do modelo em três grupos, (82, 84 e 101 moléculas) através de algoritmo genético, obtendo (r)s de 0,96, 0,93 e 0,93 respectivamente para cada grupo, tornando o cálculo por grupo mais seguro. O trabalho de CHEN ET AL. foi o último publicado desenvolvido nos moldes aproximados deste trabalho, com programas, método semi-empírico de modelagem e propriedades diferentes.

4.2 – Validação

Os dados utilizados para a validação estão na Tabela 2, abaixo. São dados de 80 moléculas extraídos do total das 322 numa escolha sistematizada de que sendo todos colocados em ordem crescente de solubilidade os de validação foram escolhidos 1 em cada 4 seguidamente. Vamos agora analisar o erro cometido pelo modelo proposto, isto é, o que representam os valores de logS experimental menos o logS calculado (Residual), entendendo a expectativa ideal de valor zero.

Tabela 2 – Tabela de dados inerentes às substâncias selecionadas para validação do modelo desenvolvido.

Nome	logS	logCalcV	Residual	PM	AREA	VOLUME	PI	HOMO	LUMO	MD	ALFA	Hdr	Hacr
3-Metilcolantreno	-7,970	-5,9792	-1,9908	268,36	289,76	323,18	7,898137	-7,898	-0,476	0,823	30,6655	0	0
Trifenileno	-6,730	-5,3986	-1,3314	228,29	246,96	268,84	8,498183	-8,498	-0,198	0,000	24,5427	0	0
Fenbuconazol	-6,226	-3,5951	-2,6309	336,82	347,66	404,62	9,377168	-9,377	-0,106	5,592	26,7132	0	3
Danazol	-5,507	-3,9287	-1,5783	337,46	315,90	374,79	8,961238	-8,961	-0,164	3,737	24,8606	1	3
Equilenin	-5,249	-4,2080	-1,0410	266,34	278,07	320,35	8,482982	-8,483	-0,200	2,750	23,5760	1	2
Estriol	-4,955	-3,7256	-1,2294	288,39	284,71	329,26	8,631287	-8,631	0,619	2,741	21,4320	3	3
Fluorene	-4,920	-4,0460	-0,8740	166,22	198,92	204,61	8,557149	-8,557	0,031	0,372	16,6832	0	0
Indoprofeno	-4,824	-3,7070	-1,1170	281,31	304,38	336,59	8,679600	-8,680	0,195	3,287	23,7181	1	3
Diidroequilenin	-4,642	-3,6140	-1,0280	268,35	278,88	327,98	8,735220	-8,735	0,489	3,233	22,1280	1	2
17Alfa-etinilestradiol	-4,484	-4,3287	-0,1553	296,41	298,56	346,91	8,730860	-8,731	0,536	1,639	23,6901	2	2
Diidroequilin	-4,402	-3,8871	-0,5149	270,37	281,93	332,38	8,615458	-8,615	0,596	2,252	22,2337	2	2
Proclorperazina	-4,398	-4,2104	-0,1876	373,94	339,17	420,24	7,713784	-7,714	0,239	1,326	30,5027	0	3
Difenil	-4,340	-3,4886	-0,8514	154,21	207,66	215,72	8,802082	-8,802	-0,259	1,595	16,7638	0	0
Naproxeno	-4,155	-3,4961	-0,6589	230,26	258,58	275,37	8,564367	-8,564	-0,196	2,593	20,1572	1	3
Perfenazina	-4,155	-4,2256	0,0706	403,97	384,44	474,16	7,974793	-7,975	-0,120	1,667	33,7409	1	4
Testosterona	-4,070	-3,3021	-0,7679	288,43	295,70	367,25	9,849744	-9,850	0,400	4,264	22,2447	1	2

