7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

Description of the cat-SAR® Expert System

Below is a summary of the cat-SAR expert system followed by a listing of models currently available or under development. Leave-one-out validation (LOO) results are included as sensitivity, specificity, and observed correct predictions (OCP). The models listed were parameterized to return models with near-equal sensitivity and specificity. Other parameters could be used to develop models with potentially higher OCP values or higher specificity values (useful for risk aversion).

Briefly, structure-activity relationship (SAR) models are powerful tools to investigate the relationship between the chemical structure of a compound and its biological or pharmacological activity. The categorical-SAR program— cat-SAR®—is a computational SAR or *in silico* toxicity analysis and prediction "expert system". Cat-SAR® stands alone from other computerized SAR expert systems in its openness, flexibility, routine for identifying important attributes of biological activity or inactivity and its method for predicting the activity of untested compounds.

The cat-SAR® system is completely open with every detail of modeling transparent to the user. Gnarus does not employ "black box" technology. The cat-SAR® approach allows the user to select and/or adjust many parameters during the modeling process from learning set makeup, to selection of types of fragment attributes to consider, to ultimately what numerical or statistical considerations are employed in developing the final model.

The cat-SAR® approach is also a very data- and information-intensive SAR expert system. During model development and the creation of the final model, all fragments associated with the categories are presented. This leaves the user with an unbiased view of all chemical fragment features associated with the biological endpoint.

The approach we have taken in developing cat-SAR® clearly diverges from existing SAR expert systems and is more in tune with modern QSAR techniques. For instance, the user is presented with a number of selectable and adjustable modeling parameters. The notion of having selectable and adjustable modeling parameters facilitates the ability to explore rigorously the relationships between chemical structure and biological activity.

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

About Gnarus Systems

Gnarus Systems, Inc. was organized in May 2009 by Al Cunningham, Ph.D. and Suzanne Cunningham, M.S., in conjunction with MetaCyte Business Lab LLC.

Management Team

Albert R. Cunningham, Ph.D. is Founder and President of Gnarus Systems. He is an Associate Professor of Medicine and Pharmacology & Toxicology at the University of Louisville and a Scientist at the University's James Graham Brown Cancer Center. Dr. Cunningham holds a B.S. in Biology and a B.A. in Philosophy from Slippery Rock University of Pennsylvania. He received his Ph.D. in Environmental and Occupational Health from the University of Pittsburgh's Graduate School of Public Health.

Suzanne Cunningham, M.S. is Founder and Vice President of Gnarus Systems. She is the original author of the cat-SAR program. Suzanne earned a B.S. in Mathematics, Specialization in Computer Science, from Slippery Rock University of Pennsylvania and holds a M.S. in Biostatistics from the University of Pittsburgh's Graduate School of Public Health.

Andrew Steen, Vice President of Business Development, also serves in the same capacity for MetaCyte Business Lab LLC. He has been involved directly and indirectly in the creation of several life science start-up companies. Andrew holds a B.A. from Transylvania University and an M.B.A. from the University of Louisville.

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

The cat-SAR® Models

Below is a listing and summary of the cat-SAR models currently available or under development. Leave-one-out validation (LOO) results are included as sensitivity, specificity, and observed correct predictions (OCP). The models listed were parameterized to return models with near-equal sensitivity and specificity. Other parameters could be used to develop models with potentially higher OCP values or higher specificity values (useful for risk aversion).

MUTAGENESIS/GENOTOXICITY
Salmonella, NTP
Salmonella, Benchmark*
Micronucleus, Rosenkranz
Micronucleus, NTP

CARCINOGENESIS
Mouse Carcinogenesis
Rat Carcinogenesis
Non-Mutagenic Carcinogenesis

DEVELOPMENTAL TOXICITY
Human Teratogenicity
Mouse Developmental
Rat Developmental
Rabbit Developmental
Human Development

SENSITIZATION AND IRRITATION
Allergic Contact Dermatitis: Marginal as positive
Allergic Contact Dermatitis: Marginal as negative
Mouse Local Lymph Node
Ocular Irritation, Rosenkranz
Respiratory Sensitization, Karol
Corrosion (Severe Skin Burns and Eye Damage)*
Skin Irritation*
Serious Eye Damage*
Serious Eye Irritation*