Oxazepam	-3,952	-4,3491	0,3971	286,72	277,74	313,09	9,163613	-9,164	-0,541	2,498	23,7712	2	3
5-Etil-5-octilbarbiturato	-3,943	-3,2962	-0,6468	268,36	315,84	346,65	11,065069	-11,065	-0,077	1,597	19,3170	2	3
Tenoxicam	-3,875	-3,3471	-0,5279	337,38	307,83	347,63	9,037456	-9,037	-1,496	2,244	28,8298	2	7
Ácido Mefenamico	-3,770	-3,7194	-0,0506	241,29	264,15	294,27	8,297186	-8,297	-0,048	3,244	22,4577	2	3
Linuron	-3,521	-3,7669	0,2459	249,10	271,24	286,65	9,453993	-9,454	-0,790	2,167	16,2192	1	2
Atrazina	-3,489	-2,1904	-1,2986	215,69	231,60	234,04	9,337321	-9,337	0,044	3,716	16,0310	2	5
Fluometuron	-3,463	-2,4193	-1,0437	232,20	243,51	251,60	9,018234	-9,018	-0,075	4,426	15,8444	1	4
Deoxicorticosterona	-3,450	-3,5026	0,0526	330,47	329,76	415,47	9,973500	-9,974	0,286	0,885	24,8616	1	3
Azatioprina	-3,443	-2,5152	-0,9278	277,26	259,98	286,82	9,697032	-9,697	-1,112	5,821	22,1388	1	6
Isoguanina	-3,401	-1,7517	-1,6493	151,13	163,80	158,44	8,997591	-8,998	-0,370	8,659	11,8385	3	4
Stricnina	-3,330	-3,4076	0,0776	334,42	304,02	384,04	9,025694	-9,026	0,493	2,392	26,7151	0	3
Cortisona	-3,270	-2,5198	-0,7502	360,45	337,70	429,54	9,9855	-9,985	0,337	5,218	25,5968	2	5
Prednisolona	-3,180	-2,5748	-0,6052	360,45	337,99	433,65	9,271698	-9,272	0,934	4,894	25,8498	3	5
Ciclohexano-spirobarbiturato	-3,168	-2,3020	-0,8660	210,23	213,57	239,94	10,927137	-10,927	-0,080	1,683	14,7043	2	3
Morfina	-3,154	-2,9415	-0,2125	285,34	268,57	325,95	8,640620	-8,641	0,375	3,087	22,1554	2	4
Norfloxacin	-3,057	-2,1116	-0,9454	319,33	313,36	364,34	8,743064	-8,743	-0,514	7,259	26,5317	2	7
Hidroflumetiazida	-3,043	-0,3538	-2,6892	331,28	259,79	289,89	10,446622	-10,447	-1,466	9,979	17,0877	3	10
Clortetraciclina	-2,940	-2,7007	-0,2393	478,89	396,39	498,87	9,249197	-9,249	-0,954	2,147	35,2230	6	10
Sulpirida	-2,876	-2,5780	-0,2980	341,42	330,26	400,62	9,133806	-9,134	0,001	1,360	24,5512	2	6
Sulfatiazol	-2,805	-2,1960	-0,6090	255,31	240,34	268,17	9,189481	-9,189	-0,541	8,635	18,2939	2	5
Epinefrina	-2,740	-2,1789	-0,5611	183,21	217,61	220,64	8,718742	-8,719	0,428	3,189	14,0547	4	4
Fenilbutazona	-2,644	-4,4182	1,7742	308,38	328,42	380,00	9,056984	-9,057	-0,200	1,567	27,0193	0	2
Pteridina-2-tiol	-2,629	-3,4564	0,8274	164,18	176,36	174,71	9,4392	-9,439	-1,294	2,935	14,4532	2	3
Cefazolina	-2,616	-1,8085	-0,8075	454,50	398,32	488,85	9,362355	-9,362	-0,802	7,209	34,5017	2	9
5_5-Dipropilbarbiturato	-2,527	-2,4713	-0,0557	212,25	235,67	254,08	11,061028	-11,061	-0,093	1,697	15,0099	2	3
Xantina	-2,483	-2,5050	0,0220	152,11	159,74	153,08	9,377155	-9,377	-0,368	3,942	10,9507	3	3