ENDOCRINE DISRUPTORS/ESTROGEN MIMICS
ESCREEN, MCF-7
FDA NCTR Estrogen Receptor

*Under Development

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

MUTAGENICITY/GENOTOXICITY

- **NTP *Salmonella* Mutagenicity:** This model is based on data from the National Toxicology Program (NTP)[1]. It is a compilation of test results from different TA strains (with and without S9 activation). The model is developed from 1678 chemicals, 573 of which were classified as mutagens and 1105 as non-mutagens.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
1678 (573/1105)	0.79 (377/477)	0.79 (629/797)	0.79 (1006/1274)

- **Ames Mutagenicity Benchmark Dataset:** This model is based on a dataset of 6512 compounds published by Hansen et al.[2]. It is a compilation of published *Salmonella* mutagenicity results with 3503 mutagens and 3009 non-mutagens. This dataset is currently being validated.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
6512 (3503/3009)	TBD	TBD	TBD

- **Induction of Micronuclei, *in vivo* (MNT):** Two models have been developed for *in vivo* MNT based on a dataset developed by Dr. Herbert Rosenkranz[3] and another obtained from NTP data.

  - **MNT Rosenkranz:** The model based on the Rosenkranz dataset consists of 238 chemicals, 157 of which were classified as capable of inducing micronuclei and 81 inactive**.**

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
236 (155/81)	0.78 (117/150)	0.71 (51/72)	0.76 (168/222)

  - **MNT NTP:** The model is based on data obtained from the NTP website and consists of 185 compounds tested in mice (with analyses in bone marrow or peripheral blood), 86 of which are classified as capable of inducing micronuclei and 99 as inactive.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
185 (86/99)	0.72 (43/60)	0.71 (47/66)	0.71 (90/126)

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

CARCINOGENESIS

- **Carcinogenic Potency Database Models:** The Carcinogenic Potency Database (CPDB) analyzes and consolidates into a single resource the world's diverse literature and NTP Technical Reports of chronic long-term animal cancer bioassays [4]. To date, analyses of 6540 experiments on 1547 chemicals are available on the CPDB [5] and the EPA DSSTox [6]websites. Separate rat and mouse CPDB carcinogenesis models have been developed from these sources.

    - **Mouse Cancer:** This model consists of 816 chemicals, 400 of which are categorized as mouse carcinogens and 416 as non-carcinogens.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
816 (400/416)	0.68 (272/398)	0.69 (258/373)	0.69 (530/771)

    - **Rat Cancer:** This model consists of 924 chemicals, 531 of which are categorized as rat carcinogens and 393 as non-carcinogens.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
924 (531/393)	0.70 (343/490)	0.69 (250/362)	0.70 (593/852)

    - **Non-Mutagenic Rat Cancer:** This model consists of chemicals that were tested and determined not to be Salmonella mutagens or that did not have structural alerts for mutagenicity. This dataset was originally published by Malacarne[7] and recently modeled by us[8].

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
390 (145/204)	0.70 (343/490)	0.69 (250/362)	0.70 (593/852)

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

DEVELOPMENTAL TOXICITY (updated 7/14/14)

- **Human Teratogenicity (*i.e.*, developmental toxicity):** This model is derived from a dataset of compounds analyzed for potential human teratogenicity developed by Dr. Donald Mattison [9]. The model contains developmental toxicity calls for 323 chemicals, 130 of which are classified as human teratogens and 193 as non-teratogens.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
323 (130/193)	0.80 (93/117)	0.78 (142/184)	0.78 (235/301)

Gnarus fragments

- **Mammalian Developmental Toxicity:** Models for mice, rats, guinea pigs, and humans have been derived using data from a set of SAR models originally published by Takihi et al [10] (w/ Mattison and Rosenkranz). From the original publication, "Although there is no generally recognized peer-reviewed authoritative compilation of developmental toxicants that have been tested under quality-controlled rigorous experimental conditions, Jelovsek and associates [11] have assembled a database of approximately 125 chemicals that have been tested in mice, rats, and rabbits and for which there is evidence (or a lack thereof) of developmental toxicity in humans." Although that database is not quantitative, the authors do list the strength of the evidence for developmental toxicity [10].

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
Mouse 74 (37/37)	0.63 (22/35)	0.66 (23/35)	0.64 (45/70)
Rat 134 (82/52)	0.69 (56/81)	0.64 (30/47)	0.67 (86/128)
Rabbit 66 (26/40)	0.64 (16/25)	0.63 (24/38)	0.64 (40/63)
Human 119 (49/70)	0.90 (26/29)	0.89 (75/84)	0.89 (101/113)

Gnarus fragments

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

SENSITIZATION AND IRRITATION-1

- **Allergic Contact Dermatitis (ACD):** The model is based on chemical allergenicity data compiled by Dr. Meryl Karol [12]. Two version of the model have been developed, one in which compounds determined to be marginal allergens were classified as positive allergens (ACD+M) and the other where marginal allergens were classified as non-allergens (ACD-M).

  - **ACD+M** model consists of 767 compounds, 355 of which are classified as positive allergens and 412 as non-allergens.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
767 (355/412)	0.82 (261/320)	0.82 (300/367)	0.82 (561/687)

  - **ACD-M** model consists of 767 compounds, 326 of which are classified as positive allergens and 441 as non-allergens.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
767 (326/441)	0.81 (228/280)	0.82 (279/344)	0.82 (507/624)

Local Lymph Node Assay (LLNA): This model is based on compounds with skin sensitization potential compiled and published by Hopfinger, Gerberick et. al [13-15]. The latest model is based on data for 229 compounds, 63 of which tested positive and 164 tested negative for skin sensitization potential.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
229 (63/164)	0.75 (42/56)	0.76 (118/155)	0.76 (160/211)

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

SENSITIZATION AND IRRITATION-2

- **Ocular Irritation**: This model is based on a published dataset of chemicals tested for eye irritation compiled by Dr. Herbert Rosenkranz [16]. The model is based ocular irritation data from 297 chemicals, 147 of which are classified as ocular irritants and 150 as non-irritants.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
297 (147/150)	0.79 (53/67)	0.74 (68/92)	0.76 (121/159)

- **Respiratory Sensitization:** This model is based on a dataset originally published by Graham et al.[17] and subsequently modeled by us[18]. The dataset consists of 40 documented human respiratory sensitizers and 40 chemically similar presumed non-sensitizers based on negative human skin sensitization results.

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
80 (40/40)	0.94 (34/36)	0.87 (34/39)	0.91 (68/75)

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

SENSITIZATION AND IRRITATION-3

- **Corrosion (Severe Skin Burns and Eye Damage) H314:** Liew and Yap dataset of 2108 tested compounds. "Causes severe skin burns and eye damage (corrosion); the production of irreversible damage to the skin; namely, visible necrosis through the epidermis and into the dermis, following the application of a test substance for up to 4 hours. Corrosive reactions are typified by ulcers, bleeding, bloody scabs, and, by the end of observation at 14 days, by discoloration due to blanching of the skin, complete areas of alopecia, and scars[19]."

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
2108 (220/1888)	TBD	TBD	TBD

- **Skin Irritation H315:** Liew and Yap dataset of 2108 tested compounds. "Causes skin irritation; the production of reversible damage to the skin following the application of a test substance for up to 4 hours[19]."

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
2108 (350/1758)	TBD	TBD	TBD

- **Serious Eye Damage H318:** Liew and Yap dataset of 2108 tested compounds. "Causes serious eye damage; the production of tissue damage in the eye, or serious physical decay of vision, following application of a test substance to the anterior surface of the eye, which is not fully reversible within 21 days of application[19]."

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
2108 (239/1869)	TBD	TBD	TBD

- **Serious Eye Irritation H319:** Liew and Yap dataset of 2108 tested compounds. "Causes serious eye irritation; the production of changes in the eye following the application of test substance to the anterior surface of the eye, which are fully reversible within 21 days of application[19]."