Adenina	-2,432	-1,4165	-1,0155	135,13	153,87	147,45	8,886783	-8,887	-0,066	2,223	10,9896	2	4
Pteridina-4-metil-tiol	-2,365	-2,5122	0,1472	178,21	188,82	192,30	9,167728	-9,168	-1,273	4,102	15,4822	0	4
Glutetimida	-2,337	-2,7793	0,4423	217,27	234,37	264,43	9,678035	-9,678	0,257	3,673	17,0772	1	2
Secbutabarbitol	-2,333	-2,1375	-0,1955	212,25	225,93	256,15	11,083425	-11,083	-0,054	1,745	14,7746	2	3
2-Aminopteridina	-2,298	-1,8755	-0,4225	147,14	165,36	160,44	9,195457	-9,195	-0,877	0,756	13,2233	1	5
Ciclobarbitol	-2,273	-2,8734	0,6004	236,27	243,12	279,41	9,711245	-9,711	-0,034	1,986	17,2900	2	3
Ácido 5-Aminosalicílico	-2,259	-2,0604	-0,1986	153,14	170,84	167,28	8,058154	-8,058	-0,314	3,719	12,0283	3	4
Idobutal	-2,172	-2,2012	0,0292	224,26	246,65	276,05	10,271094	-10,271	-0,122	1,672	16,1528	0	3
7-Hidroxipteridina	-2,124	-1,4145	-0,7095	148,12	161,69	157,37	10,168779	-10,169	-1,318	3,893	12,0160	1	5
Butalbital	-2,119	-2,5842	0,4652	224,26	239,11	272,92	10,185651	-10,186	-0,140	1,848	16,1863	2	3
Minoxidil	-1,978	-1,7574	-0,2206	209,25	235,78	248,16	8,523272	-8,523	0,135	5,882	18,5417	3	6
Ciproheptadina	-1,898	-4,7687	2,8707	287,40	307,91	365,39	8,538689	-8,539	0,118	1,170	28,5210	0	1
Cicloethano-spirobarbiturato	-1,886	-1,4154	-0,4706	154,13	165,79	165,36	11,297229	-11,297	-0,207	1,694	10,3808	2	3
Meprobamato	-1,807	-1,0475	-0,7595	218,25	253,59	272,03	10,822055	-10,822	1,197	3,852	14,2123	2	4
Lidocaina	-1,768	-3,0133	1,2453	234,34	279,89	317,64	8,813650	-8,814	0,614	4,148	19,9807	1	2
Pteridina-2-metil-tiol	-1,754	-2,8494	1,0954	178,21	193,65	193,23	9,063779	-9,064	-1,197	2,619	16,1481	0	4
7-Butil-8-metilteofilina	-1,745	-3,0215	1,2765	250,30	280,82	306,00	8,776708	-8,777	-0,046	4,322	20,3105	0	3
5_5-Dimetilbarbiturato	-1,742	-1,1063	-0,6357	156,14	170,58	174,22	11,244386	-11,244	-0,130	1,284	10,3389	2	3
1-Butilteobromina	-1,625	-2,8831	1,2581	236,27	278,29	299,11	8,911863	-8,912	-0,119	3,632	19,3190	0	3
Sulfacetamida	-1,507	-2,0143	0,5073	214,24	224,62	229,73	9,5228	-9,523	-0,457	10,944	14,8365	2	4
4-Hidroxipteridina	-1,471	-1,3389	-0,1321	148,12	159,81	155,54	10,247350	-10,247	-1,247	3,962	11,7802	1	5
Isopropilbarbiturato	-1,456	-1,3195	-0,1365	170,17	185,18	196,37	11,232619	-11,233	-0,138	1,408	11,3996	2	3
Salbutamol	-1,224	-2,7528	1,5288	239,31	283,08	312,58	8,866164	-8,866	0,477	3,508	18,4470	4	4
Metronidazol	-1,212	-1,4050	0,1930	171,16	188,41	194,78	9,976591	-9,977	-0,958	3,654	12,3226	1	4
Citosina	-1,159	-0,7624	-0,3966	111,10	135,84	124,65	9,395555	-9,396	0,015	6,274	8,5957	2	2
Acetaminofen	-1,074	-2,1767	1,1027	151,16	182,44	183,10	8,843680	-8,844	0,146	3,442	12,4793	2	2