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
2108 (291/1817)	TBD	TBD	TBD

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

ENDOCRINE DISRUPTORS/ESTROGEN MIMICS

- **ESCREEN, MCF-7 Relative Proliferate Effect (RPE):** The ESCREEN relative proliferative effect (RPE) model for estrogen-like activity is based on a set of 122 chemicals tested and published by Soto and colleagues [20-22]. The RPE learning set consisted of 73 active and 49 inactive chemicals. Cat-SAR models based on these data have been published [23].

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
122 (73/49)	0.87 (61/70)	0.89 (39/44)	0.88 (100/114)

- **FDA National Center for Toxicological Research Estrogen Receptor Binding (NCTER ER):** This model is based on estrogen receptor binding data developed by the National Center for Toxicological Research[24, 25]. The data for this model was obtained from the EPA's Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network [6] and consisted of 323 compounds, of which 139 are designated as ligands and 93 as non-ligands (note: eight marginal compounds were classified as ligands for this model).

Model and LOO Validation Data

Compounds (A/I)	Sensitivity	Specificity	Concordance
232 (139/93)	0.92 (126/137)	0.87 (75/86)	0.90 (201/223)

gnarus systems

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

REFERENCES

1. NTP, *Results, Status and Publication Information On All NTP Chemicals*, http://www.predictive-toxicology.org/data/ntp/original_ntp_data.txt.

2. K. Hansen, S. Mika, T. Schroeter, A. Sutter, A. ter Laak, T. Steger-Hartmann, N. Heinrich, and K.-R. Müller, *Benchmark Data Set for in Silico Prediction of Ames Mutagenicity*, J. Chem. Inf. Model. 49 (2009), pp. 2077-2081.

3. W.-L. Yang, G. Klopman, and H. S. Rosenkranz, *Structural basis of the in vivo induction of micronuclei*, Mutat. Res. 272 (1992), pp. 111-124.

4. L. S. Gold, T. H. Slone, and B. N. Ames, *Overview and update of analyses of the carcinogenic potency database*, In *Handbook of Carcinogenic Potency and Genotoxicity Databases*, (L. S. Gold and E. Zeiger, Eds.), CRC Press, New York, 1997, pp. 661-693.

5. CPDB, *Carcinogenic Potency Database*, http://potency.berkeley.edu.

6. DSSTox, *Distributed Structure-Searchable Toxicity (DSSTox) Database Network* http://www.epa.gov/NCCT/dsstox/index.html.

7. D. Malacarne, M. Taningher, R. Pesenti, M. Paolucci, A. Perrotta, and S. Parodi, *Molecular fragments associated with non-genotoxic carcinogens, as detected using a software program based on graph theory: their usefulness to predict carcinogenicity*, Chem. Biol. Interact. 97 (1995), pp. 75-100.

8. A. R. Cunningham, C. A. Carrasquer, S. Qamar, J. M. Maguire, S. L. Cunningham, and J. O. Trent, *Global structure-activity relationship model for non-mutagenic carcinogens using virtual ligand protein interactions as model descriptors*, Carcinogenesis doi: 10.1093/carcin/bgs197 (2012), pp. 1940-1950.

9. M. Ghanooni, D. R. Mattison, Y. P. Zhang, O. T. Macina, H. S. Rosenkranz, and G. Klopman, *Structural determinants associated with risk of human developmental toxicity*, Am. J. Obstet. Gynecol. 176 (1997), pp. 799-806.

10. N. Takihi, H. S. Rosenkranz, G. Klopman, and D. R. Mattison, *Structural determinants of developmental toxicity*, Risk Analysis 14 (1994), pp. 649-657.

11. F. R. Jelovsek, D. R. Mattison, and J. J. Chen, *Predicition of risk for human developmental toxicity: How important are animal studies for hazard identification*, Obstet. Gynecol. 74 (1989), pp. 624-636.

12. C. Graham, R. Gealy, O. T. Macina, M. H. Karol, and H. S. Rosenkranz, *QSAR for allergic contact dermatitis*, Quant. Struct. Act. Relat. 15 (1996), pp. 224-229.

13. Y. Li, Y. J. Tseng, D. Pan, J. Liu, P. S. Kern, G. F. Gerberick, and A. J. Hopfinger, *4D-Fingerprint Categorical QSAR Models for Skin Sensitization Based on the Classification of Local Lymph Node Assay Measures*, Chem. Res. Toxicol. 20 (2007), pp. 114-128.

7403 Abbott Glen
Crestwood, KY 40014
(502) 851-1024
arc@gnarus-systems.com

14. Y. Li, D. Pan, J. Liu, P. S. Kern, G. F. Gerberick, A. J. Hopfinger, and Y. J. Tseng, *Categorical QSAR Models for Skin Sensitization based upon Local Lymph Node Assay Classification Measures Part 2: 4D-Fingerprint Three-State and Two-2-State Logistic Regression Models*, Toxicol. Sci. 99 (2007), pp. 532-544.

15. P. S. Kern, G. F. Gerberick, C. A. Ryan, I. Kimber, A. Aptula, and D. A. Baskettel, *Local Lymph Node Data for the Evaluation of Skin Sensitization Alternatives: A Second Compilation*, Dermatits 21 (2010), pp. 8-32.

16. H. S. Rosenkranz, Y. P. Zhang, and G. Klopman, *The development and characterization of an SAR model of the Draize eye irritation test*, ATLA 20 (1999), pp. 779-809.

17. C. Graham, H. S. Rosenkranz, and M. H. Karol, *Structure-activity model of chemicals that cause human respiratory sensitization*, Regul. Toxicol. Pharmacol. 26 (1997), pp. 296-306.

18. A. R. Cunningham, S. L. Cunningham, D. M. Consoer, S. T. Moss, and M. H. Karol, *Development of an information-intensive structure-activity relationship model and its application to human respiratory chemical sensitizers*, SAR QSAR Environ. Res. 16 (2005), pp. 273-285.

19. C. Y. Liew and C. W. Yap, *QSAR and Predictors of Eye and Skin Effects*, Molecular Informatics 32 (2013), pp. 281-290.

20. C. Sonnenschein, A. M. Soto, M. F. Fernandez, N. Olea, M. F. Olea-Serrano, and M. D. Ruiz-Lopez, *Development of a marker of estrogenic exposure in human serum*, Clin. Chem. 41 (1995), pp. 1888-1895.

21. A. M. Soto, T.-M. Lin, H. Justicia, R. M. Silvia, and C. Sonnenschein, *An "in culture" bioassay to assess the estrogenicity of xenobiotics (E-SCREEN)*, In *Chemically-Induced Alterations in Sexual Development: The Wildlife/Human Connection*, (T. Colborn and C. Clement, Eds.), Princeton Scientific Publishing, Princeton, NJ, 1992, pp. 295-309.

22. A. M. Soto, C. Sonnenschein, K. L. Chung, M. F. Fernandez, N. Olea, and F. O. Serrano, *The E-SCREEN assay as a tool to identify estrogens: An update on estrogenic environmental pollutants.*, Environ. Health Perspect. 103(Suppl7) (1995), pp. 113-122.

23. A. R. Cunningham, D. M. Consoer, S. A. Iype, and S. L. Cunningham, *A structure-activity relationship analysis for the identification of environmental estrogens: The categorical-SAR (cat-SAR) approach*, In *Endocrine Disruption Modeling*, (J. Devillers, Ed.), CRC Press, 2009, pp. 173-198.

24. H. Fang, W. Tong, L. M. Shi, R. M. Blair, R. Perkins, W. Branham, B. S. Hass, Q. Xie, S. L. Dial, C. L. Moland, and D. M. Sheehan, *Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens*, Chem. Res. Toxicol. 14 (2001), pp. 280-294.

25. R. M. Blair, H. Fang, W. S. Branham, B. S. Hass, S. L. Dial, C. L. Moland, W. Tong, L. Shi, R. Perkins, and D. M. Sheehan, *The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands*, Toxicol. Sci. 54 (2000), pp. 138-153.