Fluorouracil	-1,028	-1,0490	0,0210	130,08	139,02	129,35	9,856279	-9,856	-0,544	3,743	8,2408	2	3
Cafeina	-0,951	-2,3416	1,3906	194,19	209,50	219,37	8,929403	-8,929	-0,127	3,638	15,2475	0	3
7-Isobutilteofilina	-0,942	-2,7254	1,7834	236,27	257,41	286,10	8,8804	-8,880	-0,093	3,841	18,7379	0	3
Ácido Nicotínico	-0,850	-0,6267	-0,2233	123,11	146,95	139,65	10,305953	-10,306	-0,661	1,053	8,9423	1	3
Aminopirina	-0,619	-2,8829	2,2639	231,30	261,44	291,08	8,488619	-8,489	0,043	4,420	21,2786	0	3
5-Metilcitosina	-0,560	0,7477	-1,3077	125,13	159,19	160,16	9,7821	-9,782	1,483	5,027	8,9625	2	2
4-Metilpteridina	-0,466	-1,4862	1,0202	146,15	168,95	166,08	9,994756	-9,995	-1,148	3,024	12,4074	0	4
Cicloserina	-0,009	2,0871	-2,0961	102,09	125,29	113,51	9,726789	-9,727	0,707	2,926	6,1174	2	3
Isoniazida	0,009	-0,9572	0,9662	137,14	165,69	160,57	9,847347	-9,847	0,019	4,232	10,3714	2	3
Difilina	0,118	-2,3495	2,4675	254,25	264,01	286,53	8,803728	-8,804	-0,059	5,453	18,8214	2	5
Nicotinamida	0,913	-1,0114	1,9244	122,13	150,30	143,42	10,224855	-10,225	-0,473	1,464	9,2609	1	2
Cloral hidratado	1,700	-0,0274	1,7274	165,40	147,04	146,11	11,250475	-11,250	-0,844	4,081	6,5535	2	2

Notas 2- legenda para o cabeçalho da Tabela 2

log S – logaritmo da solubilidade das substâncias em água (S em mol/L), obtido da literatura [11];

logScalV – logaritmo decimal da solubilidade calculado a partir da equação geral obtida no item 4.1

Residual= logS – logScalV

PM- Peso molecular (em g/mol), obtido da literatura [11];

VOL* – volume das moléculas (A^3); AREA* – área das moléculas (A^2); PI*- potencial de ionização (EV);

HOMO* – “High occupied molecular orbital” (EV); LUMO* – “Lowest unoccupied molecular orbital” (EV);

MD* – momento dipolar (debye); ALFA* – polarizabilidade (alfa) (a 0,5 EV; em ANG.³)

Hdr – Número de Hidrogênios doadores de pontes de Hidrogênio em cada molécula.

HAc- Número de átomos aceptores de pontes de hidrogênio em cada molécula.

*(Todos os dados com cálculos desenvolvidos pelo MOPAC a partir dos parâmetros das moléculas desenhadas no HyperChem).

O erro médio cometido pela previsão é -0,1233, indicando um valor de $\log S_{\text{calculado}}$ abaixo do valor de $\log S_{\text{experimental}}$. Isso representa percentualmente um erro médio de 4,5%, o que pode se considerar aceitável do ponto de vista prático. Já o desvio padrão do erro associado é de 1.11 unidades de $\log S$, que pode ser considerado relativamente elevado.

Para a validação em si, foi considerado um teste de hipótese, t-student, para amostras pareadas, entre o $\log S$ calculado e o $\log S$ experimental. No teste consideramos um intervalo de confiança de 95%. Os resultados, apresentados na tabela abaixo, indicam p-valores maiores que 5%, mostrando não haverem indícios de diferenças estatisticamente significativas entre o $\log S$ calculado e o $\log S$ experimental. Além disso revela um alto grau de correlação, $(r) = 75\%$. Isso evidencia o potencial de previsão do modelo.

Tabela 3: Teste de hipótese, t-student pareado, para validação do modelo. Mostra não haver indícios de diferença estatisticamente significativa entre o logS calculado e o logS experimental.

<i>Teste t-Student Pareado</i>	<i>logS experimental</i>	<i>logS Calculado</i>
Média	-2.6812375	-2.5579025
Variância	2.812653728	1.760758861
N	80	80
Correlação de Pearson	0.750677135	
Hipótese testada	0	(Igualdade)
Graus de liberdade	79	
t Stat	- 0.993741228	
P(T<=t) unicaudal	0.161691472	
t Critico unicaudal	1.66437141	
P(T<=t) bicaudal	0.323382944	
t Critico bicaudal	1.990450177	

Apesar do modelo ter um comportamento aceitável na previsão da solubilidade, a literatura [30] mostra que há mecanismos de melhorá-lo, conforme descrito antes, possivelmente com divisões de sub-grupos das moléculas. A divisão em sub-grupos deve elevar a precisão da previsão da solubilidade mas por outro lado cria uma dificuldade em termos de classificação a priori em qual sub-grupo se encontra a molécula, o que por si só já é passível de erros. Outro aspecto importante é que a divisão dos sub-grupos dá-se simplesmente por uma questão matemática, dissociada das propriedades das moléculas em si, o que seria desejável. Por esses motivos, torna-se mais interessante que se obtenha uma única equação, como proposto e concretizado neste trabalho, pois assim temos um modelo universal.

Capítulo 5

Considerações finais

O propósito deste trabalho não foi explicar os mecanismos que envolvem a solubilidade de moléculas diversas em água mas confirmam-se conceitos já trabalhados de que estes mecanismos são multifatoriais [2]. A idéia simplificada de solubilidade, já exposta neste trabalho, de que precisamos ter uma dissociação de uma molécula, A-A, para interagir com água, B-B, e formar A-B envolve não “uma” propriedade específica das moléculas mas propriedades e que, mesmo assim, essas propriedades não explicam tudo. É possível através das propriedades, tradicionalmente vistas na literatura [29], que se relacionam com solubilidade, calcular-se uma previsão de solubilidade em água a partir de sua estrutura molecular, com erro estatisticamente aceitável.

O conceito envolvendo entropia da mistura, também embora não tenha sido propósito deste trabalho, não ganha maior esclarecimento. Não foram trabalhadas variações envolvendo “quantidade de calor” e temperatura. Este trabalho não envolveu dinâmica e sim trabalhou propriedades “estáticas”, propriedades inerentes a moléculas isoladas. Não foi visto, por exemplo, a energia envolvida na formação do cristal para cada sólido formado.

Ao que os dados indicam todas as propriedades trabalhadas estão bem envolvidas no processo de solubilidade mas o que é mostrado também, de que a média da solubilidade calculada é inferior a solubilidade experimental, é de que falta alguma “coisa” para complementar. Se esse resultado fosse inverso, seria pior. Sugeriria que alguma propriedade estaria atrapalhando, a mais.

Sem aprofundar a questão da química quântica que está envolvida no método semi-empírico RM1 e somente abstraindo de que a química quântica não foca predominantemente partes das moléculas e sim as moléculas como um todo, parece relevante o que este trabalho conseguiu de exprimir a previsão de solubilidade *in silico* numa única equação para um conjunto de moléculas pretensamente universais como orgânicas e virtualmente não ionizáveis.

5) Referências bibliográficas

- 1) Aulton, M. E.; *Pharmaceutics: The Science of Dosage Form Design, Harcourt Publishers Limited, Second Edition, Leicester, UK, pp. 2-11 e 16, 2002.*
- 2) Yalkowsky SH, Valvani, SC.; Solubility and partitioning I: Solubility of nonelectrolytes in water; *J. Pharm Sci* 69:912-922, 1980.
- 3) Klopman, G., Wang, S., Balthasar, DM.; Estimation of aqueous solubility of organic molecules by the group contribution approach. Application to the study of biodegradation. *J Chem Inf Comput Sci* 32:474-482, 1992.
- 4) Myrdal, PB., Manka, AM., Yalkowsky, SH.; AQUAFAC 3: Aqueous functional group activity coefficients; application to the estimation of aqueous solubility. *Chemosphere* 30:1619-1637, 1995.
- 5) Lipinski, CA., Lombardo, F., Dominy, BW., Feeney, PJ.; Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Delivery Rev* 23:3-25, 1997.
- 6) Jorgensen, William L.; Duffy, Erin M.; Prediction of drug solubility from Structure, *Advanced Drug Delivery Reviews*, 54, 355-366, 2002.
- 7) Delaney, John S.; Predicting aqueous solubility from structure, *Drug Discovery Today*, Volume 10, Number 4, February 2005.
- 8) Hansch, C., Muir, R. M., Fujita, T., Maloney, P. P., Geiger, E., Streich, M.; The Correlation of Biological Activity of Plant Growth Regulators and Chloromycetin Derivatives with Hammett Constants and Partition Coefficients. *J. Am. Chem. Soc.* 85, 2817-2824, 1963.
- 9) Hansch, C., Fujita, T. F.; A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* 86, 1616-1626, 1964.
- 10) Hansch, C.; On the Structure of Medicinal Chemistry. *J. Med. Chem.* 19, 1-6, 1976.
- 11) Rytting Erik, Lentz Kimberley A., Chen Xue-Qing, Qian Feng e Venkatesh Sridini; Aqueous and Cosolvent Solubility Data for Drug-like Organic Compounds; *The AAPS Journal* 7 (1) Article 10, 2005.
- 12) Huibers, PD., Katritzky, AR.; Correlation of the aqueous solubility of hydrocarbons and halogenated hydrocarbons with molecular structure. *J Chem Inf Comput Sci* 38:283-292, 1998.
- 13) Nelson, TM., Jurs, PC.; Prediction of aqueous solubility of organic compounds. *J Chem Inf Comput Sci* 34:601-609. 1994
- 14) Huuskonen, J., Salo, M., Taskinen, J.; Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. *J Chem Inf Comput Sci* 38:450-456, 1998.
- 15) Huuskonen, J.; Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *J Chem Inf Comput Sci* 40:773-777, 2000.
- 16) Sutter, JM., Jurs, PC.; Prediction of aqueous solubility for a diverse set of heteroatom-containing organic compounds using a quantitative structure property relationship. *J Chem Inf Comput Sci* 36:100-107, 1996.
- 17) Bodor, N., Huang, M.; A new method for the estimation of the aqueous solubility of organic compounds. *J Pharm Sci* 81:954-960, 1992.

- 18) Mitchell, BE., Jurs, PC.; Prediction of aqueous solubility of organic compounds from molecular structure. *J Chem Inf Comput Sci* 38:489–496, 1998.
- 19) Jain, N., Yalkowsky, SH.; Estimation of the aqueous solubility I: Application to organic nonelectrolytes. *J Pharm Sci* 90:234–252, 2001.
- 20) Fini, A., Fazio, G., Feroci, G.; Solubility and solubilization properties of non-steroidal anti-inflammatory drugs. *Int J Pharm* 126:95–102, 1995.
- 21) Kenneth, C. James; *Solubility and Related Properties*, Marcel Dekker, Inc. New York, New York. 1986.
- 22) United States National Library of Medicine – National Institutes of Health - Division of Specialized Information Services, <http://chem.sis.nlm.nih.gov/chemidplus/>, Ago a Nov 2009
- 23) Sung Jin Cho, Mark, A. Hermsmeier J.; Genetic Algorithm Guided Selection: Variable Selection and Subset Selection; *Chem. Inf. Comput. Sci.* 42, 927-936, 2002.
- 24) Renxiao Wang, Ying Fu, and Luhua Lai; A New Atom-Additive Method for Calculating Partition Coefficients; *J. Chem. Inf. Comput. Sci.*, , 37 (3), pp 615–621, 1997; Institute of Physical Chemistry, Peking University, Beijing 100871, P.R. China
- 25) Morgon, Nelson H., Coutinho, Kaline; Métodos de Química Teórica e Modelagem Molecular, *Editora Livraria da Física, São Paulo, 2007.*
- 26) Maitland Jones, Jr.; Organic Chemistry, *W.W. Norton & Company, Second Edition, New York, 2000, pp 6-7.*
- 27) Rogers D, Hopfinger AJ.; Application of genetic function approximation to quantitative structure activity relation and quantitative structure property relationships. *J Chem Inf Comput Sci* 34:854–866, 1994
- 28) Abraham MH, Le J.; The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. *J Pharm Sci* 88:868–880, 1999.
- 29) Martin, Alfred; *Physical Pharmacy, Lippincott Williams & Wilkins, Philadelphia, 1993, pp.86-87.*
- 30) Chen, Xue-Qing, Cho, Jin Sung, Li, Yi, Venkatesh, Srin; Prediction of Aqueous Solubility of Organic Compounds Using a Quantitative Structure-Property Relationship. *J. of Pharmaceutical Sciences*, vol, 91, no. 8, August 2002.
- 31) Pagano, Marcello, Gauvreau, Kimberlee; princípios de Bio estatística, tradução da 2ª Edição Norte-Americana, Thomson, Brasil, 2004.
- 32) Martins, Gilberto Andrade; Estatística Geral e Aplicada, 3ª Edição, São Paulo, Editora Atlas S.A, 2008.
- 33) Sartorius, Alexandre; Estatística e introdução à econometria. São Paulo, Saraiva, 2003